

The Bio-Digital Footprint (BDF) Protocol

Verifiable Records for Transformed Bioassets via Bio-Digital Twins and Tiered Attestation

Fair Food Data (FFD)

Version 1.0 · 24 September 2026

Abstract

This paper introduces the *Bio-Digital Footprint* (BDF) protocol, a framework for the verifiable provenance of transformed biological assets. BDF models each bioasset as a *Bio-Digital Twin* (BDT), a typed digital object whose identity is derived from its data through deterministic canonical serialisation and whose lifecycle is recorded as a signed append only chain. Independent attestations covering the agronomic, logistic, process, and recipe stages of the supply chain are anchored to the BDT as typed signed vectors. Industrial transformation consumes inputs via burn semantics; distribution instantiates derivative package twins that preserve their parent’s lineage.

The contribution of this paper is the *Tiered Attestation Model*, a structural decomposition of the evidence supply problem in food traceability into four progressively rigorous tiers, ordered by source of evidence, marginal cost, and probative value, and additive in their detection capabilities. We define detectors at each tier as typed functions on the BDT, prove the additivity of the detection classes, and describe how a deployment composes per tier verdicts into an aggregate verification verdict. External Attestations following the W3C Verifiable Credentials envelope extend the protocol to forensic laboratory evidence and to any other third party signer. Payloads compose with established open standards above beeson canonicalisation.

The protocol guarantees structural integrity of the audit trail under explicit network assumptions and supports composable verification verdicts at any chosen depth of evidence. It transparently delimits what it does not guarantee, including the truthfulness of physical measurements and protection against fully coordinated collusion. Applications include high integrity carbon credits for recovered food, verifiable provenance for premium markets, and ESG and Scope 3 reporting under CSRD and CBAM.



Contents

1	Introduction: The Cost of Trust	5
I	Foundations	6
2	The Bio-Digital Twin	6
2.1	Lifecycle Semantics	6
2.1.1	Genesis	7
2.1.2	Signed Ownership Chain	7
2.1.3	Transformation (Burn)	7
2.1.4	Distribution and Lineage	7
3	Attestation Vectors	8
3.1	Agronomic Vector: Biological Context at Origin	10
3.2	Logistic Vector: Spatio Temporal Context Between Origin and Plant	11
3.3	Process Vector: Industrial Telemetry of the Transformation	11
3.4	Recipe Vector: Operator Commitment to Process Parameters	11
3.5	Vector Independence and Cross Referencing	12
4	The Tiered Attestation Model	12
4.1	Tier 1: Cryptographic Floor	12
4.2	Tier 2: Telemetry Anomaly Detection	13
4.3	Tier 3: Finished Product Spectroscopy	13
4.4	Tier 4: Laboratory Forensic via External Attestations	13
4.5	Tier Comparison and Additivity	14
II	Operation	15
5	Per-Tier Detection Capabilities	16
5.1	Structural Verification at T_1	16
5.2	Statistical Anomaly Detection at T_2	17
5.3	Spectroscopic Classification at T_3	18
5.4	Bayesian Forensic Verification at T_4	18
5.5	Composition of Detectors and Verification Verdict	18
6	Architecture and Data Governance	19
6.1	Three Layer Architecture	19
6.2	Layered Data Placement	19
6.3	Encryption, Crypto Shredding, and Selective Disclosure	20
6.4	Retrospective Access Revocation via DREAM	20
6.5	Ownership versus Access	21
6.6	Regulatory Compliance	21
7	Adversarial Model and Per Tier Guarantees	22

7.1	Threat Landscape	22
7.2	Guarantees by Tier	22
7.3	Signer Independence and Reference Set Integrity	23
7.4	Residual Limitations	23
III	Application and Practice	23
8	Empirical Pilot Results	23
8.1	Pilot Configuration	24
8.2	Agronomic Profile	24
8.3	Logistic Vector and Mass Flow	24
8.4	Process Vector	25
8.5	Tier 2 Detector Demonstrations	25
8.6	Limitations and Ongoing Work	27
9	Applications	28
9.1	Verifiable Provenance for Premium Markets	28
9.2	Carbon Markets and dMRV for Recovered Food Carbon	28
9.3	ESG Reporting and Scope 3 Disclosure	29
9.4	Quality Assurance for Producers	30
9.5	Fraud Deterrence	30
9.6	Extensibility to Other Verticals and Evidence Types	30
10	Roadmap and Future Work	31
10.1	Implementation Roadmap	31
10.2	Hardening Pathways	31
10.3	Open Research and Extensions	32
11	Conclusion	33
A	Mathematical Formalisation	33
A.1	Canonical Object	33
A.2	Attestation Vector	33
A.3	Ownership Chain	34
A.4	Burn and Distribution Conditions	34
A.5	Provenance Graph	35
B	Feature Inventory	35
B.1	Agronomic Vector	35
B.2	Logistic Vector	36
B.3	Process Vector	36
B.4	Recipe Vector	36
	Document History	37

1 Introduction: The Cost of Trust

Global food supply chains suffer from a structural problem: the chain of evidence about a product degrades at every handoff. Certificates are PDFs, temperature logs may have been copied, laboratory reports are valid only until the lot they describe is gone. Consider a batch of freeze dried mango at a destination market, claimed to be grown in a specific region, harvested three days before processing, processed within forty eight hours, delivered cold. None of that can be verified today without trusting every intermediary. The fruit carries its history physically, in its cellular structure and its mass, but the fruit cannot speak. Its history is implicit, not legible.

The *Bio-Digital Twin* (BDT) closes this gap. It is a structured digital object that travels alongside the physical bioasset and accumulates, in cryptographically verifiable form, the same history the asset accumulates physically. This paper introduces the *Bio-Digital Footprint* (BDF) protocol as the framework that defines and operates this twin. BDF treats each bioasset as a typed digital object whose identity is derived from its data through deterministic canonical serialisation. Ownership evolves through a signed append only chain. Industrial transformation consumes input objects via burn semantics; distribution instantiates derivative package objects that preserve their parent’s lineage. The integrity of the audit trail is structural, not declarative.

The voluntary carbon market is the most documented case of the failure that motivates this work. When trust broke in 2023, the bulk of the market broke with it. The high integrity tier that survived, around 35 percent of credits, now commands roughly four times the price of paper credits, and the gap is widening. The market that will exist in 2030 is the one that produces evidence, not assertions. Premium markets defending denomination of origin claims, ESG and Scope 3 reporting under CSRD and CBAM, and fraud deterrence in commodity supply chains face the same erosion of paper based evidence.

The contribution of this paper is the *Tiered Attestation Model*, a structural decomposition of the evidence supply problem into four progressively rigorous tiers, organised by source of evidence, marginal cost, and probative value, and additive in their detection capabilities. Around it, BDF organises evidence into four *attestation vectors*, $\vec{A}$ Agronomic, $\vec{L}$ Logistic, $\vec{P}$ Process, $\vec{R}$ Recipe, each signed by an independent attesting party at a distinct stage of the lifecycle. T_1 provides cryptographic integrity by construction. T_2 runs statistical anomaly detection over the signed telemetry. T_3 introduces spectroscopic classification of the finished product. T_4 anchors forensic laboratory evidence via External Attestations. A deployment selects the depth of evidence its use case requires; the protocol is identical at every depth. Payloads compose with established open standards (SenML, GS1 EPCIS 2.0, FoodOn, SAREF4AGRI, PROV-O, W3C Verifiable Credentials) above beeson canonicalisation [Tron et al., 2026].

