An Introduction to

Data Verification Platform

The executive summary of a new model for sharing verified data and quality signals

Stephen Wilson
George Peabody

"Just give us the facts."

Version 1.0-US
May 25, 2023

LOCKSTEP

Executive Summary

In this paper, we propose a general unified model for verifying data which uses the latest techniques for sharing verifiable credentials in a general-purpose network business model.

Open data, open banking, the rights of access to publicly-funded research, and so many other plans to "unleash the power of data" are being promoted across government, business, and social institutions. But what do we know about this data? For starters, how do we know any data is legitimate? And how can we know the important finer-grained properties such as the jurisdiction of origin, the algorithms used in processing, and whether there's consent to share?

There is clearly a need for all organizations to be able to verify the data they're relying on.

To achieve these lofty data-sharing goals, we need an environment which is both orderly and scalable. The users of data need to be reassured about its quality, its provenance, the permissions for its use, the processes that created it, and more.

Developments in the digital identity industry are instructive. With ever more emphasis on provenance, authority, fidelity, privacy, and agency, the parties to a transaction are focusing on credentials, affiliations, and other attributes—in other words, the metadata surrounding traditional "identity" data such as date of birth or national ID numbers.

In both the FIDO Alliance and in verifiable credentials, "identity" is less prominent, or even absent altogether. This is progress.

The data structures and signatures which are already being used for verifiable credentials can be broadened. A verifiable credential is, in fact, an attestation by a respected source about a fact about a subject—and that could be *any* fact. Indeed, it could be an attestation to any facts about *anything*, including non-human subjects, internet of things (IoT) devices, and data more generally.

But that raises another question: How do we scale up the acceptance of verifiable credentials and data when the entities who rely on the data are distant from the data's origins, whether that's geographically or legally? How do they know they can trust the entity making these attestations?

Data verification requires more than just technology. It also needs an *infostructure* that includes rules and scalable processes for distributing meaningful facts. We have researched and designed an infostructure to bring together the users and originators of data under a

uniform set of platform rules so that they can interoperate without needing to negotiate bilateral legal arrangements.

We set out here a *data verification platform* (DVP) which provides the operating principles and core functions needed for trustworthy data and credential sharing.

The platform intermediates between the sources of verifiable facts about ***data subjects***—the ***data origins***—and the ***risk owners*** who use those facts, via a new type of specialist business we call a ***data distributor***.

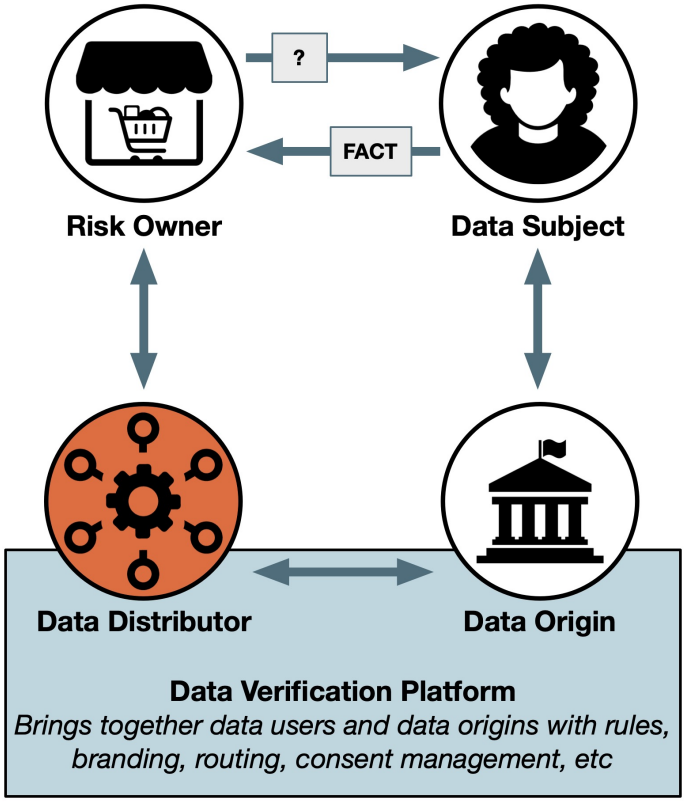


Figure 1: High-level data verification platform architecture.

