The “Immutable Me”

Data Provenance Fuelling New Value Chains

George Samman & Katryna Dow

5 May 2016

Problem Statement

How can people link all the data they generate to their identity and control the components in a verified, connected fashion and have all those identity points when they need them most? (Identity theft, man in the middle attacks)

Contextual Identity is an important element of identity. Can linking those attributes on a ledger(s) provide a greater degree of "I am who I say I am?"

The Opportunity

Up until now the power to capture, analyse and profit from personal data has resided with business, government and social networks. What if you and I had the same power? (Meeco Manifesto 2012[1]).

According to the PWC Report Identity 3.0[2], "Developed economies are moving from an economy of corporations to an economy of people. More than ever, people produce and share value amongst themselves, and create value for corporations through co-creation and by sharing their data. This data remains in the hands of corporations and governments, but people want to regain control. Digital identity 3.0 gives people that control, and much more."

Identity is moving beyond issued instruments like passports and social security cards and ID cards. It is moving towards ***contextual identity*** which is "I can prove who I am" (persona) in the context of what I am doing. Government issued identity instruments are relatively easy to forge. Stolen identities are equally exploited through organised crime and on-line hacking.

Conversely, personal attributes, behaviour, social and reputational data is difficult to forge or replicate over an immutable timeline. Increasingly the sum of the parts of our digital exhaust and social presence creates a strong identity. However the opportunity for individuals to use this is limited.

Proposal

The movement from User Centric Identity to Self Sovereign Identity is underway and becoming a key trend for the future of how individuals will control their own attributes. Using blockchain technology to empower data provenance based on the individual controlling his own attributes, Meeco plans to be at the forefront of this movement and is currently working on a proof of concept that allows individuals to add verification and provenance to the existing control and permission of their attributes (and the personal information they choose to associate with their identity).

Meeco aims to be Blockchain agnostic (since what type of ledger is used will be use case dependent), thus enabling individuals to link provenance to data/attributes to support a range of personas and progressive disclosure. This approach also supports the option for individuals to use private, public, permissioned and permissionless chains.

[1] Meeco Manifesto: https://meeco.me/manifesto.html

[2] PWC Report Identity:

http://www.chairdigitaleconomy.com.au/wp-content/uploads/2015/08/Digital_Identity_30_20150831_FINAL.pdf

The identity pieces (data and attributes) that are shared will be context sensitive. Individuals can thus create context-based personas through unifying relevant attributes across different chains with the added value of provenance.

Personal control is central to increasing the power individuals hold over the range of attributes and claims that make up their identity. Enabling identity markers to be thin sliced, refined and contextual provides added privacy protection prior to sharing. The combination of attribute, verification and provenance provides the capability for data governance to move from data collection, i.e. personally identifiable information (PII) to binary pull requests, i.e. date of birth versus over 18 = yes/no.

This approach provides protection as the individual solely has the power to bring these attributes together with the added value of verification. For the relying party the option exists to store the provenance rather than the attribute on public and private blockchains and distributed ledgers, thus providing an immutable audit trail for assurance with out the compliance risk of collecting, managing and holding the data.

What Is Provenance?

Provenance refers to the tracking of supply chains and provides a record of ownership via an audit trail, which can be ordinated (the specific order matters). In the case of attributes and claims it is important that all data points link back to a master record of identity (the golden record) and that this is shown to be immutable.

It's important for purposes of integrity, transparency and counterfeiting that this asset and its path be known or provable every step of the way. A supply chain refers to the creation of a network in which an asset is moved, touching different actors before arriving at a destination. It helps bring time and distance together in a manageable way. The tracking of this asset has big value to the actors involved for the reasons mentioned above. This equally applies to identity and all the components that make up that identity.

This approach is a pathway to turning data (single attributes) into networked data flows that compound in value and become assets similar to financial synthetics such as asset backed securities (ABS) as much as anything else in the world considered to be an asset including: Norwegian salmon, diamonds, prescription drugs or Letters of Credit (LOCs) and Bills of Lading (BOLs).

In order for data provenance to be effective from a technology standpoint it must fulfill certain requirements and characteristics that would upgrade supply chain management and monitoring. According to Sabine Bauer in his paper titled "*Data Provenance in the Internet of Things*" they are:

Completeness: Every single action, which has ever been performed, is gathered.