The paper is organised in three parts. *Foundations* formalises the typed twin and its transformation and distribution semantics (Section 2), the four attestation vectors and the standards stack (Section 3), and the Tiered Attestation Model with the External Attestation mechanism and the additivity of detection classes (Section 4). *Operation* develops the detectors at each tier (Section 5), the three layer architecture and the data governance framework that reconciles audit integrity with the right to erasure (Section 6), and the adversarial model with per tier guarantees (Section 7). *Application and Practice* reports empirical pilot results (Section 8), surveys the deployment applications enabled by the protocol (Section 9), and outlines the implementation roadmap, hardening pathways, and open research directions (Section 10). The paper closes with conclusions (Section 11). Three companion papers extend this

corpus: a formal security analysis with full threat model and reductions [Fair Food Data, 2026b], the methodology for recovered food carbon credits aligned with major registries [Fair Food Data, 2026a], and an implementation reference for the underlying open standards stack [Fair Food Data, 2026c].

The protocol guarantees structural integrity of the audit trail under explicit network assumptions and supports composable verification verdicts at any chosen depth. It does not guarantee the truthfulness of physical measurements nor protect against fully coordinated collusion across all data sources. What it offers is reproducibility: any verifier with access to the public record and the relevant reference materials can independently retrace the chain of evidence and arrive at the same verdict.

Part I

Foundations

2 The Bio-Digital Twin

BDF is the protocol; the *Bio-Digital Twin* (BDT) is its primitive object. The central innovation is modelling each biological asset as a canonical typed digital object with a unique tamper proof identity. While the worked examples in this paper concern food lots carried in containers, the BDT abstraction applies to any biological entity that can be identified, instrumented, and bound to a digital identity: agricultural batches, livestock, aquaculture, microbial cultures, biotech samples, or any organic asset whose history is worth preserving.

A BDT is described in beeson [Tron et al., 2026], an object notation expressing structured data as a typed attribute value matrix. Beeson defines a deterministic canonical serialisation, which is hashed to produce the unique identifier:

$$H_\alpha = \mathcal{H}_{\text{BMT}}^+(\text{Serialise}(\alpha)), \quad \text{such that } \alpha \text{ is a fully specified object of type } \tau.$$

Any modification of the data, however small, yields a different hash, so tampering is structurally detectable. The schema is embedded in the representation, so identical data always yields the same identifier regardless of who computes it.

Ownership of the BDT is represented through an append only chain of signatures rather than by contract enforced runtime validation. This separates structural integrity from measurement validation and allows reproducible verification without embedding business logic into smart contracts. It also allows data to be verifiable without leaking private information (Section 6). At each stage of the supply chain the collected data is packaged into a typed BDT and stored on a content addressed network, where it functions as a Real World Asset record, a node in the transformation history, a tradeable digital token, and a substrate for selective disclosure.

2.1 Lifecycle Semantics

The BDT has four operations on its lifecycle: genesis (mint), transfer (signed ownership), transformation (burn), and distribution (fan-out with conservation).

2.1.1 Genesis

A BDT comes into existence through a single atomic event we call *genesis*: the first signed attestation that instantiates the typed object’s canonical identity. Before genesis, there is sensor data without identity; after genesis, there is a BDT with permanent canonical hash H_O and a public ownership record. The genesis attestation is typically the Agronomic Vector $\vec{A}$ signed at harvest, though the protocol does not constrain which vector plays this role. The actor performing the mint becomes the initial owner A_0 . Genesis is conceptually equivalent to the mint operation in token systems, converting an unbounded stream of raw measurements into a bounded, identifiable, transferable object.

The genesis attestation declares the type of its minter (`human_operator`, `automated_iot`, `proplet_attestation_certified_lab`) so downstream verifiers can apply the trust profile appropriate to the source. This is auditable and attributable; a verifier reading a BDT can distinguish a hand-typed origin weight from a hardware-attested measurement and weigh them differently. Section 10.2 discusses how Proplet compliant sensors shift the genesis principal from human operator to tamper resistant device.

2.1.2 Signed Ownership Chain

Every transfer of a BDT, from farmer to transporter, transporter to processor, processor to buyer, is recorded as a signed statement:

$$\sigma_i = \text{Sign}_{A_i}(H_O \parallel A_{i+1})$$

where A_i is the current owner and A_{i+1} the new owner. The signature can only be produced by A_i ’s private key. Transfers are published under a deterministic Single Owner Chunk identifier $\text{SOC}_O = \mathcal{H}(H_O \parallel \text{“transfer”})$. Under replicated content addressed storage [Trón, 2024], anyone can reconstruct the complete ownership history; conflicting transfers from the same signer are publicly detectable.

2.1.3 Transformation (Burn)

Industrial transformation is represented as a terminal ownership transfer to a designated transformation address, marking the input as consumed:

$$\text{Sign}_A(H_O \parallel \text{TransformAddr})$$

Once burned, the input cannot be reused. The processing plant signs a `TransformationEvent` that consumes one or more burned inputs (fan-in) and instantiates a new output BDT whose identity is derived from the joint payload of inputs, recipe, and process telemetry (Figure 1).

2.1.4 Distribution and Lineage

Most supply chains do not end at the production of a single processed lot. A batch is typically distributed as multiple retail or industrial packages, often across shifts or days rather than in a single atomic packaging event. The protocol represents this fan-out with a hybrid pattern combining NFT identity with UTXO style lifecycle.

Each BDT is a non fungible object with a unique identity; it cannot be subdivided in the NFT sense. A *distribution event* consumes the parent BDT entirely (analogous to a burn) and instantiates a set

$$\{O_{p,1}, \dots, O_{p,M}, R\}$$

of M package BDTs and an optional *remainder twin* R . Each output is a unique BDT with its own identity and ownership chain, and a cryptographic reference back to the parent. The remainder R is itself a finalised BDT eligible for its own distribution event later, producing a chain of distributions over time until the lineage is exhausted. The protocol enforces a quantitative conservation invariant on the metadata quantity declared in the schema:

$$\sum_{k=1}^M q(O_{p,k}) + q(R) = q(O_p),$$

where $q(\cdot)$ is the declared quantity (mass, volume, count) and the remainder term is zero when no remainder is produced. The invariant operates on metadata, not on token fungibility: each output BDT remains a unique non fungible object. The pattern is structurally analogous to UTXO change semantics in Bitcoin and is preceded in compositional NFT designs, ERC-1155 contracts, and Bitcoin Ordinals. Each output anchors its own forked subfeed of the parent’s content addressed record [Trón, 2024], preserving the full vector lineage (Figure 2).

