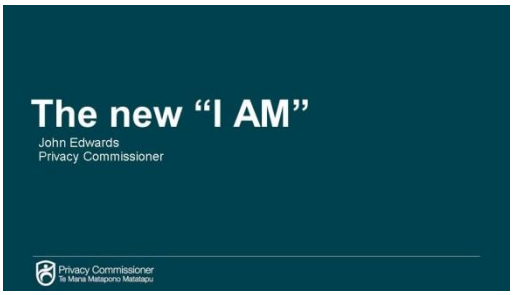
The New “I AM”: Identity and anonymity in an untrusting world

PRESENTATION BY PRIVACY COMMISSIONER JOHN EDWARDS

2015 IDENTITY CONFERENCE:
ENABLING DIGITAL IDENTITY AND PRIVACY IN A CONNECTED WORLD
TE PAPA TONGAREWA - 18-19 MAY 2015



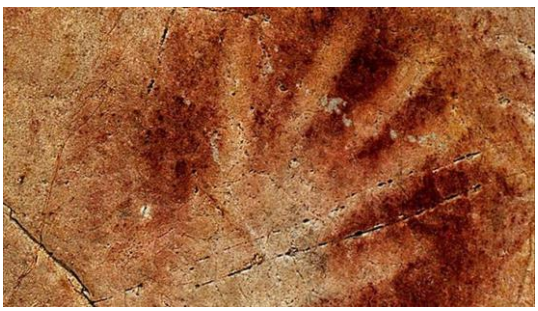
In a world of constant connection, identity is becoming less fluid; we are tying ourselves ever tighter to our pasts. Facilitating anonymity is one answer, but is it the best one? And if we achieve a world where anonymity is effortless and free, what will that mean for the intricate web of systems we rely on every day?

In the mid 1980s, when I was in the process of forming my own identity, almost every day I walked past this imposing artwork, Gate III, by Colin McCahon. It used to hang in the Maclaurin Building at Victoria University, and can now be found in Rutherford House. It is a landscape overwritten with biblical text, from Isaiah, and Jeremiah of the Old Testament, but the dominant text or image, or both, is the massive “I AM”. This text is a fundamental assertion of selfhood, of identity. Clipped from its original New Testament contexts “I am the truth, the way and the light:

I am the resurrection” it appears as an existential crie de coeur, a statement of individuation and differentiation from the crowd, a call to be noticed, to be taken account of.

I thought I’d take it as a bit of a theme for my introductory comments to the conference, fitting, because VUW is one of the three organisers of this event, together with my Office and DIA, and fitting because of the venue, Te Papa, which is the keeper of our cultural identity.

This impulse to assert and differentiate oneself is a feature of the very earliest artistic endeavours produced by our species and perhaps is one of the features that differentiates us from other species.



It’s a stencil of a hand, from the Gua Ham cave in Borneo it’s around 40,000 years old. Could this be the first biometric? A Mesolithic biometric?

A sense of identity is something about our innate humanity. It is irrevocably bound to other fundamental characteristics of humanity, dignity, autonomy.

Today, like everybody else here, I am going to talk about “identity”. What is it? How does it form, how is it assigned? Can it change? How do institutions respond to assertions of identity?

Although assertions of identity are as old as humanity, the concept of identity is changing. I want to look at some of those changes and their implications. I hope to explore the dichotomy between the need for some actors in society and the economy to be able to rely upon a trusted and “verified” assertion of “objective identity” of the individuals with whom they are interacting, and the desire for others to preserve their autonomy by going about their business unidentified, and unobserved, without any assertions of identity.

I want to consider the phenomenon of “identity” as an accretion of records, to look at the means by which we “identify” individuals, or “de-identify” datasets.

To the beginning. “Identity” means different things to different people, different disciplines. A philosopher, a psychologist, a political scientist, a database controller and a regulator will all have quite different but equally legitimate understandings of the term. Descartes, for example in his famous proclamation “Cogito ergo sum”, or “I think therefore I am”, meant something quite different from his “I am” than did McCahon.

Psychological constructs give us the origin of the word, identity is the entity which contains the id, the personality structure that contains a human's basic, instinctual drives.

We have seen a lot in popular discourse recently about “gender identity”.

When I was at university the academic journals were full of hand wringing dissertations on whether New Zealand could be said to have a distinct “national, or cultural identity”.

