

Actors, Objects, and Linked Data

Daniel Hardman

2022-11-16

Abstract

This paper argues that data about people is ethically special, and that the W3C Verifiable Credentials data model mishandles that specialness by binding credentials to Linked Data. The argument takes a deliberately philosophical path: the old distinction between things that act and things that are acted upon, the human capacity for agency, and the social and technological harms that follow from objectifying people — treating them as resources to be processed rather than actors with boundaries of their own. Computationally there is no difference between data about a person and data about a stock price, but in awareness, consent, accountability, relationship, and change, human data is unlike any other. By requiring `@context` at issuance, the VC specification grafts human subjects into a global, ungovernable semantic graph built for correlation — whether or not anyone ever processes the credential as Linked Data. The author recommends making `@context` optional, at least for credentials about people, and shifting the community’s emphasis from perfect, web-scale data toward extensible actions and the humans the system is meant to serve.

(TL;DR)

Data about individual people is special — not in raw computation, but in the ethical implications it carries. The current W3C data model for credentials grafts such data into the Semantic Web in a way that I consider dangerous. When this is ignored, people are objectified. This is not just troubling, but downright dangerous. We can do better...

...The Full Story

Recently, an old debate has flared again in W3C circles.[1] At issue is whether, how, and how much to associate the next generation of W3C’s formal verifiable credential data model with Linked Data (LD).

I suggested that a weak association would be best. Proponents of strong LD binding characterized positions like mine as being rooted in a desire for developer convenience and less formal data processing. They asked if reducing friction would make the debate go away.

This is an excellent question. It caused me some pondering. My conclusion: No. Although friction matters, and although others might accept that resolution, I personally don’t want to make friction the focus of the conversation. For me, that’s not the heart of the matter. I said I’d try to explain why.

To keep my promise, I need to take you on a philosophical journey. I promise it will bend back to verifiable credentials.

I. Things to Act, Things to Be Acted Upon

Years ago I was impressed by wisdom from an ancient prophet-philosopher during a teaching moment with his children. He said that the universe could be divided into things to act, and things to be acted upon — and that the human race emphatically belonged in the former category.[2] Thinkers like Kant,[3] Levinas,[4] and Carl Rogers[5] have articulated similar thoughts.

The older I get, the more I find this mental model profound. Although humans can be modeled as objects in some contexts, it's crucial that we view them primarily as actors. Objectifying people — treating them as things to be manipulated or processed, rather than as action-capable others[6] with boundaries around an independently worthwhile reality of their own — is a theme running through many tragedies in micro and macro human history.

- Patterns of abuse are rooted in the abuser objectifying the victim, and always involve efforts to limit the victim's desire and/or ability to act.[7]
- War and genocide are preceded and followed by a mental model that divides the world into friends and enemies. Enemies are dehumanized, rendered as helpless as possible, and then brutalized.[8]
- Self harm and eating disorders arise as the suffering person internalizes objectified views of their self and their body, and seeks desperately for actions that redirect the pain.[9]
- Sexual harrasment, pornography, child pornography, prostitution, sexual assault, and human trafficking all rest on a foundation of objectifying people, dissociating them from their bodies or body parts, and manipulating their ability for independent action.[10]
- A hallmark of dictators and repressive governments is to view citizens merely as economic cogs, threats to the state, or cannon-fodder. These are object categories. Even in freer systems, politicians objectify and manipulate voters with sound bites and cynical coalitions.
- Racism, sexism, ageism, and other forms of prejudice are essentially driven by reducing human beings to an objectified stereotype, and then refusing to interact with them in the same way as an "in" group.[11]
- The mental health crisis triggered by the COVID-19 pandemic is associated with loneliness, isolation, and feelings of objectification and helplessness by many.[12]

II. Actors and Objects in Tech

As a computer professional, I've also observed with distress how this same actor-versus-object theme manifests in aspects of technology and the web where I make my living.