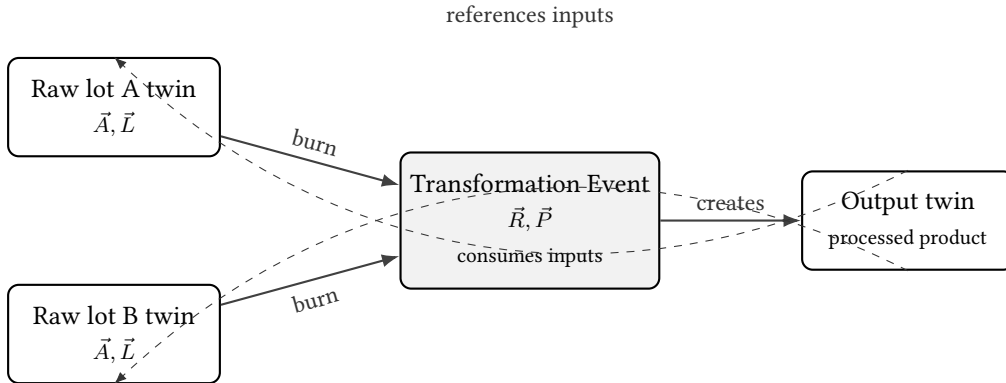


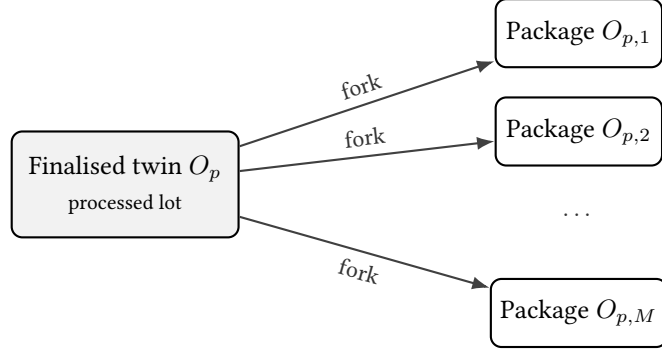
Figure 1: Provenance graph and burn semantics. Two input twins are consumed via burn into a Transformation-Event. The event instantiates a new output twin cryptographically linked to its predecessors. Dashed arrows indicate the references the output carries back to the burned inputs.

3 Attestation Vectors

The protocol binds together evidence drawn from independent sources, each with its own attesting party, its own measurement standards, and its own juridical semantics. We formalize this evidence as typed *attestation vectors*, structured records of measurements signed by a specific attesting party at a specific stage of the asset’s lifecycle.

Definition 1 (Attestation Vector). *An attestation vector is a tuple*

$$\vec{V} = (a, \tau, \vec{x}, t, \sigma)$$



Each package twin carries its own identity and ownership chain.
Quantities sum to $q(O_p)$ via the conservation invariant.

Figure 2: Distribution and lineage semantics. After finalisation, a parent twin O_p instantiates a set of M package twins (and optionally a remainder twin) through a single atomic distribution event. Each package is transferable, claimable, and verifiable independently of its siblings.

where a identifies the attesting party, τ is the schema type to which the vector conforms, $\vec{x}$ is the structured payload of measurements that satisfies τ , t is the temporal anchor of the attestation, and

$$\sigma = \text{Sign}_a(\mathcal{H}_{\text{BMT}}(\vec{x}, \tau, t))$$

is the signature of a over the canonical hash of the payload, schema, and timestamp under the binary Merkle tree hash $\mathcal{H}_{\text{BMT}}$ associated with the beeson canonical serialisation [Tron et al., 2026].

The four core vectors we define cover the physical lifecycle of a transformed bioasset from field to finished product. We denote them $\vec{A}$ (Agronomic), $\vec{L}$ (Logistic), $\vec{P}$ (Process) and $\vec{R}$ (Recipe), with corresponding schema types $\tau_A, \tau_L, \tau_P, \tau_R$ and attesting parties a_A, a_L, a_P, a_R . The four core vectors are not exhaustive: the protocol admits extension by additional vectors and by external attestations from third parties, formalised in Section 4.

The signature attached to each vector is not a generic stamp. It carries the identity of a specific attesting party, who commits to a specific class of evidence under a specific accountability regime. Cross referencing evidence across vectors is forensically powerful precisely because the parties, the instruments, and the failure modes are independent. Fabricating a coherent narrative across all of them simultaneously requires coordinated capture of the entire evidence supply chain, an adversarial threshold substantially higher than the forging of a single certificate.

Generality of the model. Although the worked examples in this paper use freeze drying as the transformation step, the four vectors are not specific to lyophilisation. The same structure applies to any industrial transformation of a biological asset where field, transport, process, and recipe are distinct, identifiable phases with their own data sources. Roasting of coffee and cocoa, conching of chocolate, fermentation of wine, beer, and dairy products, extrusion of cereals and feed, pasteurisation and ultra high temperature processing of liquids, and tunnel drying of fruits and vegetables all share the same canonical decomposition. We treat freeze drying as the reference transformation throughout the paper because it admits the cleanest exposition of the four vector model and because it sits in a methodological white space for the carbon market discussed in Section 9. The protocol is transformation agnostic by design.

Standards stack. The protocol is agnostic to the payload format of $\vec{x}$ inside any attestation vector. The reference deployment we describe adheres to a layered stack of open standards above beeson canonicalisation. Raw sensor readings follow SenML (IETF RFC 8428). Supply chain events follow GS1 EPCIS 2.0 in JSON-LD serialisation. Semantic vocabularies are sourced from the GS1 Web Vocabulary, FoodOn for food taxonomies, SAREF4AGRI (ETSI) for agricultural quantities, and PROV-O (W3C) for provenance genealogy. Signatures follow W3C Verifiable Credentials 2.0 with Data Integrity Proof. Beeson provides the deterministic canonical serialisation required for content addressed identification above all these formats. The cryptographic guarantees of the protocol depend on beeson and on the underlying signature scheme. The semantic interoperability of the payloads depends on the standards layered above. Adopting these standards is a deployment choice rather than a protocol requirement, but it is the choice we recommend in the implementation roadmap of Section 10.

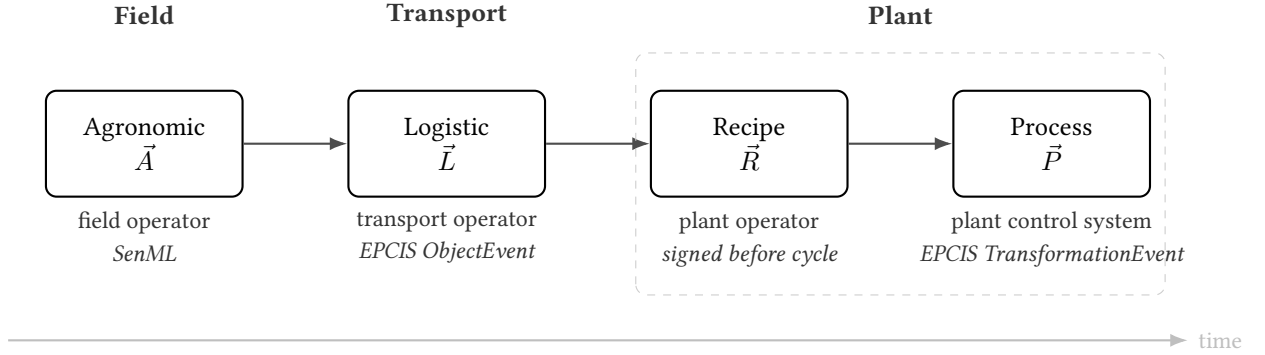


Figure 3: The four core attestation vectors of the BDF protocol. Each vector is signed by an independent attesting party at a distinct stage of the lot’s lifecycle. The dashed region groups the two plant side vectors. The Recipe Vector $\vec{R}$ is committed before the cycle begins; the Process Vector $\vec{P}$ is signed automatically by the control system after the cycle ends. Cross referencing $\vec{R}$ with $\vec{P}$, and $\vec{L}$ with $\vec{P}$, is the canonical pattern of execution fidelity exploited by the T_2 detectors (Section 5.2).