Is identity subjective, how we feel about ourselves, or is it objective, how and who we are observed to be? What happens when we decide that identity is or can only be one of these?

A new addition to the lexicography is “identity theft”. The greater the confidence institutions have in the reliability of their system of “identity” and “identification” the more devastating the consequences of “identity theft”. The less a business or organisation requires us to do to “prove” or verify, or authenticate our identity, the more vulnerable we are to having someone else impersonate us, acquire some advantage based on the misidentification, and leave us tainted and burdened by the interloper’s mischief. We need strong and reliable and secure means of demonstrating our identity to avoid those taints and burdens.

Lets go back to that great biblical, Cartesian, artistic assertion “I AM”.

In one sense we might consider our identity (the “I”) to be a combination of Actions and Memories.

First comes the impulse from that Mesolithic artist who stood in that cave in Borneo and said, in an moment “I will make my mark”.

Second the action; red ochre blown across his hand onto the wall;

Third the memory held by the space, the actual physical mark left on the wall

Those three things, the impulse and the action, and the memory of that action, make up identity.

And identity is therefore a very particular statement

I am here, and this is me.

The world into which that statement is made has changed utterly. In 2015 when you're seized by a notion to say 'I am here' you can write a book, post a blog, write a song. A moment of impulse can produce a tweet, a Snapchat, YouTube video or comment, Instagram, Facebook, Vine. But those impulses, giving rise to actions endure. They become part of the collective memory of who and what you are, they become your identity.

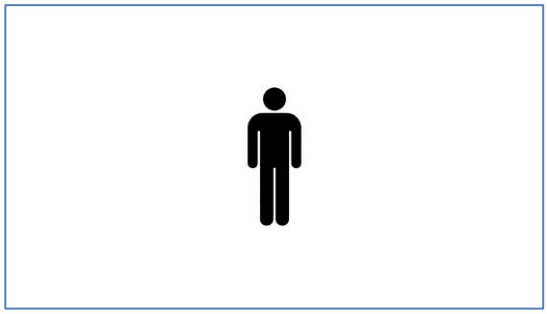
Of course you don't need to tweet, or post comments online, or click a "like" button. But even if we do none of those things, we are, every moment recording our actions in a way that has never occurred before. Our location is known by not just the phone we carry with us and the network provider (who requires "evidence of identity" before issuing a SIM card to us), but increasingly a range of other sensors, and devices, in our cars, on our computers, on our wrists, in our credit cards and in the labels on our clothes. All of these contribute to our identity as perceived by that telecommunications carrier, web platform, credit card company.

This is not necessarily happening in a particularly sinister way. Most of the time these watchers don't care too much who you really are, what they care about is a person-shaped someone who looks like you. The watchers are more often not even human, just algorithmic suckers and aggregators of data, linked by assumed unique and constant features. A name. An IP address. A telephone number.

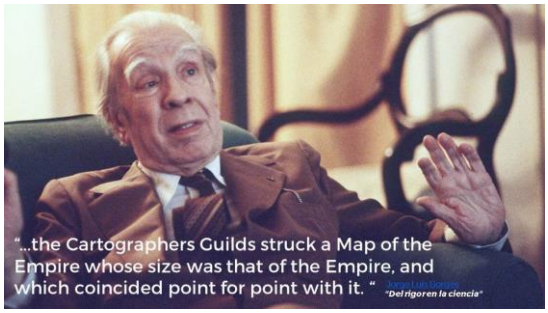
They do this because when they have enough information they can be reasonably sure that person shaped someone will also act like you. And that knowledge is worth money, if you're a business, and it's worthwhile if you're the government, because knowing your population means you are better at helping them or doing whatever else you decide you need to do.

There is a constant evolutionary pressure for both business and government to get better and more information about what people have done, and what they are, and who they are to read the traces of action, and collate the memories that people have left about themselves.

And inevitably, it seems to me, as the databanks grow and join up the person shaped block of data that looks like you will come to resemble and be relied upon as if it were an exact portrait of you that it is indistinguishable from you, the person.



Jorge Luis Borges once wrote a story about a gigantic map. This map grew, and grew, and kept growing along with the ever-expanding desire of its cartographers to accurately represent the space it depicted until eventually the map was exactly the same size as the country it described, point for point. Hundreds of miles across. Naturally, the map was completely and utterly useless. The story ends with the map being left to rot.



There is a parallel with the endlessly detailed digital portraits that the world is painting of each of us.