- Some of the most important companies on the Internet give away services in exchange for the ability to farm their users by analyzing behaviors and selling insights and access. This embodies the observation "If the service is free, YOU are the product" and is even more true today than when it was first suggested in 1973. Perhaps the business models of these companies can be made ethical, if user consent is truly informed, and if there are not unreasonable terms and conditions that present a Hobson's choice. But even so, the troves such systems assemble invite even more objectification when abused (e.g., Cambridge Analytica, distorting public discourse in an election).[13]
- Our quest for artificial intelligence is measured by the Turing Test, which takes as its standard for success the ability of a machine to masquerade as a human in a human interaction.[14] Engineers work hard to make Alexa and Siri tell us jokes, but Alexa and Siri objectify us by treating us only as input givers or purchasers. Personal assistants run on algo-



Figure 1: Photo by Obi - @pixel6propix on Unsplash

rithms and data rather than on empathy. They have no loyalty, and they don't laugh with us. SEO and clickbait optimize for "eyeballs" and "impressions" rather than for usefulness to people.

- Avatars have become pervasive. They create a layer of indirection: real human beings with tangible bodies and lived experiences are replaced with simplified, idealized projections. Connect this to violent video games, which invite players to spend hours killing avatars. Avatars may spatter blood but don't suffer, have backstories or relationships, or invite our empathy.
- Spam — 85% of all email by volume[15] — is an abuse of a recipient's interest and intentionality. It objectifies the recipient as a potential buyer whether they want to be seen that way or not. It also circumvents reasonable actions that a recipient takes in their own interest.
- Data breaches lead to human identities being auctioned on the dark web, with no recourse for victims other than depersonalized class action lawsuits.
- Fake news often objectifies the subjects of the news articles, who are reduced to heroes and villains — and it always objectifies the audience by reducing them to entities for manipulation.
- Chat bots, fake social media accounts, touchtone phone menus, and automated email responses connect people to technological projections of an identity that lacks empathy and human insight. Often the hard-coded features of these systems presents people with unrealistic or unhelpful choices that dehumanize them even as they optimize support costs.[16]

III. Agency

As I think about objectification, my thoughts go to Viktor Frankl's important book, *Man's Search for Meaning*. Frankl, a formally trained psychiatrist, describes his dehumanizing experiences in a concentration camp during World War II. He suffered, and so did so many others. But he tells how his suffering triggered an important epiphany about humans as actors. It informed his personal and professional activities for the rest of his life. Another author, summarizing Frankl's life work, distilled the idea in a way that's become a popular meme on the internet: "Between stimulus and response there is a space. In that space is our power to choose our response. In our response lies our growth and our freedom." [17]

Philosophers call this deeply human trait "agency." [18] It is the defining characteristic of things that act, and I believe it deserves our reverence and advocacy. This is because, although agency cannot be pried out of us, systems can deny agency its outward expression.

When systems fail to model and enable agency properly, suffering ensues. That's the essence of objectification.

IV. Data in Decentralized Identity

If we made a tag cloud of conversations in our community, one of the bolded words in a huge font would surely be "data." DIDs and DID docs are important forms of *data* to us. Credentials are *data*. The word *data* appears 257 times in the DID spec, and 263 times in the VC spec. Both specs standardize a *data* model and draw upon linked *data* theory. Microsoft identity hubs and web5 decentralized web nodes are all about *data* sharing. The VC API offers "protocols to issue, verify, present, and manage *data*." DIF Credential Manifests are a *data* format. The "D" in SOLID is *data*. So is the "D" in ACDCs. The CCG is working on encrypted *data* vaults. Wallets are described as

secure containers for our *data*. And I have heard numerous statements like these:

- “Whoever controls the *data* controls the identity.”
- “At the heart of SSI is the idea that a person should control their own *data*.”
- “It’s all about authentic (verifiable) *data*.”
- “We can’t have interoperability unless we have interoperable *data*.”
- “Empower customers to manage their own *data*.”