3.1 Agronomic Vector: Biological Context at Origin

Definition 2 (Agronomic Vector). *The Agronomic Vector $\vec{A}$ is an attestation vector of schema type τ_A , attested by a_A , the operator of the field IoT infrastructure, whose payload $\vec{x}_A$ aggregates environmental and edaphic measurements taken on a defined agricultural plot during a defined production cycle. Individual sensor readings within $\vec{x}_A$ follow the SenML format (IETF RFC 8428).*

A typical sensor inventory covers soil moisture and temperature at multiple depths, soil bulk conductivity, matric potential, ambient temperature and humidity, atmospheric pressure, photosynthetically active radiation, rainfall, and wind. Some deployments add in line water quality probes, multispectral imagery, and at line portable instruments such as pXRF or LIBS for elemental screening. The full feature inventory is enumerated in Appendix B. The signature σ_A commits to the integrity of the readings as they leave the gateway; sensor calibration and geographical placement are addressed by the trust assumptions of T_1 and the consistency analysis of T_2 , with T_4 available when forensic resolution is required.

3.2 Logistic Vector: Spatio Temporal Context Between Origin and Plant

Definition 3 (Logistic Vector). *The Logistic Vector $\vec{L}$ is an attestation vector of schema type τ_L , attested by a_L , the transport operator in custody of the lot, whose payload $\vec{x}_L$ records the geographic origin and destination of the lot, the elapsed transit time Δt_f , the cold chain trajectory, and any handover events between custody parties. The payload conforms structurally to a GS1 EPCIS 2.0 ObjectEvent with embedded SenML readings inside the sensorElementList.*

The forensically active component of $\vec{L}$ is the logistics latency Δt_f : biological degradation between harvest and processing is time dependent, and process telemetry at the plant is sensitive to it. A lot that sat in a warehouse for five days responds differently in the freeze dryer than one processed within twenty four hours, and that response is detectable as an anomaly under T_2 . Independent attestation by the transport operator isolates handling effects from those under the farmer's control: a producer can attest $\vec{A}$ honestly while disclaiming what happens once the lot leaves the gate.

3.3 Process Vector: Industrial Telemetry of the Transformation

Definition 4 (Process Vector). *The Process Vector $\vec{P}$ is an attestation vector of schema type τ_P , attested automatically by a_P , the control system of the transformation plant, whose payload $\vec{x}_P$ records the time integrated telemetry of the cycle and the post cycle at line measurements. The payload conforms structurally to a GS1 EPCIS 2.0 TransformationEvent that consumes a set of input EPCs and produces a set of output EPCs.*

In freeze drying, $\vec{x}_P$ contains the shelf temperature curve $T_s(t)$, chamber pressure $P_c(t)$, Pirani Baratron differential, product temperature $T_b(t)$ in witness vials, condenser temperature, mass loss curves, and time integrated quantities such as the energy density E_ρ per kilogram of water removed; at line measurements after the cycle (Karl Fischer humidity, water activity a_w , color, texture) extend the same vector. Signing is automatic at the PLC or SCADA level: the data path from sensor to signature is short, instrumented, and auditable, which makes $\vec{P}$ a strong forensic source. The same structural pattern applies to other transformations (coffee roasting, chocolate conching, fermentation, extrusion); the variable inventory differs but the signed object does not.

3.4 Recipe Vector: Operator Commitment to Process Parameters

Definition 5 (Recipe Vector). *The Recipe Vector $\vec{R}$ is an attestation vector of schema type τ_R , attested by a_R , the plant operator, whose payload $\vec{x}_R$ encodes the configurable parameters governing the transformation. The Recipe Vector is signed before the cycle begins and is referenced from the corresponding TransformationEvent in $\vec{P}$. The protocol supports selective disclosure: a BDT may carry only the canonical hash $\mathcal{H}_{\text{BMT}}(\vec{x}_R)$ as proof of commitment, with the full payload disclosed to authorised parties as required.*

In freeze drying, $\vec{x}_R$ encodes the programmed shelf temperature schedule, target chamber pressure, freezing protocol, annealing time, tray geometry, particle size, and any other operator selectable parameter that influences the outcome. The semantic distinction between $\vec{P}$ and $\vec{R}$ is forensically central: $\vec{P}$ reports observation, signed automatically; $\vec{R}$ reports intention, signed by the operator before

the work begins. Comparing the two yields an auditable record of execution fidelity that no single vector on its own can produce. Recipe parameters are often proprietary process knowledge; selective disclosure preserves the IP of the operator while keeping the integrity of the audit trail intact.

3.5 Vector Independence and Cross Referencing

The four core vectors are not redundant. Each addresses a different forensic question with a different attesting party and accountability regime: $\vec{A}$ what conditions the raw asset experienced at origin, $\vec{L}$ what happened to it between origin and plant, $\vec{P}$ what the transformation actually did, $\vec{R}$ what the operator committed to do. Cross referencing across vectors is what gives the protocol its power: an adversary attempting to falsify a claim must produce a coherent fabrication across all four sources, signed by all four independent parties, before downstream verifiers retrieve them and check.

The detection capabilities arising from this cross referencing are developed in Section 5 and mapped to the four tiers of Section 4. The protocol does not impose a universal level of attestation; each deployment selects the depth of evidence its use case requires.

4 The Tiered Attestation Model

The protocol does not impose a single standard of evidence. It parametrises one. A deployment in which two commercial partners simply want a tamper evident audit trail does not need the same machinery as a regulated denomination of origin defending its claim before an administrative tribunal. Both are legitimate uses of BDF, served by the same protocol without forcing one to bear the cost structure of the other.

Definition 6 (Tier). *A tier is a layer of attestation defined by the source of evidence it consumes, the verification operations it enables, and the class of inconsistencies it can detect. We index tiers by $n \in \{1, 2, 3, 4\}$ and write T_n for the tier of index n . The detection class $\mathcal{D}_n$ is the set of fraud patterns or inconsistencies that T_n can flag.*

The four tiers are organised by data source, marginal cost, and probative value, and are additive rather than alternative. T_1 is the cryptographic floor that every deployment inherits by construction. T_2, T_3 and T_4 are opt in capabilities selected according to use case, regulatory exposure, and budget. A buyer of recovered food carbon credits may operate at $T_1 \cup T_2$ and be fully served. A premium denomination of origin defending against substitution claims will reach into T_4 . The protocol is identical in both cases; what differs is the depth of the evidence stack the deployment chooses to populate.

4.1 Tier 1: Cryptographic Floor

T_1 is the integrity layer the protocol provides by definition. Every typed object is canonically serialised and content addressed (Section 2). Every ownership transfer is a signed statement on a deterministic Single Owner Chunk identifier. Every transformation terminates the input via burn semantics; every distribution forks a parent into a finite set of derivative packages. No deployment of BDF operates below this tier.

The class $\mathcal{D}_1$ is the set of inconsistencies expressible as failures of the structural predicates: hash integrity, ownership chain consistency, non reusability of burned inputs, atomicity of distribution events, provenance graph validity. It does not include any claim about the truthfulness of measurements: a farmer who signs a soil moisture reading commits the reading to the chain, but T_1 does not prove the sensor was calibrated or correctly placed.