But the decaying part is not going to happen with big data because we have conquered space when it comes to data storage. Moore's Law, that postulates exponential growth of computing power over time, is well known.

What's equally important is Kryders Law, which suggests a similar growth in storage capacity and drop in storage cost. The gigabyte that cost a few hundred thousand dollars in 1980 costs a few cents now.

We are always going to have more space to store data, and cheaper space to store data, and we are always going to have more memory. So we have conquered memory, and therefore, with our machines, we have conquered forgetting. As Bruce Schneier suggested 'we're embarking on a great experiment of never forgetting'.

These objectively observed Actions and Memories, these digital traces that are aggregated to give what Europeans call the "data controller" a version of your identity that can be misleading. We have less control than any time in history over what facts are recorded and aggregated to make up what others take to be our identity. We have created machines that are so much better than us at remembering that we threaten to make any subjective notion of identity obsolete. All can be tabulated by observing our actions as we move through the world.

To get a feeling for the nature and pace of change, think back to your own infancy.

A baby is born, consists of name (maybe), sex, date of birth, weight. That's pretty much all you can say about a baby and that information is passed among friends like it's precious because of that, but from birth little stories start happening, and we in turn record those.

I don't know about you, but I don't believe I could easily lay hands on a single photo of myself up to the age of five. I have a handful of photos of my mother as an infant, and one or two of her mother as a child around about the turn of last century.

By comparison a modern parent with a smartphone and a fast finger on the shutter button might have 15,000 photos of their child by the time they are ready go to school. But it's not just the number of the photos, its their nature and location. They are digital, and many will be stored in the cloud. If just one of those is "tagged" with a name, very soon, the capacity will exist for all those to be linked by facial recognition software and connected with a vast array of other data points to give a rich and diverse longitudinal record of that child's identity into adulthood and forever more. Add to that the hundreds of thousands of data points collected and retained by the government, business, social networks,

When a 21st century baby is all grown up he or she will have a life that has been recorded, scanned, and meticulously curated by careful machines. They'll have a fantastically intricate, logged, searchable external record of their identity, to go with their internal sense of self. There will be value in that.

Perhaps we will see an increasing demand for "a right to be forgotten" (or the "right to oblivion" as the French so cheerfully call it) more tightly regulating the circumstances in which agencies must delete, or de-link personal information, or a right to at least take data with you when you leave a platform.

We might see people taking greater steps to reduce the amount of data that can be attached to their identity. Might we all start conducting our online interactions using anonymising TOR browsers? We've seen an enormous appetite for WhatsApp, with its confidential messaging service protected by end to end encryption enjoying phenomenal growth. In April they hit a record of 800 million monthly active users.

Is guaranteed anonymity online the answer, or even a social good? Plato's famous and mythical ring of Gyges, made the wearer invisible. Granting them the power to do what they wanted without consequence and in the story, transformed the wearer into a sociopath.

I've seen the same phenomenon on Twitter, and ask.fm. Bullying, abusive behaviours shielded by anonymity, and then the almost equally obnoxious "doxing", revealing the identity of others trying in vain to maintain their anonymity.

Or perhaps we will see a change in societal and cultural values that will see the next generation being more tolerant of the universally known; more forgiving of the still recorded but now irrelevant historical indiscretion.

Perhaps it will be those who take steps to minimise their digital footprint, to "opt out" of the endlessly enumerated identity that will arouse suspicion. We already regard those who insist on using cash in this increasingly electronic economy as anomalous anachronisms – even to the extent that anti-money laundering laws assume a presumption of criminality for those whose mistrust of banks leads them to do their business with wads of notes.

We are already seeing companies betting the farm on the market appetite for "the quantification of the self", the devices that record your workout, heartrate, food intake, blood pressure and the like. But there is also a risk that conflating observable objective characteristics and actions with "identity" can reduce the space for self identification. For individualisation.

An imposed identity denies self identification, denies nuance and context and undermines dignity. Its extreme denies humanity – who and what you think you are means nothing.

We have decided you are the sum of your online interactions, or purchases, or contacts, or your religion.

Here we see the ultimate in individual identity at once denied, and bureaucratic identity imposed, the dehumanising "just a number".

Don't just tick "Godwin's Law" on your privacy conference bingo card. Although we have almost become inured to these kind of images today, and it is easy to consider them hyperbole when invoked in any discussion about "where things might end", it is impossible to understand the prescriptive, rights-based Western European approach to data protection and privacy without understanding the crucible from which the modern human rights instruments were formed.

