Securing the Semantic Supply Chain

Meaning as a Foundation of Trust Infrastructure

Executive Brief

Dr. Carsten Stöcker
Spherity GmbH
carsten.stoecker@spherity.com
ORCID 0009-0003-7417-8872

Version 1.0, 16 November 2026 doi:10.5281/zenodo.24293613



Except where otherwise noted, this work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). © 2026 Dr. Carsten Stöcker.

This brief accompanies the research article of the same title and its supplements S1–S5.

The answer in brief

Verifiable credentials are digitally signed statements: a company's role in the energy market, a person's age, the chemistry of a battery. When a signing key is reliably linked to the named issuer, a valid signature authenticates the content covered by that signing method. It does not by itself establish what that content means to the system reading it (the verifier).

The meaning comes from elsewhere. A credential names its terms; dictionaries, vocabularies, validation rules, rulebooks and code tables define them. These definitions are published by someone, change over time, and are copied, cached and replaced outside the signed bytes. Together they form a supply chain of meaning, the *semantic supply chain*. Each link can be swapped, rolled back to an old version, frozen, shown differently to different readers, or withdrawn without the readers noticing.

W3C Data Integrity, one of the World Wide Web Consortium (W3C) standards for securing verifiable credentials, already protects part of this chain: for credentials secured with it, each verifier must check the dictionaries (JSON-LD contexts) it uses against approved copies, or apply equivalent protections. That protects one verifier at one moment. Who approves a meaning, how an update or a withdrawal reaches every verifier, and how a decision can later show which meaning it applied are left open.

> **A valid signature is not a valid meaning.** Meaning needs the discipline that software and cryptographic keys already have: authorised publishers, signed releases, freshness, withdrawal and audit.

Why it matters now

Credentials are moving from pilots into legal and economic infrastructure. The dates below are as of 14–15 November 2026.

- **Battery passports.** From 18 February 2027, light-means-of-transport batteries, industrial batteries above 2 kWh and electric-vehicle batteries placed on the EU market or put into service carry a battery passport (Regulation (EU) 2023/1542, Article 77). Parts of it are restricted to parties with a legitimate interest.
- **Product passports.** The Ecodesign Regulation (EU) 2024/1781 sets the framework; requirements follow in delegated acts. On 30 October 2026, the Commission's product-passport registry published its dictionaries, vocabularies and validation shapes.
- **Wallets for people and businesses.** Under Implementing Regulation (EU) 2025/1569 of the EU electronic-identification rules (eIDAS), whose Articles 6–9 apply from 19 August 2026, the Commission must publish a sealed catalogue of attestation schemes; a request to add or change a scheme must state its status and version. The European Business Wallet proposal, COM(2025) 838, is in negotiation between Council and Parliament.
- **Business-partner records.** WE BUILD, a large-scale pilot of the EU Digital Identity Wallet for business and payment use cases, pilots "know your business partner" checks on customers and suppliers. Its semantics group has published a vocabulary for business attestations whose terms include legal representatives and their powers, ownership and control, and beneficial owners. In our paper-only analysis, signed records using this vocabulary could support processing without re-keying for onboarding, business-partner master data, anti-money-laundering checks and export controls. Removing manual handling could also remove an opportunity to notice changed definitions; we have not tested or deployed this use case.
- **Customs.** 4.6 billion low-value e-commerce consignments entered the EU in 2024, almost twice as many as in 2023. A white paper we co-authored with TU Delft and the PARSEC, CIRPASS-2 and BORDERLINK projects notes that product-passport data helps customs only if it is verifiable and machine-readable "with semantic meaning".
- **AI.** AI agents increasingly act on credentials and on linked sets of signed evidence (evidence graphs). An agent applies the meaning it is given. For high-risk AI systems, the AI Act's record-keeping duty (Article 12) applies, after the 2026 amendment of the Act, from 2 December 2027 or 2 August 2028, depending on the category.
- **Long-lived labels.** New W3C drafts compress credentials and put them into barcodes such as QR codes. On products, such labels can stay in use for 10 to 20 years.

A wrong meaning can then become a wrong access decision, a wrong onboarding, anti-money-laundering or customs decision, or a wrong action by an AI agent, with a valid signature on it.

What goes wrong: three short stories

1. The grid operator that became a transmission operator. An energy-sector credential states that a grid operator is a *distribution* system operator with access to *public*

battery-passport data. A battery manufacturer's portal checks the credential and grants access. The credential names its dictionary by a web address. If the document at that address is changed, for instance by a compromised server, a cache or a careless update, the same signed bytes now say *transmission* system operator and *restricted* data. The signature check still says "valid". Nobody broke a key or revoked anything, yet authority changed (Figure 1).