4.2 Tier 2: Telemetry Anomaly Detection

T_2 runs statistical analysis on the telemetry that $\vec{A}$, $\vec{L}$, $\vec{P}$ and $\vec{R}$ already capture and sign under T_1 . It introduces no new sensors, no new equipment, and no new attesting parties; its incremental cost is software. Formally, $\mathcal{D}_2$ is the class of inconsistencies expressible as statistical or threshold tests over the signed payloads of the four core vectors.

The detection patterns fall into three families: conservation tests (a fresh mango lot at 80 to 85 percent water by mass that registers a final mass loss of 60 percent is inconsistent with the declaration), envelope tests (a sublimation rate curve that lies outside the band predicted by $\vec{R}$ and the input declaration triggers an anomaly), and cross vector coherence tests (the comparison of $\vec{R}$ with $\vec{P}$ for execution fidelity, of the logistics latency Δt_f in $\vec{L}$ with the freshness signal in $\vec{P}$ for batch substitution). T_2 is the canonical layer for recovered food carbon credit issuance, internal QA, fraud deterrence, and Scope 3 reporting. It does not authenticate origin at fine geographic resolution: two parcels in the same valley with the same cultivar will produce telemetry that T_2 cannot statistically distinguish.

4.3 Tier 3: Finished Product Spectroscopy

T_3 introduces a new physical measurement performed on the finished product: near infrared spectroscopy, hyperspectral imaging, or Raman, with a statistical classifier trained on a reference base. Unlike T_2 , T_3 requires equipment; unlike T_4 , that equipment can sit at the plant and read every batch automatically. The class $\mathcal{D}_3$ extends $\mathcal{D}_2$ with the inconsistencies expressible as classifier disagreements with declared identity at confidence threshold θ_3 , plus out of distribution flags.

Classification accuracy of 80 to 93 percent for geographic origin discrimination of freeze dried matrices is documented in literature precedents on blueberry, durum wheat, and meat cuts. T_3 is screening grade, not juridical: the signal captures the consequence of agronomic and processing conditions, not their cause, and is sensitive to cultivar drift, vintage effects, and equipment ageing. It depends on T_4 for its initial training and for periodic recalibration, even when it runs daily and T_4 runs annually.

4.4 Tier 4: Laboratory Forensic via External Attestations

T_4 is the forensic layer. It consists of evidence generated by accredited laboratories using mass spectrometric and isotope ratio techniques (ICP MS multielemental with rare earth elements, EA IRMS for $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{34}\text{S}$, MC ICP MS for $^{87}\text{Sr}/^{86}\text{Sr}$, optionally LC HRMS for untargeted metabolomic profiling), and delivered to the protocol as signed objects under a generic mechanism we call *External Attestations*.

Definition 7 (External Attestation). *An External Attestation is a typed signed object*

$$\vec{E} = (a_e, \tau_e, \vec{x}_e, r_e, t_e, \sigma_e)$$

where a_e is the identity of an external signer (an accredited laboratory, certifying body, regulatory auditor, or other trusted third party), τ_e is the schema type, $\vec{x}_e$ the payload, r_e a reference to the BDT the attestation pertains to, t_e the temporal anchor, and

$$\sigma_e = \text{Sign}_{a_e}(\mathcal{H}_{\text{BMT}}(\vec{x}_e, \tau_e, r_e, t_e)).$$

External Attestations follow the W3C Verifiable Credentials 2.0 envelope and may be canonicalised either via beeson for on chain anchoring or via JCS (RFC 8785) for off chain interoperability. We denote the (possibly empty) set of External Attestations anchored to a BDT O by $\mathcal{E}(O)$.

External Attestations are not a fifth attestation vector: they are a generic extension mechanism through which any third party signer can anchor evidence to the BDT under the same cryptographic regime as a core vector. The accreditation references carried within $\vec{x}_e$ (laboratory ISO 17025 accreditation number, signing date, method standard) give the attestation its juridical weight. The class $\mathcal{D}_4$ extends $\mathcal{D}_3$ with the inconsistencies expressible as bayesian rejections of the null hypothesis of declared origin at posterior threshold θ_4 (typically 0.99 for forensic deployments). Its strength is parcel level origin authentication and detection of organic versus synthetic fertilisation, geographic substitution, adulteration, and hydroponic origin masquerade. Its limit is coordinated collusion across all signers including the laboratory itself.

4.5 Tier Comparison and Additivity

Table 1 summarises the four tiers along the dimensions a deployment needs to choose between them.

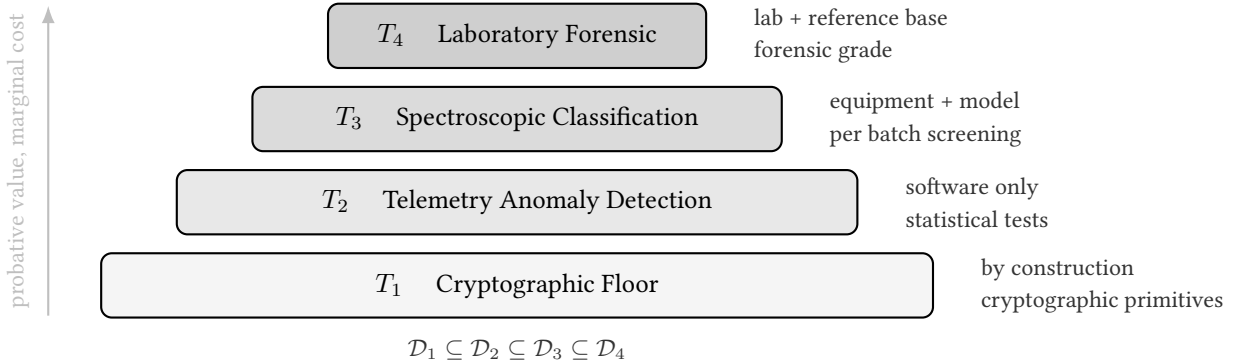


Figure 4: The four tiers of the BDF attestation model as a hierarchy of evidence. Lower tiers form the floor on which higher tiers are added. The detection classes are monotone increasing (Proposition 1). Probative value and marginal cost both grow with tier index.

Proposition 1 (Tier Additivity). *The detection classes of the four tiers form a monotone increasing chain*

$$\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \mathcal{D}_3 \subseteq \mathcal{D}_4.$$

A deployment that operates at tier T_n inherits, by construction, all detection capabilities of $T_1, \dots, T_n$. Each T_n extends T_{n-1} with a new source of evidence and a new class of detectable inconsistencies, without invalidating any of the lower tiers' guarantees.

Table 1: The four tiers of the BDF attestation model. Tiers are additive: a deployment running T_n inherits all guarantees of $T_1, \dots, T_n$.