An official sanctioned, assigned or mandated, or even a commercially monopolistic identity can become a trap which limits one's economic or social participation. How does an individual assert an identity that is at variance with the "official" or even commercial record?

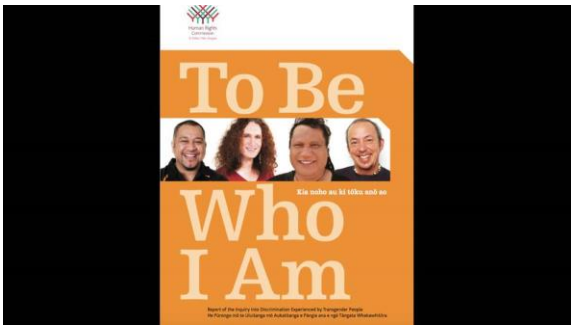
The French equivalent of my office, the Commission Nationale de l'informatique et des Libertés, was created in 1978 after a proposal known as Safari caused great consternation in France. The proposal was to identify each citizen with a unique number, and use that number to interconnect all government records. The plan was scrapped leaving only the CNIL to watch out for people's liberties, and a brand name that Apple could grab for its browser 30 years later. Now many of those liberties are up for grabs again in France, as legislators there struggle to address home grown terrorism in the wake of the Charlie Hebdo attacks.

Similarly here, access to identity documents, and the connectedness of government information systems is under scrutiny, after convicted murderer and child rapist Phillip Smith/Traynor managed to get a passport and leave the country while on short release from prison.

There is an inquiry under way in relation to that so I won't speculate on the issues confronting the inquirers, or their possible recommendations in response, but it is hard to imagine the processes by which passports are available to inmates will be loosened as a result? Does that mean it will be more difficult for all of us? Will there be a check to see whether I or someone who looks like me, is serving a jail term next time I need to apply to renew my passport to urgently travel?

Passports might become more difficult to get for some as a result, and we have already increased the circumstances in which the passport, that most definitive identity document can be cancelled where a Minister things the holder might be involved in foreign jihads. The pressures on identity assertion, verification and authentication are myriad, and not just as a result of technology as we've seen. Global geopolitics can affect how we, in New Zealand see our rights and liberties.

Now that we have a more comprehensive than ever aggregation of supposedly "objective" data, of "proven" evidence of our interactions in the world, is there a risk that our level of confidence in that third party assignment of identity will go beyond what is justified? Will we crowd out any capacity for the subjective assertion of self?



We've already seen this in a number of institutions in relation to gender identity - an inability of rigid bureaucracies to accommodate the difference between the observed, supposedly objective characteristics of an individual, and that individual's subjective experience and sense of identity. This inflexibility can cause exclusion and significant distress. Note the title of this report produced by the Human Rights Commission on discrimination experienced by transgender people.

As I discussed earlier, one of the great anchors of "official" identity is evidence of birth. The date, the parents, the place, the name given to the child recorded in an official, incontestable register and certificate.

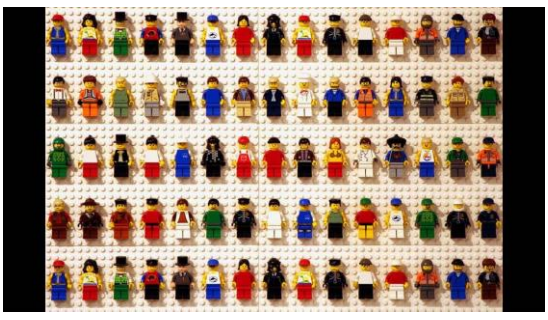
Could it be wrong? Can a date of birth change?

I am in the process of investigating a case in point at the moment. I haven't concluded that investigation so I need to slightly obscure some facts, and not identify the parties involved. But imagine this scenario. A refugee arrives in New Zealand from a failed state, without evidence of identity.

Anyone who has worked with refugees knows that this is a not an uncommon phenomenon. Family members attempt to procure documentation to satisfy New Zealand authorities, and later produce, from the refugee's village, a statement from the village authority that a person with the name of that name was born in that village on a specified date. That document is taken as the official record of identity and the age of the refugee is recorded as 15. But he is quite big. Doubts grow. He feels older, is out of sync with those officials have deemed to be his "peers". Medical examiners test and conclude he must be at least three years older. Bone density scans support that judgement. A visit back to the village suggests an error might have been made.