I agree that data is important. We need standards about our data, and paying attention to data is an important precondition to making identity better. However, I assert that identity is NOT all about data. It’s all about the identity subject. I’m not just throwing that out as a slogan or an opinion; the very definition of identity requires it to be *all about a subject*. And when that subject is human, then the identity is *all about an actor*. How big is *that* word in our tag cloud?

I think the reason we talk about data so much is because data is what the information economy can process to extract value. And processing to extract value is something that we do to objects, not actors. That worries me. It really worries me.

V. Human Data vs. Other Data

There is zero difference, computationally, between data about humans and data about any other subject. This is why the abstract of the DID spec makes the following point:

A DID refers to any subject (e.g., a person, organization, thing, data model, abstract entity, etc.) as determined by the controller of the DID.

It is also why the VC Use Cases Note includes a section about credentials for devices, why there are active discussions about VCs in IoT, and why GLEIF is pushing so hard for institutional credentials. We can process all these kinds of data the same way.

But notice my qualifying word up above. Right after “zero difference,” I said “computationally.”

There are other dimensions to data besides how computers process it. And in many of them, data about humans is unusual or even unique. Consider how uniqueness manifests in questions like these:

- Is the data subject *aware* of this data? If so, what is their degree of *comprehension*? What are their emotions about it? Does the data subject agree with it?
- Can the subject of the data change this data? To what extent is it *appropriate* to hold the subject *accountable* for whatever this data says or implies?
- If you see one piece of data about this subject, what other data would the subject like you to have as well?
- What ethical obligations does it create to see or process data about this subject?
- Do other data subjects claim to have a stake in the data about this subject (e.g., by virtue of a human *relationship*)? If so, on what basis? To what extent do we consider their claim *valid*?
- How does the data subject evolve with time and circumstances, and in what ways does that color the answers to all the previous questions?
- How do relationships between the data subject and others change with time and circumstances, and in what ways does that color the answers to all the previous questions?

Please note that that list isn't privacy-centric. I've been known as a privacy hawk in the community, and I am. But I deliberately didn't emphasize privacy in the list, because I'm making a point that is much larger than just privacy. I don't want the list to be distilled down to that one issue.

I claim that when we build data processing systems that do not account adequately for ways that data about humans is unique, we are on dangerous ground as far as objectification is concerned.

VI. Identity Data as Semantic Web Data

Semantic web proponents are another community that loves the word data. A W3C vision statement about it, dating all the way back to 2001, says:

The Semantic Web is about two things. It is about common formats for integration and combination of *data* drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the *data* relates to real world objects. That allows a person, or a machine, to start off in one *database*, and then move through an unending set of *databases* which are connected not by wires but by being about the same thing.

Nowadays, work on the semantic web lives in the W3C's Data Activity. This is the locus of RDF and Linked Data research and expertise. The community tagline is "Building the Web of Data." The group describes its intentions like this:

Data is increasingly important to society and W3C has a mature suite of Web standards with plans for further work on making it easier for average developers to work with graph *data* and knowledge graphs. Linked *Data* is about the use of URIs as names for things, the ability to dereference these URIs to get further information and to include links to other *data*. There are ever increasing sources of Linked Open *Data* on the Web, as well as data services that are restricted to the suppliers and consumers of those services. The digital transformation of industry is seeking to exploit advanced digital technologies... W3C is seeking to make it easier to support enterprise wide *data* management and governance, reflecting the strategic importance of *data* to modern businesses.

Now, I am grateful that someone is thinking about these things. I believe everybody benefits from better structure, better tools, and better connections among data. I want to turbocharge the information economy and unlock new value. So I am not down on the semantic web in theory.

However, I begin to feel some dissonance when we make human identity data part of a web that enables a "machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing," and where this semantic web aims to help industry and enterprise "manage" and "exploit" what they process. Those are W3C's words, not mine.

Let me be clear. I'm not imputing bad intentions to champions of the semantic web. Not one little bit. But I'm concerned because, among all the data types that might be web-ified, I see no provision for the specialness of data about humans. If human data is grafted into the semantic web, next to data about stock prices and the GDP of Mauritius and CC camera feeds, then the people described by the data are objectified, just like the rest of the subjects in the global graph.