Tier	Source of Evidence	Marginal Cost	Deployment Moment	Probative Value	Typical Cases	Use
T_1	Cryptographic protocol layer	Negligible	By construction, always	Structural integrity of audit trail	Bilateral commercial trust, baseline of any deployment	
T_2	Statistical analysis on signed telemetry	Software only	Continuous, on-line	Detection of gross fraud and inconsistency	Carbon credit issuance, QA, fraud deterrence, Scope 3 reporting	
T_3	Spectroscopic measurement of finished product	Equipment plus per batch operating cost in euros	Per batch, post production	Origin classification at medium resolution, screening grade	Continuous quality control, premium markets, pre shipment authentication	
T_4	Laboratory mass spectrometry and isotope ratios via External Attestations	Hundreds to over a thousand euros per sample	Periodic audit, dispute, recalibration	Forensic origin authentication, juridical grade	Denomination of origin defence, regulatory enforcement, reference base construction	

Proof sketch. By construction. Each tier consumes the evidence of all lower tiers as input to its detectors and adds at least one independent source. Detectors of T_n thus include, as a special case, every detector of T_{n-1} that depends only on lower tier evidence. The strict containment $\mathcal{D}_{n-1} \subset \mathcal{D}_n$ holds in any deployment where the additional source of T_n is populated; the inclusion is non strict in deployments that opt out of populating the higher tier. $\square$

The tiers are not a marketing taxonomy. They are a structural decomposition of the evidence supply problem in food traceability, anchored on the data sources available at each step and the cost structure of obtaining them. Cheaper T_4 instruments, more expressive T_3 models, and tighter T_2 detectors are technological refinements within the model; new tiers, if they emerge, will be additions, not replacements.

Part II

Operation

5 Per-Tier Detection Capabilities

This section formalises the detectors at each tier as typed functions on the BDT. The verification verdict at any tier is a composable, reproducible computation rather than a discretionary judgement.

Definition 8 (Detector). *A detector at tier T_n is a function*

$$D_n : \mathcal{O} \longrightarrow \{\text{pass}, \text{flag}\} \times [0, 1]$$

that maps a BDT $O \in \mathcal{O}$ to a verdict and a confidence score. The verdict is flag when the detector identifies an inconsistency in $\mathcal{D}_n$ and pass otherwise. The confidence score is the magnitude of the deviation under the detector's metric, normalised to $[0, 1]$. The detection class $\mathcal{D}_n$ is the set of inconsistencies for which there exists a detector at tier T_n that returns flag with non zero confidence.

A deployment may run multiple detectors at the same tier in parallel; the aggregate verdict at a tier is the disjunction of the individual detector verdicts. Composition across tiers is treated in Section 5.5.

5.1 Structural Verification at T_1

The detectors at T_1 are predicates over the cryptographic state of the BDT. The class $\mathcal{D}_1$ contains exactly those inconsistencies expressible as the failure of one of these predicates.

Hash integrity. For each attestation vector $\vec{V} = (a, \tau, \vec{x}, t, \sigma)$, the verifier recomputes the canonical hash $\mathcal{H}_{\text{BMT}}(\vec{x}, \tau, t)$ and checks that σ verifies against the public key of a . Any tampering of $\vec{x}$, τ or t breaks the signature.

Minter type consistency. Each attestation vector declares its minter type (Section 2). The verifier checks that the declared type is consistent with the registered profile of a (a key registered as a Proplet device cannot sign as `human_operator`). A mismatch is flagged. This predicate does not enforce truth of the underlying measurement; it ensures that downstream trust profiles cannot be silently spoofed by mislabelling the source.

Ownership chain consistency. For each ownership transfer $\sigma_i = \text{Sign}_{A_i}(H_O \parallel A_{i+1})$, the verifier checks that A_i was the legitimate previous owner and that no conflicting signature from the same A_i exists in the public record. Two conflicting signatures constitute a publicly provable double signing.

Burn non reusability. For any object O whose ownership chain terminates in a transformation address, no further valid transfer of O exists in the public record.

Provenance graph integrity. For any output object O' that references input objects $O_1, \dots, O_k$ in its TransformationEvent, the verifier checks that each referenced input exists, that each is burned, and that the input is referenced by exactly one output. For any package object $O_{p,k}$ or remainder R that references a parent O_p in a distribution event, the verifier checks that O_p is consumed by exactly one distribution event in the public record, that all $M + 1$ derivative twins of that event are jointly defined, and that the conservation invariant $\sum_k q(O_{p,k}) + q(R) = q(O_p)$ holds within the schema's declared tolerance.

These predicates are necessary conditions for any higher tier verdict to be meaningful: a deployment in which T_1 already returns flag should not advance to T_2 before the structural inconsistency is resolved.

5.2 Statistical Anomaly Detection at T_2

The detectors at T_2 are statistical and threshold tests over the signed payloads of the four core vectors. They organise into three families.

Conservation tests. Physical quantities that must be conserved across the transformation must conserve within their measurement uncertainty. The canonical case is the water mass balance: with m_0 initial mass, w_0 declared initial water fraction, and m_f final mass after the cycle,

$$m_0 \cdot w_0 - (m_f - m_0 \cdot (1 - w_0)) = \int_{t_0}^{t_f} \dot{m}(t) dt$$

within the measurement uncertainty of the load cells and mass flow sensor. A residual outside the uncertainty band is a T_2 flag for input misdeclaration. The same test admits an energy version.

Envelope tests. Time series must sit inside the envelope predicted by the recipe and the input declaration. Given $\vec{x}_R$ and an input declaration $\vec{x}_A^{\text{decl}}$, an empirical model f predicts an admissible band $[\underline{y}(t), \bar{y}(t)]$ for each observable $y(t)$. The envelope test flags any $y(t) \notin [\underline{y}(t), \bar{y}(t)]$ for a sustained interval. The model f is calibrated against the deployment's own historical batches.

Cross vector coherence tests. Two independently signed vectors must agree on quantities they both encode. The cleanest case is $\vec{R}$ versus $\vec{P}$: the recipe encodes $T_s^{\text{rec}}(t)$ and the telemetry encodes $T_s^{\text{real}}(t)$; the detector flags any sustained $|T_s^{\text{real}}(t) - T_s^{\text{rec}}(t)| > \theta_{T_s}$. The same pattern applies to chamber pressure and any other shared parameter. A second cross vector test compares $\vec{L}$ with $\vec{P}$: a freshness statistic $\phi(\vec{x}_P)$ derived from the early phase resistance is predicted by an empirical model $g(\Delta t_f, \vec{x}_R)$ that decays with logistics latency; a fresh tissue signature under a long Δt_f is the canonical *batch substitution anomaly*. A third test detects intra batch heterogeneity by flagging excess variance across parallel trays.

The thresholds θ_y and the predictive models f, g are calibrated empirically against the deployment's own historical batches through correlation mapping, degradation modelling, and anomaly thresholding. The mathematical shape of the relationships is determined by the data, not assumed in advance.

5.3 Spectroscopic Classification at T_3

Let $\vec{s}$ be the spectroscopic measurement of the finished product and $\mathcal{C} = \{c_1, \dots, c_K\}$ the set of classes the model discriminates. The classifier outputs a posterior distribution $\hat{P}(c_k \mid \vec{s})$, and the detector returns

$$D_3(O) = \begin{cases} (\text{pass}, \hat{P}(c^* \mid \vec{s})) & \text{if } c^* = \arg \max_k \hat{P}(c_k \mid \vec{s}) \text{ matches the declaration and } \hat{P}(c^* \mid \vec{s}) \geq \theta_3, \\ (\text{flag}, 1 - \hat{P}(c^* \mid \vec{s})) & \text{otherwise.} \end{cases}$$

The threshold θ_3 is chosen on the validation set; typical operating points place it in $[0.85, 0.95]$. A complementary detector flags out of distribution inputs based on the distance to the nearest class centroid in the model’s latent space, returning flag when $\vec{s}$ lies outside the support of the training distribution.