The individual's self identity is that of a maturing adolescent ready for social and economic interactions which are denied him because of his recorded age. His self identification, supported by extrinsic medical evidence is that he is of an age at which he is entitled to receive certain benefits, enrol in certain educational programmes, enjoy a range of other rights entitlements and privileges which are denied to him based on the "official", and "officially accepted" version of his identity.

Which identity should prevail?



When we speak of identity it is assumed that we are speaking of information associated with "identifiers". The Privacy Act for example is only concerned about "information about an identifiable individual".

An "identifier" is a name, or unique combination of data elements that allows us to connect an Individual with a set of data.

As we rush to embrace the undeniable benefits of “big data” and enhanced analytics we do need to pause from time to time and contemplate identity.

Does stripping name, date of birth, and address out of a data set effectively “de-identify” it?

Perhaps not.

Scott Peppet, a professor of law at Colorado School of Law, wrote in a 2014 paper on the Internet of Things that researchers at MIT recently analysed data on 1.5 million cellphone users in Europe over 15 months and found that it was relatively easy to extract complete location information about a single person from an anonymised data set containing more than a million people. In a stunning illustration of the problem, they showed that to do so required only locating that single user within several hundred yards of a cellphone transmitter sometime over the course of an hour four times in one year. With four such known data points, the researchers could identify 99 percent of the users in the data set. As one commentator on this landmark study put it, for sensor-based data sets “it’s very hard to preserve anonymity”.

Think about that. If I know where you live, and where you work, and make a couple of educated guesses about where else you might have been in a year (everyone comes to Te Papa once a year right? To a gas station somewhere close to you house?) I can have precise details of your movements over a whole year.

The other famous example is Lantanya Sweeney’s from as far back as the mid 1990’s. I’ve used this story before so forgive me if you’ve already heard it.

The Massachusetts Group Insurance Commission decided to release anonymised health data on state employees. Its aim was to help health researchers to improve healthcare. Obvious identifiers such as name, address and social security number were removed from the data. The Massachusetts Governor at the time, William Weld, assured the public that the Group Insurance Commission had protected patient privacy by deleting identifiers.

A graduate student in computer science at MIT, Latanya Sweeney, requested a copy of the data, and got to work. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For $20, she purchased the complete electoral rolls for Cambridge. These included the name, address, ZIP code, birth date and sex of every voter. Only six people in Cambridge shared Governor Weld’s birthdate. Only three of those six were men, and of them, only he lived in his ZIP code. Dr Sweeney had the Governor’s detailed health records, including diagnoses, prescriptions and details of hospital visits, delivered to his office.

Latanya Sweeney has continued to research in this area, and has revealed that our intuitive beliefs about how easy it is to identify an individual from a supposedly anonymous set of data are often misplaced. Among her findings she has demonstrated that 87% of the US population can be identified by birth date, sex and ZIP code alone. This is particularly startling when you keep in mind that the average ZIP code has a population of around seven and a half thousand. To put that in the New Zealand context, the average population of a Statistics New Zealand mesh block is about 90. For the next largest statistical unit, the “area unit”, the average population is 2100.

As more data sets are linked together there are an increased number of vectors for identifying a target. So, while you might not have information about a target's birth day, you might know what they studied at University. Or how many children they have, or whether they've ever been convicted of an offence.

As well as the risk to individuals, the ability to identify individuals within a large dataset can jeopardise the dataset's objectives. This is a feature of last year's debates in the UK over the government's decision to make detailed NHS data available to researchers through its care.data initiative.

Privacy campaigners pointed out that where the details of treatments were in the public domain, such as then Labour leader Ed Milliband's nose operation to cure a sleep apnoea, or then Deputy Prime Minister Nick Clegg's wife's broken elbow, it will be possible to identify the individual and 'read across' their broader NHS record. Well known medical science commentator, and advocate for greater use of data in public policy Ben Goldacre summed it up in the Guardian with the title "Care.data is in chaos. It breaks my heart." As he put it: "When you're proposing to share our most private medical records, vague promises and an imaginary regulatory framework are not reassuring."

The public outcry forced the government to delay the launch of care.data. An online tool developed by a private company has already been shut down. There's been a Parliamentary select committee inquiry.