Consider some other language from this community (emphasis mine):

Any IRI or literal denotes *something* in the world (the “universe of discourse”). These *things* are called resources. *Anything* can be a resource, including physical *things*, documents, abstract concepts, numbers and strings... (RDF 1.1 Concepts and Abstract Syntax, section 1.2)

“Resource” is an important word for this community. It is the “R” in IRI (URI) and RDF. In the HTTP spec, it is the noun category upon which all CRUD operations operate. Notice how passive and object-minded this word is; it means “something available for use.” And note its association with that other word, “thing,” in the quote... Per my observation that there is zero difference, computationally, between data about humans and data about other subjects, one of the physical things that can be a resource denoted/referenced by a URI is a human being.

This ability to refer to people with URIs is the basis of grafting W3C VCs into the semantic web.[19] For VCs where the subjects are not people, this feels fine to me. But for VCs about people (the vast majority of our use cases in the VC Use Cases Note and our community conversations; also the major locus of VC adoption), the effect is to place those human subjects into a global data graph to support the most ambitious correlation that our smartest data scientists can conceive. Correlation is, after all, what is meant by the W3C vision verbiage about “unending set of databases which are connected not by wires but by being about the same thing.” And this correlation is enabled whether the issuer, the holder, and the verifier know it or not, and whether they “process” the VCs as JSON-LD or not. The VC spec guarantees this grafting, by requiring @context at issuance. Contextualizing data into the larger web is the very purpose of @context.

Do not equate “correlation” with me harping about privacy here. Privacy is one thing affected by correlation — an important thing — but it is by no means the only thing. Other examples include the ability of a person to assert or curate a different reputation from proximate human friends, the ability of a person to impose terms and conditions on those who consume her corner of the global graph, the ability to tailor how the graph is traversed according to jurisdiction rules and according to the accountability and age of the subject, and the ability of a person to undo self-reinforcing graph cycles. What all of these problems have in common is that they arise because there is no mechanism in the semantic graph that a human could use to mediate between stimulus and response; consumed nodes and edges are just passive data objects.

How about this quote, from section 1.3 of the RDF 1.1 spec?

Perhaps the most important characteristic of IRIs in web architecture is that they can be dereferenced, and hence serve as starting points for interactions with a remote server.

Yes. I love that we can dereference URIs! The web that permeates our lives owes almost all of its power to that one little sentence. However, I don’t love that the only interaction that’s imagined is with a remote *server*. Maybe I have Moxie Marlinspike’s review of web3 ringing in my ears:

It’s probably good to have some clarity on why centralized platforms emerged to begin with, and in my mind the explanation is pretty simple: People don’t want to run their own servers, and never will... I don’t think this can be emphasized enough — that is not what people want. People do not want to run their own servers. Even nerds do not want to run their own servers at this point. Even organizations building software full time do not want to run their own servers at this point. If there’s one thing I hope we’ve learned about the world, it’s that people do not want to run their

own servers. The companies that emerged offering to do that for you instead were successful, and the companies that iterated on new functionality based on what is possible with those networks were even more successful.

So follow the logic chain with me here...

DIDs are URIs that VCs use to reference human identities. Those URIs are going into the semantic web, regardless of what “processing” we do; the grafting is done by issuers, who are required to specify a valid @context for each credential when it’s created. And in the W3C’s view, URIs help us have conversations with *servers*. And if Moxie is right, individual humans will never run servers, but enterprises will. Doesn’t that strongly encourage conversations about humans to be had with... centralizing enterprises?

This all dovetails nicely with decentralized web nodes, and encrypted data vaults, and SOLID pods. But I think it’s bad news for actors.

You might say, “Well, that’s why we need governance frameworks and regulation. We can’t solve this problem with technology.”