The classifier degrades over time and requires biannual recalibration against T_4 ground truth. The deployment’s T_3 detector at any point in time is therefore versioned, and the version identifier is part of the detector’s signed output.

5.4 Bayesian Forensic Verification at T_4

Let $\vec{e}$ be the laboratory measurement vector for the lot in question and $\mathcal{B}$ the reference base. Let H_0 be the null hypothesis that the lot belongs to the declared parcel p^* , and H_1 the alternative. The detector computes

$$P(H_0 \mid \vec{e}, \mathcal{B}) = \frac{P(\vec{e} \mid p^*, \mathcal{B}) \cdot P(H_0)}{P(\vec{e} \mid p^*, \mathcal{B}) \cdot P(H_0) + P(\vec{e} \mid \neg p^*, \mathcal{B}) \cdot P(H_1)}$$

and returns flag when $P(H_0 \mid \vec{e}, \mathcal{B}) < 1 - \theta_4$, with θ_4 typically 0.99 for forensic deployments. Multivariate likelihoods are typically modelled as multivariate normal in the appropriate transformed space, with covariance estimated from the reference base; robust alternatives based on Mahalanobis distance with Hotelling T^2 tests are equivalent under standard distributional assumptions.

The reference base $\mathcal{B}$ is itself a structured artifact, signed and anchored to the protocol as a special form of External Attestation. Its versioning, revision history, and method standards are part of the public record so that any T_4 verdict is reproducible by any verifier with access to the same base.

5.5 Composition of Detectors and Verification Verdict

A deployment running tiers $T_1, \dots, T_n$ produces an aggregate verdict by composition: pass if and only if every tier returns pass, flag otherwise. The aggregate confidence score is the maximum of the per tier scores when the verdict is flag, and the minimum when it is pass. A single high confidence flag at any tier dominates; a pass requires the weakest tier to remain robust.

Detector composition is independent of evaluation order, but the cost structure typically motivates an ascending evaluation: T_1 runs by construction on every transfer, T_2 runs continuously per batch, T_3 runs per batch with bounded equipment cost, T_4 runs on selected batches at audit cycles or on dispute. A deployment can short circuit higher tiers when a lower tier already returns flag, concentrating expensive analytical resources on the lots where additional evidence is needed. The verdict, the per

tier scores, and the version identifiers of the detectors used are themselves recorded as a typed signed object anchored to the BDT: reproducibility of verification, not authority of a single arbiter, is what gives the protocol its standing as a public verification rail.

6 Architecture and Data Governance

The protocol's verification capability sits on a three layer architecture, with data flowing through layers that have distinct persistence and access semantics. This section describes the architectural layers and the privacy mechanisms that govern data placement, retention, and selective disclosure.

6.1 Three Layer Architecture

Layer 1: Data signing and aggregation. Where attestation vectors are produced. IoT providers sign aggregated field readings at the gateway, transport operators sign logistics events, the plant control system signs aggregated process telemetry at the PLC or SCADA level after each batch, and the plant operator signs the recipe before the cycle. Aggregation happens at the source: raw time series are summarised into structured metrics before signing, because the full series would be impractical to handle on chain. No party signs data on behalf of another. This layer is the operational floor of T_1 .

Layer 2: Cryptographic anchoring. Where signed objects are committed to public verification rails. Canonical hashes are anchored to a content addressed storage network for retrieval; ownership transfers are committed to public chains for dispute resolution and timestamping. The blockchain does not perform runtime validation of measurements; it serves as an immutable reference layer. Single Owner Chunk identifiers and burn semantics from Section 2 are realised here.

Layer 3: Post hoc verification. Where the detectors at T_2, T_3, T_4 operate. Statistical anomaly detection consumes the canonicalised payloads of signed vectors. Spectroscopic classification consumes the signed measurements of the finished product. Forensic laboratory verification consumes the External Attestations anchored to the BDT. Verification verdicts are themselves recorded as signed objects on the public record so that any third party can reproduce them. The protocol provides the structured data; the verification logic runs anywhere, including as a standalone tool, a service, or embedded in a marketplace application.

The architecture relies on cryptographic signing at the software level. Hardware roots of trust progressively strengthen the integrity of Layer 1 (Section 10).

6.2 Layered Data Placement

The three architectural layers map to three data placement layers with distinct persistence and access semantics. The distinction is essential to reconcile the integrity of an append only audit trail with the right to erasure recognised by GDPR Art 17 and analogous frameworks.

Public chain layer. Carries only cryptographic commitments: hashes of canonical objects, signatures on ownership transfers, SOC identifiers. Persistence is permanent by construction. No PII, no commercial sensitive data lives here.

Content addressed storage layer. Carries the encrypted payloads of attestation vectors and External Attestations. Each chunk’s persistence is conditional on a valid postage stamp; when the postage balance is depleted, chunks become eligible for garbage collection [Trón, 2024]. Persistence is economically sustained, not permanent. Retention policies can be enforced natively: ceasing to renew postage is equivalent to deletion at the storage layer.

Private operator systems. Carry raw sensor data, employee records, and pseudonymous identity mappings. PII never appears on chain or on Swarm; private systems follow standard data protection practice (GDPR Art 25, privacy by design).

6.3 Encryption, Crypto Shredding, and Selective Disclosure

BDF deployments encrypt the payload of every attestation vector before storage. The encryption uses the chunk-level mechanism of Swarm: each chunk is encrypted with a key, and the chunk reference embeds both the address and the decryption key [Trón, 2024]. Without the key, the chunk is indistinguishable from random data, providing plausible deniability.

Crypto shredding is the protocol’s retrospective deletion mechanism. The encrypted payload remains in storage but the encryption key is destroyed in the operator’s private system; the data becomes uniform random bytes to any reader. The destruction event is recorded as a signed object on chain, providing auditable evidence that erasure occurred.

Commercial confidentiality is preserved by selective disclosure of decryption keys. The Recipe Vector $\vec{R}$ is the canonical case: the recipe is committed via its hash, the plaintext lives encrypted on Swarm, and the operator decides which keys to share with whom (an auditor under NDA, a regulator under enforcement powers, a buyer under contractual right). Swarm’s access hierarchy extends this to multi-party scenarios with different keys granting access to different fields within the same BDT.

6.4 Retrospective Access Revocation via DREAM

For the strongest right to erasure case, BDF deployments use Swarm’s DREAM construction [Trón and Diedrich, 2025, Trón, 2024], which provides retrospective access revocation with quantitative guarantees. The DREAM reference distributes the chunk identity across multiple network neighbourhoods; granting access requires the grantee to follow a path through these neighbourhoods. To revoke access, the uploader publishes new chunks to a designated batch that disrupts the path. As long as at least one neighbourhood on the path is honest, the disruption is effective.

The probability of successful revocation depends on path length and the honesty rate of the network. With 90 percent honest neighbourhoods and a 6 hop path, the probability of successful revocation exceeds 99.9999 percent. This is retrospective access revocation with cryptographic backing, structurally distinct from policy based revocation in centralised systems.

6.5 Ownership versus Access

The protocol separates two concerns that other systems often conflate.

Ownership of the BDT is the on-chain status: the chain $A_0 \rightarrow A_1 \rightarrow \dots \rightarrow A_n$ with append only signed transitions. The current owner A_n holds the right to act on the BDT.

Access to the data inside the BDT is the off-chain status: who holds the encryption keys for which payload fields. Access is flexible, granular, and revocable; ownership is lineal, atomic, and append only.