Our proposed model deploys verifiable credential tools in what economists call a *two-sided market*, where all the parties' risks and roles are aligned economically and standardized contractually—characteristics lacking in existing systems of federated identity.

Our DVP provides a common foundational ruleset, regulatory posture, legal arrangements, certification framework, secure message routing, uniform UX requirements, and trust mark branding. It ensures that when a first party requires data about a second party to carry out some transaction, that data can be obtained from reliable sources, supported by a range of verified quality signals which are aligned with the risk-management needs of the transaction risk owner.

The DVP model fosters the fine-grained definition of transaction data, data sources, and verification metadata. All this information will be available in the DVP ecosystem to the designers of transaction systems so they may build in comprehensive real-time data verification.

We deliberately use new terms to clarify the essential functions and roles of each party using DVP-mediated data.

The DVP supports risk owners in obtaining verified core data about subjects—in plain language, *facts*—plus metadata about each fact's reliability, at scale, from diverse participating authoritative data origins using standard data verification protocols.

To manage specific risks, the first party always needs to know some data about the second party, and that data usually has a defined or preferred origin.

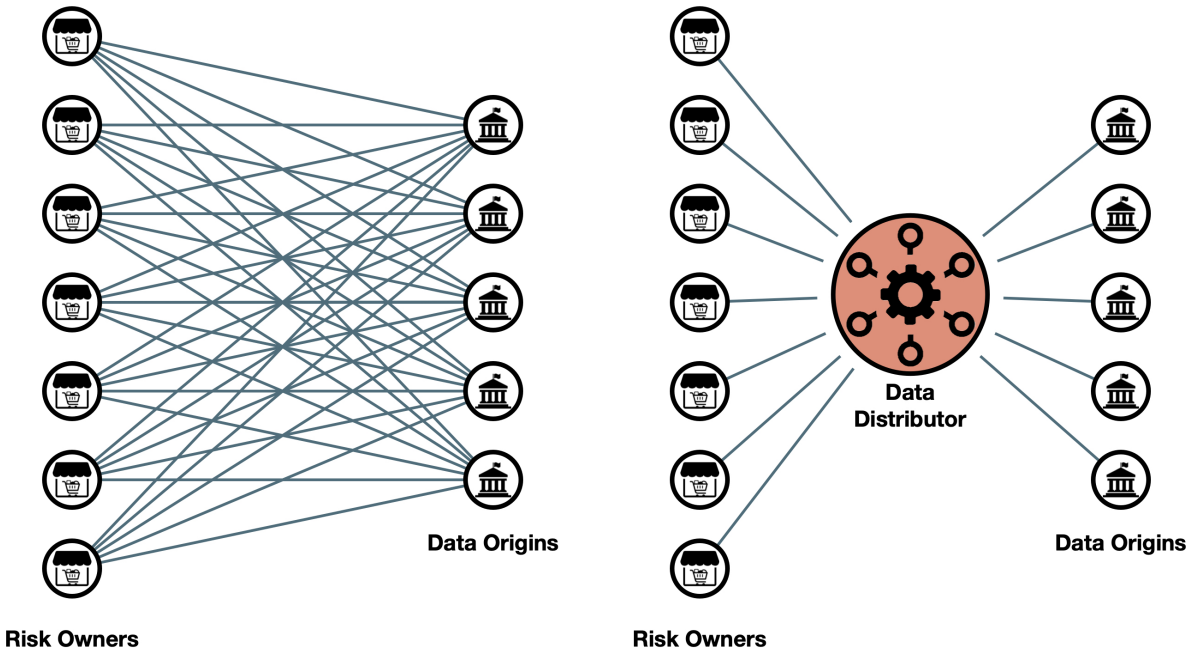


Figure 2: The data distributor simplifies the risk owners' legal and technical arrangements. On the left, without a data distributor, every risk owner must have a relationship with every data origin, a total of m×n links. On the right, the data distributor provides consistent arrangements for each data origin it represents, reducing the total number of links to m+n.