I agree that we need those things. They help. I also agree that the solution can’t only be tech. But I think tech has to do *something* about this, as well. Just pushing the problem off to governance is asking for failure, if we can’t even see the risk in our own data-centric worldviews, and if the technology we build is so strongly oriented toward objectifying humans. Besides, governance alone can’t fix this, no matter how good it is. That’s because the semantic web, just like the non-semantic web, is *global* and *ungovernable* by design. And membership in the giant graph in the sky is one-way and irrevocable. A DID owner may have to sign a nonce when presenting a VC — but after a signed version of their data is revealed and leaked or recorded once, the gap in the semantic web is plugged, and nobody ever needs to deal with the human again. Servers can just talk to servers.

Before I close this meditation about the implications of semantic web for human data, I need to say something about two other words that are prominent in the VC community’s data cloud: “extensibility” and “interoperability.” The defense of @context is partly justified with the claim that we need it to guarantee these two magic words.

I don’t agree. HTML solved useful problems and saw strong adoption long before browsers got serious about viewing documents through the lens of XHTML. JSON solves useful problems every day without formal namespaces. Markdown can’t even agree on half the syntactic details, and it’s still powering huge chunks of the web, including the very websites where W3C people argue about how important formal structure is for interoperability and extensibility.

What if we emphasized extensible and interoperable actions[20] instead, and obsessed less about standard data for the time being? I’m trying to learn French right now, and I’m told this is how you learn a language: say stuff, even if it’s not quite right, and let experience fix the data.

I’m not claiming that we can/should ignore the VC data model. I’m claiming that the existence of a field named “issuer” communicates 99% of the value needed for extensibility and interoperability, and a mechanism that turns it into <https://www.w3.org/2018/credentials#issuer> is and should be an uninteresting 1%. The VC spec includes some good common sense that agrees with me, I think:

When two software systems need to exchange data, they need to use terminology

that both systems understand. As an analogy, consider how two people communicate. Both people must use the same language and the words they use must mean the same thing to each other. This might be referred to as the context of a conversation.

When a prover is proving something with a VC, the context for the conversation is VCs, whether or not @context is present. Both parties understand this and expect to see an “issuer” field. Getting more specific, if a verifier is challenging for credentials in the “Digital Diploma 2.0” schema, then the verifier already has a natural context they can use to decide what the “GPA” field means; they don’t need @context to figure it out.

It’s only when the credential data is divorced from the natural context of VC issuance and proving interactions, and moved into the global semantic web, that processors truly need a @context. That is the very context shift that objectifies people. And that is why @context should be optional: so people who are concerned about this issue can keep VC data in natural VC contexts only.

VII. Recommendations

So, what do I think we should do about this?

1. VCs about humans and VCs about other topics should be different. They can use the same data model, except... Using @context should be optional, at least when a VC is about humans.
2. Let’s stop imagining that the pushback on the @context requirement is reducible to developer prejudice or data processing burden.
3. Let’s fix our tag cloud by putting data in its proper role — prominent but not as important as actors.
4. Let’s stop saying that governance can fix antipatterns that technology enables and reinforces. Technology should first do no harm, then add governance.
5. Let’s emphasize actions more, and relax the rhetoric and the worries about data-oriented extensibility and interop. The biggest help for extensibility and interop will be starting to accept loosely correct data, not demanding that everybody make data that perfectly conforms to a committee’s ideal recipe.
6. Let’s make VC 2.0 simpler than 1.x — not by tightening the constraints, but by pushing more of the semantic web stuff out of scope.

One of Manu’s final comments in Issue 947 says, “all alternatives proposed to date fail to deliver on key features of the specification AND harm interoperability by multiplying the necessary (optional) code paths that currently exist.” About the first half of this statement, I say he is partly correct: they do fail to deliver. At least, my proposal here does. But I’m saying that the data processing requirements he’s pointing to (basically, everything that guarantees that VC data must fit into a big graph in the sky) are not good requirements. So let’s turn them into options. Choosing not to implement a feature that lacks strong consensus isn’t failure. :-)

About the second half of Manu’s sentence, I feel some sympathy. Implementing, testing, and securing optional features is expensive — and it’s no fun to do that, only to have others announce they won’t do the same. But the best way to eliminate those code paths is to omit a feature entirely. Manu is comparing the cost of maintaining an optional feature to the cost of making it required; if my counter-proposal is accepted, we’d then be comparing the cost of maintaining an optional feature to the cost of not having it at all, which feels more appropriate to me.