Integrity: Data has not been manipulated or modified by an adversary.

Availability: The possibility to verify the collected information. In this context, availability is comparable to auditing.

Confidentiality: The access to the information is only reserved for authorized individuals.

Efficiency: Provenance mechanisms to collect relevant data should have a reasonable expenditure.

From the perspective of Meeco we believe these requirements require the additional characteristics of:

Privacy: the ability to determine the access and visibility of personal data.

Unlinkability: a state where this personal data must be secure and not be able to leave the system and get into the wrong hands,

Transparency: which is simply total transparency of the chain and total traceability of the data particularly when it comes to actions and modifications.

A blockchain contains all of these requirements. For provenance to work really well on a blockchain, the whole ecosystem (all parties involved) needs to be a part of the ledger so that network effects make for a much more robust supply chain and the associated integrity of the data that comes from it.

How Does Meeco Link Data Provenance To Attributes on a Blockchain?

Blockchain allows for Digital Identity 3.0 Quoted from the PWC paper:

> *Digital identity 3.0 is a private and integrated master record that exists independently of any immediate commercial or legal context. It empowers people to create new attributes, share these attributes selectively as they connect with others, and create experiences and value beyond what can be predicted.*
>
> *The first part of this definition refers to a privately owned and integrated master record. This is our definition of the 'digital identity' part. A master record is a set of attributes of a person or thing. Contrary to a transaction record, a master record is not transient; it has a long lifespan. Some of the attributes in the record, however, may be transient or dynamic (see fine print). A private master record is uniquely related to one person or thing, is not part of a larger repository of master records and is managed by the person or thing the attributes refer to.*
>
> *In other words, this first part of the definition refers to a set of attributes that are linked to the person or thing who owns and manages it, including which attributes and clusters of attributes the owners may want to add and maintain. Indeed, the owner can write, read, change and delete any attribute, and has full control over who else can execute these functions.*
>
> *This brings us to the second part of the definition, which states that this private master record exists independently of any immediate commercial or legal context. Most master records that exist today are created and maintained because they serve*

a commercial or governmental purpose. Organisations keep master records of their customers to have a stable point of reference for all their interactions including quotation, delivery or invoicing. Aggregating such transactions over time allows insightful customer analytics. Governments keep master records about citizens for various public services related to licenses, permissions and personal records. The individual is merely a stakeholder in all of these scenarios: a citizen, a customer, a patient, etc. The private master record we describe in this paper, however, does not exist to create value for organisations or governments—its primarily creates value for people.

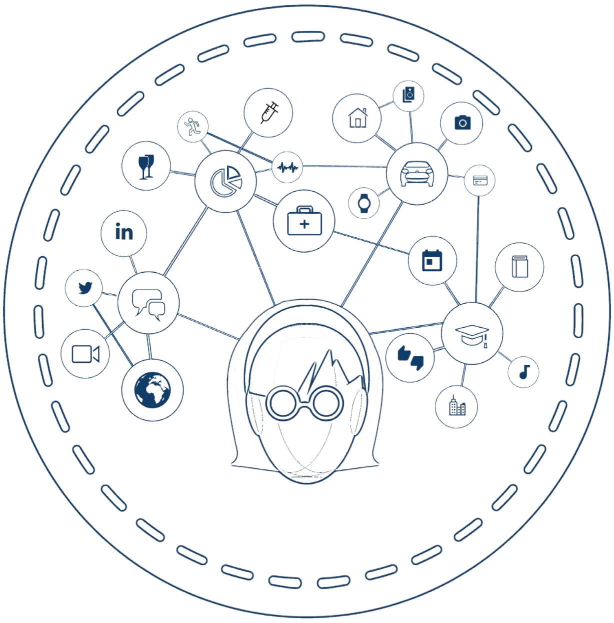
The third part of the definition refers to the platform element, stating that it empowers people to create new attributes, share these attributes selectively as they connect with others, and create experiences and value beyond what we can predict.

Digital identity 3.0 serves the purpose of enabling people to create value. That's why we call it a platform. Platforms allow users to create, exchange and consume value

People want to control how their digital identity grows and is maintained (write and change). Other parties—including devices and things—may be given rights to add, update or authorize data, but users want to be able to have full control of the user management of their digital identity.