Status: demonstrated in our laboratory on production software, with two independent implementations, in one use case. It requires a signing method that does not cover the dictionary's content and a verifier that does not check its own approved copy. Deployments that keep approved copies of their dictionaries and check against them are not shown to be affected, and we have not observed it in a deployed system.

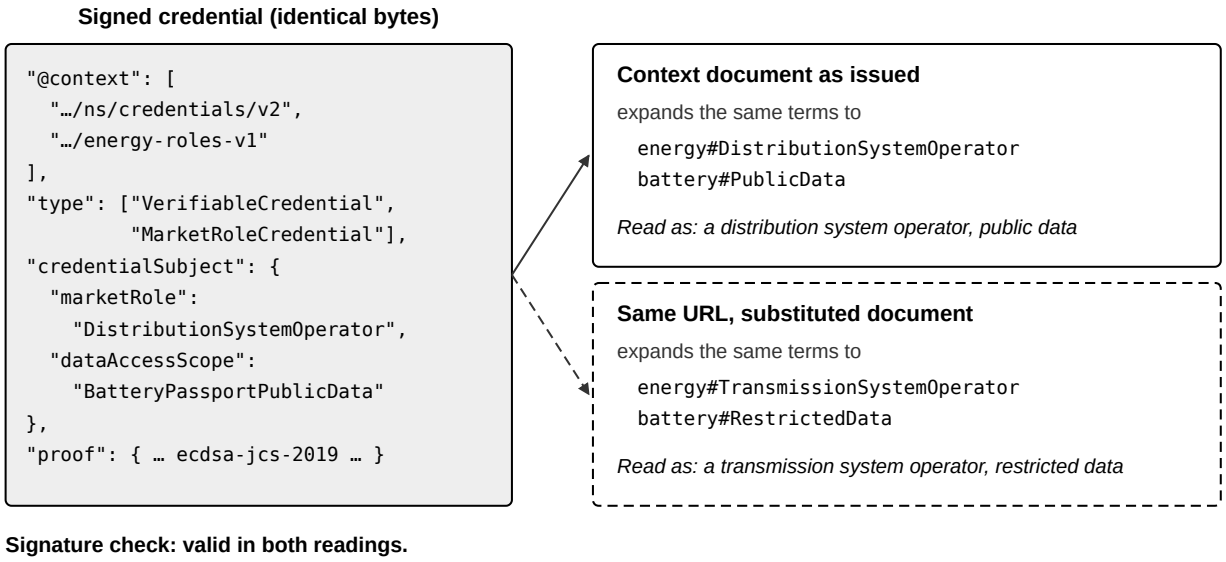


Figure 1: Same signature, two meanings. The credential on the left is signed once. Read with the dictionary its issuer intended, it grants a distribution system operator access to public data. Read with a different document placed at the same web address, the identical signed bytes describe a transmission system operator with access to restricted data. The signature is valid in both readings.

2. The battery label read in 2040. A battery's label will be read for 15 years or more, by readers running other software and other vocabulary versions. In our tests, a routine update of a vocabulary made a signed compact label unreadable, with an error message that does not name the cause. A label whose signature covers its compressed form still verified, but decoded into a different statement, with its access scope now *restricted*.

Status: demonstrated in our tests, on one CBOR-LD library with a synthetic label. That a label read in 2040 no longer means what it meant in 2026 is our projection from this result, not an observation. The binding we propose rejects such a label before decoding, and an archive restores the original meaning; this is tested in software, but no production archive exists.

3. The AI agent that acts on a changed definition. An AI agent acts for a company under a mandate and reads evidence (test results, certificates, roles) before each action. If someone upstream changes a *definition* rather than a data value, the agent applies the new meaning and does not notice.

Status: illustrative. This "definition-level poisoning" is a conjectured attack; none has been measured. For AI systems fed from large knowledge graphs, the article names the open problems and does not yet propose a solution.

This is not a format question

The problem is not specific to W3C credentials. In SD-JWT VC and ISO mdoc, the two other formats the article examines, the signature covers claim names whose meaning comes from documents defining a credential type (Type Metadata) and from rulebooks, the documents that specify each kind of attestation. An integrity hash can fix the bytes of such a document, but not its freshness, protection against rollback, its withdrawal, or which version a verifier actually applied.

Some attacks do not arise in these formats, because no dictionary is fetched during verification; the weak point moves to the rulebook and the verifier's configuration. Rulebooks are long documents, written per type of attestation and changed by people, and the same attribute (a birth date, a role, a battery chemistry) appears in many of them. Our mapping of which safeguards carry over to these formats is design reasoning; none has been specified or tested for them, and our further work concentrates on W3C credentials.