The new players, the *data distributors,* make these facts more digestible and more easily discoverable across the DVP ecosystem by onboarding the risk owners with a set of consistent legal agreements and technology with which to connect to the DVP network.

Risk owners do not need to make their own arrangements to accept far-flung data origins. Data origins, often government agencies, can remain complete strangers to risk owners and

still be trusted. Risk owners don't even need local legislation to legitimize the use of foreign data.

The DVP is about delivering the facts, direct from the source, including the metadata that tells the story of that fact, to the risk owner, with fine-grained quality signals to use for crisper risk-based transaction decisions.

The DVP does not alter any fact, yet it makes the data better. The DVP makes a rich array of quality signals available to any party that relies on facts, so that the facts are more reliable.

One major beneficial side-effect will be that businesses can cut down the amount of ancillary data they collect about the subject, because the core data they really need to know will be so much better.

The risk owner's application can instead be designed around specific data and metadata available in real time from associated data distributors.

Building the DVP will be a heavy lift, but comparable efforts have succeeded and indeed thrived in time.

By way of comparison, the payment card system has enabled a globally accepted payment and user experience through its founding principles, architecture, standards, business model, and contractual consistency. As one of the first true two-sided markets for digital services, the card system has delivered enormous value for merchants, consumers, and financial institutions, and fostered countless fintech businesses.

The FIDO Alliance is a fine contemporary example of collaboration on global security standards by competing risk owners, including banks, telcos, insurers, e-commerce and mobile platforms.

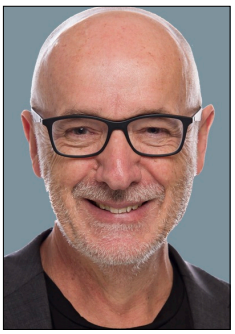
For both FIDO and the card schemes, the foundations for global scalability are their common principles, transparency, and respect for the needs of all participants.

The global card system has produced a secure yet simple to use system of Click to Pay, already in use in many countries. Our DVP model can similarly produce a secure yet simple to use system of "Click to Prove", so that any attested fact about someone can be held safe in a mobile wallet and presented by them, in-app, to a counterparty, easily, privately, and securely.

Building a large-scale data platform can be done again. Indeed, a modern general-purpose platform must be created in the interests of properly governed, economical, and uniformly experienced data-sharing worldwide.

About the authors

Stephen Wilson



Steve is a researcher, innovator and analyst in data protection. He has been a lead digital identity adviser to the governments of Australia, Hong Kong, Indonesia, Kazakhstan, Macau, New Zealand, and Singapore, and has been awarded 10 patents.

In 2018, he was **described by digital ethnographer Tricia Wang** as "one of the most original thinkers in digital identity in the world today".

Starting in public key infrastructure in 1995, Steve saw the potential for this technology in digital credentials. In 2004, he was awarded his first patent for anonymously verifiable attributes.

In 2011, he discovered an ecological explanation for the diversity of digital identities as ensembles of attributes, which in turn convinced him that all digital identity boils down to data.

George Peabody

George is a 20-year veteran of the payments industry. With 30 years in IT-based entrepreneurship and product management, he has expertise in payments strategy and market development.

His interests range across business and technology areas including mobile and point of sale payments acceptance, online and offline data security, and data verification.

Before joining Lockstep, George was partner at payments industry consultancy Glenbrook Partners, and led telecommunications research at Aberdeen Group. He co-founded payment identity firm Payment Pathways and a regional ISP. George produced and continues to co-host *Payments on Fire*®, the top-rated payments industry podcast.

In 2013, George and Steve published *Fractional Identity: An Alternative to NSTIC and Federated Identity*, which made the case to shift focus from centralized identity providers to contestable attribute providers.