People also want to control who has access to which information and for how long (the read function). People should be able to share single or logical clusters of attributes (see figure 1), and these clusters should be personalized automatically yet open to be changed by the user. It should also be possible for clusters of attributes to be shared without disclosing an identity. A person could, for example, share information for a quote about an asset that is to be insured (e.g. a car), without sharing information about the owner of the asset.

Finally, people want to be able to withdraw information after it has been shared (the delete function) and know what third parties they shared their information with.



Digital Identity 3.0, Figure 1
http://www.chairdigitaleconomy.com.au/wp-content/uploads/2015/08/Digital_Identity_30_20150831_FINAL.pdf

Master Record

In summary the primary purpose of a master record is to create value specifically for the individual. However, the existence of this master record can in return (if the existence of the record is permissioned) create significant value for receiving parties; i.e. organisations, enterprises, institutions and other individuals.

The master record will not be stored or visible on the blockchain. The intention is not to (ever) permission the master record, but to reference back to its immutable existence. This way the master record can support infinite links to data attributes of association, without linking all the attributes in one chain. This master record will have an anonymous unique identifier that is only known to the owner via private keys.

Provenance tracking

Tracking the origin and movement of data across a supply chain in a verifiable way is a difficult thing to do. In supply chains stretching across time and distance, all of these items could suffer from counterfeiting and theft. Establishing a chain of custody that is authenticated, time-stamped and replicated to all interested parties is a paramount to creating a robust solution.

The problem can be addressed using blockchains in the following way:

1. When the data is created, a corresponding digital token is issued by a trusted entity, which acts to authenticate its point of origin (the master record).

2. Then, every time the data changes hands that is associated with that identity, the digital token is moved in parallel, so that the real-world chain of custody is precisely mirrored by a chain of transactions on the blockchain. A tokenized asset, is an on-chain representations of an item of value transferred between participants. In this case an identity.

This Identity token is acting as a virtual "assertion of identity", which is far harder to steal or forge than a piece of paper. Upon receiving the digital token, the final recipient of the physical item, whether a bank, distributor, retailer or customer, can verify the chain of custody all the way back to the point of origin.

This digital identity token can also act as a mechanism for collectively recording and notarizing and linking *any* type of data back to the master identity record. A blockchain can provide a distributed database into which all the records of our assertions about a piece of information about us (data) are written and linked back to us, accompanied with a timestamp and proof of origin that ties back to the Identity token in a most verifiable way. An example could be a hash of a record that a certain element of my identity was verified when engaged in a certain type of action. This distributed database also stops corruption and theft by storing the multiple pieces of our identity in a highly distributed manner that we with the proper keys can open and put back together.

An example of this problem is with how companies take our data from us by tracking it via the identity we provide to them. They use this data to target individuals with the aim to mold and influence behavior. Currently, our identity does not control the attributes and elements that link back to it and worse, we don't get to realise the value of, or monetize the data companies collect on us.

How Will The Data Get Stored?

The Multichain blog[3] eloquently talks about how data is stored on the blockchain and this is approach that would be taken for the scenario described above:

> *"In terms of the actual data stored on the blockchain, there are three popular options:*
>
> 1. ***Unencrypted data**. This can be read by every participant in the participating blockchain, providing full collective transparency and immediate resolution in the case of a dispute.*
> 2. ***Encrypted data**. This can only be read by participants with the appropriate decryption key. In the event of a dispute, anyone can reveal this key to a trusted authority such as a court, and use the blockchain to prove that the original data was added by a certain party at a certain point in time.*
> 3. ***Hashed data**. A "hash" acts as a compact digital fingerprint, representing a commitment to a particular piece of data while keeping that data hidden. Given some data, any party can easily confirm if it matches a given hash, but inferring data from its hash is computationally impossible. Only the hash is placed on the*

[3] Four genuine blockchain use cases

http://www.multichain.com/blog/2016/05/four-genuine-blockchain-use-cases/

> *blockchain, with the original data stored off-chain by interested parties, who can reveal it in case of a dispute."*

For the purpose of Meeco, option (3), hashed data is our area of focus. Meeco advocates the data be at rest in a disturbed network, e.g. financial data might remain with the issuing bank, medical records with a physician and student records with the school admin system. In turn all these records can be augmented by the individual records, stored in personal clouds with read and write access based on permissions. Where the Meeco API of Me is used for a direct implementation between peers (individuals and organisations) connected parties can benefit from real-time, read, update, and audit records.