> **Rulebooks, vocabularies and code tables form a supply chain too.** They deserve the protection that software packages received after integrity checks of single packages proved insufficient.

What we found

We built an open test harness on production software libraries. Five results stand out. The first four cover the article's findings F1 to F5; the fifth comes from one of its worked examples (WE-5).

1. **A valid signature can carry a changed meaning** (the first story). The theory was argued before; we demonstrated it on production libraries.
2. **Software fails quietly.** A signed field that no dictionary defines stayed validly signed and was silently dropped from the machine-readable meaning; the W3C standard requires software that performs this processing to reject such a loss. In one JavaScript library, a shared cache meant that a verifier's own approved copy of a dictionary was never consulted on one cache path.
3. **An attacker without any key can make a verifier work before the signature is checked.** In our 2026 measurement on one machine, a 22 kB document cost a verifier that processes without a work limit more than 20 seconds, before any signature check. The tested rdf-canonize library, from version 4.0.0 onward, stops such input by default.
4. **That default, in turn, rejects some legitimate credentials.** Under the tested libraries' default limits, graph-based verification cannot process some lists with repeated values, such as 1, 2, 2, 3; legitimate credentials containing those lists therefore fail veri-

fication by that method under those settings. This is an availability problem wherever such lists are legitimate. We tested one family of lists on two software stacks.

5. **Time limits make decisions depend on machine load.** With a time limit set at 1.5 times the typical duration on an idle machine, a legitimate credential passed 21 of 21 times on an idle machine and 0 of 21 times on a busy one. A limit on the number of operations accepted it 21 of 21 times in both cases. We tested one configuration.

How we know. The harness is open source (Apache-2.0) and archived with a DOI: version 0.4.0 passes 136 tests. A mathematical model states the core results as theorems with outline proofs. A program then tested the safeguards that keep state, such as the update and withdrawal rules, in 66 small cases; this covers the cases tested, not every size. The work went through seven rounds of adversarial review by AI models, the last by two reviewers in parallel, and a cross-check of the last round. Every finding was checked before it was accepted: code defects by running them, standards and legal statements at the primary source. These reviews are not peer review. The demonstrations in the first two stories and all timings come from the earlier harness version 0.3.0, on one machine and one software stack.

What must change

Software faced the same questions years ago and answered them with provenance, reproducible and signed builds, update frameworks, transparency logs and security advisories. The article proposes 30 safeguards and 91 requirements that apply that discipline to meaning. In plain terms:

- **Authorised publishers.** A trusted list of who may *publish meaning*, not only of who may issue credentials.
- **Meaning built and released like software.** Definitions built from one source and tested against a fixed set of examples. A change that the examples show to alter meaning cannot be published under an existing name, and the signing key is used only after the checks have passed.
- **Freshness, rollback protection and withdrawal.** Verifiers reject outdated update information, never accept an older version without authorisation, and treat a signed withdrawal as permanent. Releases are recorded in a transparency log that witnesses check.
- **Check first, then interpret.** Where the signing format permits it, authenticate before costly interpretation; otherwise enforce work bounds during the processing needed to verify the signature. Our proposed experimental signing design makes signature checking possible before costly interpretation. Bound processing work by operations, not by time.
- **Decisions that remember.** Each decision records fingerprints of the meaning and decision context it used, and preserves the inputs and context needed for deterministic replay by an auditor. Withdrawing a meaning stops the authority that depended on it.
- **Fixed evidence for compliance decisions.** For decisions such as customs release, the evidence a decision needs is fixed in advance, and the presenting party has no choice over evidence that affects the outcome. Whether that evidence is complete remains an

obligation on authoritative sources; no safeguard here proves that evidence is complete or true.

These safeguards establish authenticity, freshness and accountability of meaning. They cannot stop an authorised publisher from releasing harmful but authorised meaning; only review, testing and transparency limit that.

The 91 requirements are our proposal for a high-assurance profile of the W3C credential standard. A W3C Community Group for such a profile was proposed on 11 November 2026 and launched on 12 November 2026; no W3C group has adopted the requirements.

What to do, by role

Issuers and rulebook authors

- Treat every vocabulary, rulebook, Type Metadata document and code table you publish as a release: versioned, signed and logged, with a named owner.
- Never change the meaning behind an existing name. Publish a new version, and withdraw old ones explicitly with a signed notice.
- For labels and codes that stay on products for years, bind the decoding tables to the label and keep an archive of them.
- If your signature covers only the credential's text and not the dictionaries it points to, add a fingerprint of each dictionary to the credential. The standard already fixes the fingerprints of its own base dictionaries.