Interoperability isn't measured in code paths. It's predicted by good will/harmony in the tech community that implements, and ultimately measured by people — not processors on the semantic web — being able to get work done.

Doing what I'm proposing here wouldn't make VCs incompatible with the semantic web; it would just make them independent from it, with bridging work falling on the shoulders of those who passionately believe in the LD value proposition and passionately disbelieve my argument. The rest of us can focus elsewhere but we can all get along.

Endnotes

[1] In the github repo for the W3C VC spec, see Issue 947 about making @context optional, Issue 948 about limiting JSON-LD optionality, and Issue 957 about VC subjects. For a previous round of debate, see Issue 184 in DID Core.

[2] The father was discussing issues of choice and personal responsibility in the Garden of Eden story that is familiar to several of the world's great religions. See <https://bit.ly/3FnYFhU>.

[3] <https://bit.ly/3FrrLNI>

[4] <https://en.dialektika.org/philosophy/emmanuel-levinas-other-responsibility/>

[5] <https://bit.ly/3WqL9jI>

[6] See this doctor's meditation about the benefits of appreciating otherness: <https://doi.org/10.7202/1058265ar>. And consider G. K. Chesterton's observation: "How much larger your life would be if your self could become smaller in it; if you could really look at other men with common curiosity and pleasure; if you could see them walking as they are in their sunny selfishness and their virile indifference! You would begin to be interested in them, because they are not interested in you. You would break out of this tiny and tawdry theatre in which your own little plot is always played, and you would find yourself under a freer sky, in a street full of splendid strangers." (In *Orthodoxy*, 1959, p. 20–21).

[7] <https://doi.org/10.1037/tra0000452>

[8] https://en.wikipedia.org/wiki/Propaganda_in_World_War_I

[9] <https://doi.org/10.1177/0011000015591287>, <https://bit.ly/3DIWAf6>

[10] See, for example, <https://doi.org/10.1080/1068316X.2016.1269902>, <https://bit.ly/3DIFw8V>, <https://doi.org/10.1177/104398629601200204>, and <https://www.genderit.org/articles/trafficking-women-female-objectification>.

[11] <https://doi.org/10.2139/ssrn.1032263>, <https://nyti.ms/3znLg5p>

[12] <https://doi.org/10.1002/hpm.3008>, <https://bit.ly/3gQiu7g>

[13] <https://techhq.com/2018/04/facebook-if-something-is-free-you-are-the-product/>, <https://cnn.it/3fkGXRM>

[14] <https://plato.stanford.edu/entries/turing-test/>

[15] <https://dataprot.net/statistics/spam-statistics/>

[16] <https://codecraft.co/a-grumble-about-buckets>

[17] https://www.viktorfrankl.org/quote_stimulus.html

[18] <https://plato.stanford.edu/entries/agency/>

[19] Claims in VCs consist of data fields for subject, predicate, and object (Jane’s DID, birthdate, 25 Oct). These map directly to subject, predicate, and object in RDF triples. The subject is (typically) a DID, which is a URI. The predicate must be a URI. The object of the triple can be a literal or another URI. In their natural form, VD data structures tend to have field names that are more human-friendly than URLs (e.g., the VC has a field named “birthDate”). By adding an LD @context, we explain to the semantic web how human-friendly values convert into URIs (e.g., “birthDate” becomes <https://schema.org/birthDate>). This guarantees that they are unambiguous and resolvable.

[20] This is the focus of DIDComm. There are philosophical reasons why I think DIDComm’s approach to standardized actions is better than the alternatives, and they have to do with objectification again. But that’s a topic for another essay. Here, I’m claiming that any focus on protocols, no matter how they’re architected, will pay off more than a focus on nailing the