Only the individual will have the ability to federate attributes into a single context specific view.

Data at rest can reside in a personal cloud and/or enterprise data store. The value created is how these data sources interoperate through an audit trail and order of tasks.

It is the sum of the parts that creates a strong, sovereign an immutable series of records. When backed by assertions, verifications and provenance, this master record of records becomes a value personal asset. In this scenario individual would enjoys the same rights and value that governments and institutions currently have.

The ultimate aim is to provide individuals with the means to collect and connect attributes that strengthen their personal and context based assertions. This may include 'I am' statements such as:

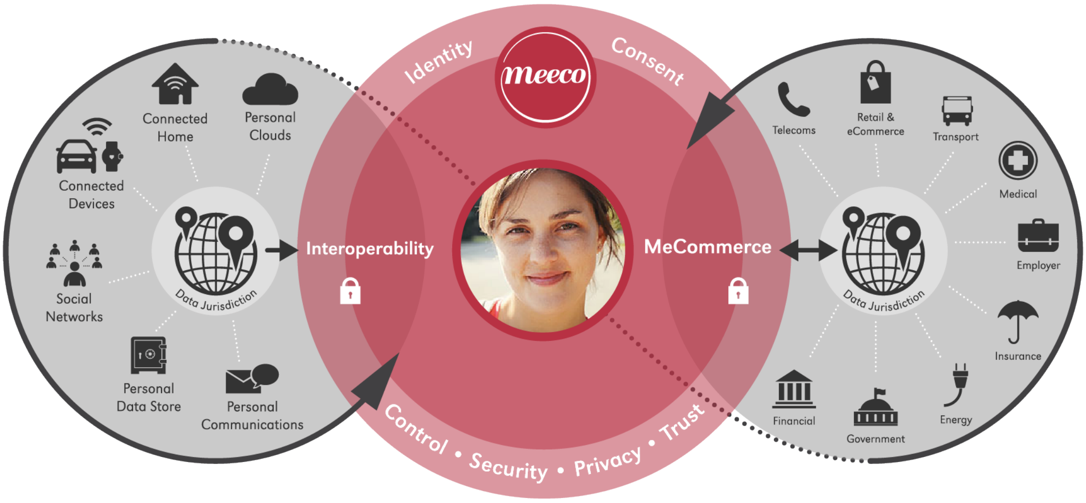
- I am a citizen of
- I am a qualified professional in the field of
- My income can support my application for
- I am old enough to access this service
- My delivery address for today is
- I can be discovered by

Conclusion

As we enter what some are calling the 4th phase of the Internet, which is defined by the right of individuals to assert things like their attributes and become sovereign a blockchain or distributed ledger will act as an enabler and supporter of that value chain.

In this context we are moving beyond send, search and social to the vantage point of self-sovereignty.

This is the vantage point by which the individual generates identity through every day activities, assertions, attributes, claims and context. By placing the individual above the attributes collected by and about them, they are able to orchestrate a value flow across the silos of their life, fuelling a multisided marketplace.



This approach supports the concept of ‘minimum viable identity information’. This is dependent on the situation, use-case and context. This is why personas become so critical as a means to protect identity, minimize the cost of collection, increase compliance and enable new networked value flows. This network of values flows enables individuals to engage in transactions for a particular need at a particular point in time for a particular desired outcome.

There will always be a core identity that is the human being (Person A) – but there are increasing security, fraud and counterfeiting issues with the current model of collecting, storing and enabling government and enterprise employees to read, handle and access personal information.

Digital Identity 3.0 together with blockchain and distributed ledger technology takes us towards the vision that we (humans), may become the custodians, issuers and provenance providers of our identity.

Glossary

The following terms are relevant to this article. These are just a subset of the terms generally used to discuss digital identity, and have been minimized to avoid unnecessary complexity.