Verifiers and relying parties

- Load definitions only from approved copies whose fingerprints match the expected versions, and check every definition your software uses, including those embedded in documents.
- Where the signing format permits it, authenticate before costly interpretation; otherwise enforce work bounds during the processing needed to verify the signature. Limit processing by operations, not by time.
- Record which meaning and decision context each decision applied, and retain the inputs and context needed to replay it deterministically in an audit or dispute.
- Fail closed: treat stale, missing or withdrawn meaning as "unknown", never as "allowed".

Wallet, platform and library providers

- Make checking against approved definitions easy to switch on, and consider making it the default; none of the libraries tested in a recent independent study applies it by default.
- Do not share cached definitions across verifiers that trust different copies.
- Document that default work limits reject some legitimate credentials, and test for it.
- For devices that verify offline, fix the longest permitted offline period and protect against rollback to old definitions.

Standards bodies

- W3C (web standards): consider the requirements in the new Community Group for a high-assurance profile; add the risk of changed decoding tables to the security considerations of CBOR-LD, the draft for compact credentials; add processing cost before signature checks to the threat model of the credential standard.
- ETSI (European telecommunications standards): add publishers of meaning to trusted lists, and the status of meaning to scheme catalogues.
- IETF (internet standards): define how semantic artefacts are recorded in transparency logs.

Regulators, supervisors and those who define compliance data

- For each compliance decision (customs, product passports, business wallets), define the sufficient evidence and how its completeness is established, and leave the presenter no choice over evidence that affects the outcome.
- Consider who may publish the meaning that credentials rely on, not only who may issue credentials.
- Consider giving published semantic artefacts, such as those of the product-passport registry, freshness information and signed withdrawal notices in addition to their version information.
- When specifying record-keeping, consider requiring systems to record which definitions they applied, not only their inputs. This is our proposal, not the text of the AI Act.

All of these asks are our proposals, not positions of any of the bodies named.

What this does not solve

- **Truth.** The safeguards bind meaning, not its truth. A well-governed vocabulary can still describe the world wrongly.
- **Authorised mistakes.** An authorised publisher can still release harmful meaning.
- **Law.** Whether meanings are legally equivalent across jurisdictions is outside the scope.
- **Hardware trust.** Protected execution on devices is simulated, not demonstrated. Attributing a device's decisions to an accountable legal person is specified, not implemented, and this work does not demonstrate an authority issuing the required records.
- **Maturity.** Most of the 30 safeguards are design proposals. Some are checked in the formal model on small instances or tested as software in the harness; none has been deployed.
- **Other formats.** For SD-JWT VC, ISO mdoc and EU wallet rulebooks, the safeguards are mapped by design reasoning only.
- **Adoption.** The requirements are a proposal. None of the use cases has been deployed or trialled.

Read more

- **The article:** *Securing the Semantic Supply Chain: Meaning as a Foundation of Trust Infrastructure*, version 1.0, with five supplements: the safeguards in full (S1), the 91 requirements (S2), the formal model with proof sketches and model checking (S3), the prior-art ledger (S4) and the method and review record (S5). Every finding and number in this brief is drawn from the article; recommendations worded as "consider" are derived from the gaps it identifies.
- **The open harness:** version 0.4.0, Apache-2.0, doi:10.5281/zenodo.24281131.
- **Earlier work:** C. Stöcker, *Proofs over Statements in Verifiable Evidence Graphs for the European Business Wallet*, Spherity Research, 2026, https://spherity.github.io/spherity-research/Stoecker-2026-Proofs-over-Statements-v1.3.pdf; B. Rukanova et al., *Cross-Border E-commerce to Consumers, Digital Product Passports (DPP), and the Role of Authorities*, 2026, https://doi.org/10.5281/zenodo.19239072.
- **Business-partner records:** WE BUILD Consortium, WP4 Semantics group, *European Business Wallet Vocabulary* v0.1, 2 November 2026, https://w3id.org/ebwv; the article's use case UC6.
- **Planned next:** a high-assurance profile for W3C verifiable credentials in which every admitted format (JSON-LD for exchange, the compact CBOR-LD for long-lived labels, YAML-LD for authoring only) binds to one graph of statements (RDF); verifiable knowledge graphs for AI; and key management for AI agents.

Acknowledgements

The author thanks Ronald Koenig (Spherity), Mikael af Hällström (Finnish Tax Administration) and the entire WE BUILD WP4 Semantics group for their work in WE BUILD, including the European Business Wallet Vocabulary; Manu Sporny (Digital Bazaar) for his leadership in W3C on JSON-LD and the Verifiable Credentials Data Model, and for our personal exchanges; and all members of the W3C Verifiable Credentials Working Group. Those named have not reviewed this work, and the acknowledgement does not imply their endorsement. The views expressed, and any errors, are the author's own.