There are new ways of identifying people that are still being identified. It is only relatively recently that it was discovered that each person has a unique gait. Back to Peppett:

This means that if I knew something about an individual Fitbit user's gait or style of walking, I could use that information to identify that individual among the millions of anonymized Fitbit users' data. I would then have access to all of that user's other Fitbit data, which would now be re-associated with her. As Ira Hunt, Chief Technology Officer of the Central Intelligence Agency, put it: "[S]imply by looking at the data [from a Fitbit] we can find out . . . with pretty good accuracy, what your gender is, whether you're tall or you're short, whether you're heavy or light, . . . [and] you can be one hundred percent . . . identified by simply your gait—how you walk."

These examples illustrate why we need to take care over concepts of identification and de-identification, and why we, while acknowledging the inevitability and desirability of deriving value from public data sets, have been very supportive of the Data Futures Forum process and its cautious and orderly recommendations for the oversight of innovative use of supposedly anonymised data.

We are increasingly seeing tensions between what individuals want of their online environment, and what the online environment seems to demand of them. A dichotomy is developing between individual consumers' and citizens' desires to safely transact, and have the means of asserting a trusted and authenticated identity, and their ability to move through the online world anonymously; unobserved and without contributing to that aggregating avatar, the digital doppelganger that with every mouse click presumes to know more about ourselves than we do.

Even when what we are seeing is the operation of an algorithm completely unmoderated by human judgement, that makes assumptions about who we are (or perhaps precisely because it is an automated process, presuming to draw its own conclusions about our identity) we can be diminished by that presumption. The Target example springs to mind. You will all have heard the story, from the New York Times, in 2012.

An analyst at Target was looking at purchasing history of women on the Target baby registry. He found that around about the beginning of the second trimester, a lot of the women had started buying unscented lotions. If you combined that with the purchases, within the first 20 weeks of supplements like calcium, magnesium and zinc, you were pretty sure to be able to begin marketing baby products to women quite early in their pregnancy. Here's what happened next:

An angry man went into a Target outside of Minneapolis, demanding to talk to a manager:

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologized and then called a few days later to apologise again.

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."

Now what gets me about that story (if its true) was that the father in that scenario felt that he owed Target an apology!!

I want to finish with another slide of a McCahon painting on the same theme as the one I started with. This one is called Victory Over Death 2.

We see that same bold statement of identity; I AM, but look to the left third of the canvas. There is a blackness, a void in what is otherwise a fully utilised space. If you look closely you can make out that the artist has painted over what was there before. There is a large AM, now obliterated with black. That confident positive assertion, "I AM" once was inverted to the more tentative, doubtful "AM I"? The artist knows that our sense of identity can be fleeting, and fragile, that at different times we don't even know ourselves who we are. Doubt and tenuousness is part of the human condition. Identity is mutable and fluid.

As Privacy Commissioner, I'm interested in watching where today and tomorrow's discussion goes, and how we here in New Zealand will respond to the challenges faced by many countries in an uncertain geopolitical climate, with tremendous advances in technology and data collection.

I'll advocate for privacy by design, which will be covered at a session today, for a proportional response to emerging threats, a desire which I am sure is shared by the Director of the Security Intelligence Service. I'll argue that we should resist "authentication inflation" and to the greatest extent possible that we should be able to go about our business in the world and online anonymously if we wish, and that we should not arouse suspicion if we do. I'll support initiatives like RealMe, the government identity service operated by New Zealand Post that can allow users to access a wide range of services, without tracking them across them.

With a reform of the Privacy Act coming up, I'm keen on having a conversation about a right to be anonymous or pseudonymous to the greatest extent possible, and what that might look like. I'm keen on looking at a prohibition on re-identification of data from "de-identified" data sets?

But on a day to day basis how do you and I respond to these increasingly externally applied identities? How do we preserve dignity and autonomy in the face of these ever confident coders? Maybe we just need to introduce doubt into their systems. Maybe, when confronted with the certainty of the algorithmic assessment of our identity, when some enterprise makes assertions, "You are John Edwards, you are a male, 49, atheist, Wellingtonian, father, partner, lawyer, pakeha, heterosexual, downloader of Game of Thrones."

Perhaps the best response is simply to cross ones arms, lean back a little, draw a breath, nod a little, and ask "Am I?"

THE END

***Acknowledgements**: Adam Art Gallery for the Gate III image, and Art and My Life and Justin Paton for the Victory Over Death analysis https://artandmylife.wordpress.com/2008/09/22/a-constant-flow-of-light/*