The below definitions come from the blog of Chrsitopher Allan, unless otherwise referenced: *http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html*

Attributes

Every Digital Identity has zero or more identity attributes. Attributes are acquired and contain information about a subject, such as medical history, purchasing behaviour, bank balance, age and so on.[4] Preferences retain a subject's choices such as favourite brand of shoes, preferred currency. Traits are features of the subject that are inherent, such as eye colour, nationality, place of birth. While attributes of a subject can change easily, traits change slowly, if at all.

Authority

A trusted entity that is able to verify and authenticate identities. Classically, this was a centralized (or later, federated) entity. Now, this can also be an open and transparent algorithm run in a decentralized manner.

Claim

A statement about an identity. This could be: a fact, such as a person's age; an opinion, such as a rating of their trustworthiness; or something in between, such as an assessment of a skill.

Credential

In the identity community this term overlaps with claims. Here it is used instead for the dictionary definition: "entitlement to privileges, or the like, usually in written form"23. In other words, credentials refer to the state-issued plastic and paper IDs that grant people access in the modern world. A credential generally incorporates one or more identifiers and numerous claims about a single entity, all authenticated with some sort of digital signature.

Identifier

A name or other label that uniquely identifies an identity. For simplicity's sake, this term has been avoided in this article (except in this glossary), but it's generally important to an understanding of digital identity.

Identity

A representation of an entity. It can include claims and identifiers. In this article, the focus is on digital identity.

Permission / Permissionless

A **permissioned** system is one in which identity for users is whitelisted (or blacklisted) through some type of KYB or KYC procedure; it is the common method of managing identity in traditional finance.67 In contrast, a **permissionless** system is one in which identity of participants is either pseudonomyous or even anonymous. Bitcoin was originally designed with permissionless parameters although as of this writing many of the on-ramps and off-ramps for Bitcoin are increasingly permission-based.

http://www.ofnumbers.com/wp-content/uploads/2015/04/Permissioned-distributed-ledgers.pdf

Personal Data

The definition of personal data is evolving. Traditionally, that definition was pre-determined and governed through the use of a binary approach: In most jurisdictions, the use of personally identifiable information (PII) was subject to strict restrictions whereas the use of non-PII was often uncontrolled. However, what is considered personal data is increasingly contextual; it changes with personal preferences, new applications, context of uses, and changes in cultural and social norms.

Traditionally, organizations have used a variety of techniques to de-identify data and create value for society while protecting an individual's privacy. Such data was not subject to the same rules as PII, as an individual could not be identified from it. But technological advances and the ability to associate data across multiple sources is shifting boundaries of what is or is not PII, including potential re-identification of previously anonymized data.

This issue is the subject of significant debate with some arguing that this means that all data is effectively personally identifiable and should be treated as such. Others urge caution, arguing that this would curtail many of the beneficial uses of anonymous data with minimal gains in privacy. A shift in approach to thinking less about the data and more about the usage could offer a way forward. If the usage impacts an individual directly it would require different levels of governance than data which is used in an aggregated and anonymized manner.

http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf

Links

- http://db.cis.upenn.edu/DL/fsttcs.pdf
- https://en.wikipedia.org/wiki/Provenance
- http://siis.cse.psu.edu/pubs/tapp10.pdf
- http://siis.cse.psu.edu/provenance.html
- http://www.cse.psu.edu/~sem284/pubs/acsac-2012.pdf
- https://www.cs.indiana.edu/ftp/techreports/TR618.pdf
- http://cs.iit.edu/~dbgroup/pdfpubls/GD07.pdf
- https://web.sec.uni-passau.de/projects/compose/papers/Bauer_Data_Provenance_in_the_Internet_of_Things.pdf
- http://vhosts.eecs.umich.edu/db//files/ipaw08.pdf
- http://www.secureidnews.com/news-item/canadian-council-aims-for-trusted-identities/
- https://diacc.ca/wp-content/uploads/2015/05/DIACC-Building-Canadas-Digital-Future-May-6-2015.pdf
- http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html
- http://www.windley.com/archives/2016/04/self-sovereign_identity_and_legal_identity.shtml
- http://www.chairdigitaleconomy.com.au/wp-content/uploads/2015/08/Digital_Identity_30_20150831_FINAL.pdf