When a BDT changes hands, the outgoing owner has four independent options. Retain access (keep keys) and continue to verify their past contributions. Grant complete access to the new owner (share all keys) for full transparency. Grant partial access (share some keys, retain others) to preserve privacy of sensitive contributions. Destroy keys and trigger a DREAM revocation against parties with whom they previously shared them, exercising the right to be forgotten with respect to data they contributed. The separation by design lets ownership remain strictly lineal (audit unambiguous) while access remains flexible (privacy granular).

6.6 Regulatory Compliance

The mechanisms above support compliance with the most relevant data protection and supply chain regulations. Table 2 summarises the mapping. A comprehensive compliance dossier with jurisdiction specific guidance is the subject of a companion document [Fair Food Data, 2026a].

Table 2: Mapping of BDF privacy mechanisms to relevant regulatory frameworks.

Framework	Provision	BDF mechanism
GDPR (EU)	Art 17 right to erasure	Crypto shredding; DREAM revocation; postage TTL
GDPR (EU)	Art 25 privacy by design	Layered placement; PII confined to private operator systems
GDPR (EU)	Art 5 minimisation	Only hashes on chain; encrypted payloads on Swarm
CSRD (EU)	Data quality, audit trail	Signed attestation vectors with verifiable provenance
CBAM (EU)	Provenance verification	Selective disclosure to customs and regulators
FSMA Section 204 (US)	Critical Tracking Events	Lifecycle events anchored to BDT; Key Data Elements preserved with selective access
ICR, Verra, Gold Standard	dMRV requirements	Auditable evidence stack with documented privacy framework

7 Adversarial Model and Per Tier Guarantees

The protocol provides quantifiable guarantees under explicit threat models and trust assumptions. This section gives an informal summary suitable for a general technical audience. The formal threat model with adversary typology, trust assumptions, and security reductions with proof sketches is developed in a companion paper [Fair Food Data, 2026b].

7.1 Threat Landscape

The protocol considers five adversary types of increasing power: a lone fraudster (single corrupted attestor), a coalition (k of n corrupted attestors with internal coordination), a network adversary (controls a fraction of the underlying storage and routing layer), a computational adversary (polynomial time bounded, cannot break cryptographic primitives at 128-bit security), and a coordinated cartel (all attestors plus laboratory corrupted simultaneously). Guarantees against the first four are detailed below. The cartel represents residual risk addressed by the Hardening Pathways of Section 10.

7.2 Guarantees by Tier

The guarantees rest on eight trust assumptions: cryptographic primitives at standard 128-bit security, deterministic canonicalisation, sufficient routing diversity in the underlying storage network with quantified honest neighbourhood fraction, an honesty threshold among attestors, functional postage economics, anchoring chain security, and ISO 17025 accreditation for laboratories issuing External Attestations.

T_1 guarantees structural integrity of the audit trail (immutable identity, conflict detection, burn non reusability, provenance graph validity) under standard cryptographic assumptions. Object collision and signature forgery probabilities are bounded by 2^{-128} . Probabilistic fork resistance scales exponentially with the number of independent verifiers.

T_2 detects fan-in conservation violations deterministically when at least one input signer is honest, and cross vector inconsistencies statistically when at least half the relevant vectors are signed by independent attestors. Detection thresholds are calibrated against the deployment’s reference set.

T_3 classifies origin at medium resolution under the assumption that the reference base is uncorrupted and isolated from production. Reported accuracy depends on cultivar, vintage, and equipment ageing; periodic recalibration is required.

T_4 produces forensic verdicts at Bayesian posterior threshold θ_4 (typically 0.99) against single corrupted attestors, conditional on laboratory integrity. T_4 does not protect against the coordinated cartel.

DREAM revocation succeeds with probability bounded by $1 - 10^{-n}$ for n hop paths and 90 percent honest neighbourhoods.

7.3 Signer Independence and Reference Set Integrity

Cross vector detection in T_2 depends on attester independence. In deployments where signers are commercially coupled (vertically integrated supply chains), organisational independence is not always available. The protocol restores independence at three layers: cryptographic (distinct keys), cryptoeconomic (staked and slashable), and hardware (Proplet attestation). Section 10 develops these. In the absence of these mechanisms, deployments with coupling must mandate T_4 verification for high stakes claims.

The reference set used for T_2 calibration is itself a target for slow rolling adversarial training. The protocol mitigates this through three combined mechanisms: separation of reference and production sets, external validation of reference updates by third party auditors, and reputation onchain bound to the integrity of the reference set. A reference contamination later proven by audit publicly slashes the deployer’s reputation.

7.4 Residual Limitations

The protocol does not guarantee:

- Truthfulness of physical measurements before signing, in the absence of Proplet hardware.
- Authenticity of laboratory measurements outside External Attestations issued by ISO 17025 accredited laboratories.
- Protection against the coordinated cartel (Type 5).
- Security against majority capture of the underlying storage or routing network beyond the threshold honesty fraction.
- Robustness of T_3 classifiers under adaptive adversarial inputs without explicit adversarial training.

The protocol’s strength is in opportunistic and partial collusion detection, where compromising a subset of signers is insufficient to produce a coherent narrative across the others. The Hardening Pathways of Section 10 progressively narrow the residual risk as Proplets, cryptoeconomic mechanisms, and zero knowledge proofs mature in deployment. Formal proofs and per assumption reductions are documented in the security companion paper [Fair Food Data, 2026b].

Part III

Application and Practice

8 Empirical Pilot Results

This section reports the first empirical results from a production pilot of the protocol, covering the full pipeline from harvest in field to lyophilisation in plant. The data are reported with explicit limitations and discussed in the context of the Tiered Attestation Model defined in Section 4. The goal of this

section is not to prove the protocol’s superiority over incumbent methodologies but to document what the first deployment captured, what the Tier 2 detectors flagged, and what remains for subsequent pilots to address.

8.1 Pilot Configuration

The pilot instrumented a freeze drying campaign of blueberry across three cultivars (Biloxi, Ventura, O’Neal) on four agricultural blocks. The Agronomic Vector $\vec{A}$ was captured by four soil sensors (Temperature, Humidity, Conductivity, with one sensor additionally measuring pH) and a weather station with ten variables, over a span of approximately seven weeks. The Logistic Vector $\vec{L}$ was captured for each basket through harvest and reception events, including origin and destination coordinates, transit time, and weight measurements at both ends. The Process Vector $\vec{P}$ was captured at the freeze drying plant by the SCADA system of the lyophiliser, sampling chamber pressure, chamber temperature, nine heater temperatures, instantaneous power consumption, and cumulative process energy at approximately sixty second intervals over a single 43.5 hour cycle covering seven sequential stages.

The genesis attestation in this pilot was signed by human operators in the field. No Proplet hardware (Section 10.2) was deployed. The genesis attestation declared its minter type as `human_operator`, which a downstream verifier interprets accordingly when applying trust profiles to the contained measurements.

8.2 Agronomic Profile

Across the seven week observation window, the weather station logged 1,051 readings. Mean ambient temperature was 16.7 degrees Celsius (range 4.7 to 38.0), mean relative humidity 77.6 percent, Vapor Pressure Deficit averaged 0.74 kPa, and Reference Evapotranspiration averaged 0.80 mm per interval. Soil sensors registered humidity in the range 54.6 to 67.5 percent, soil temperature 18.2 to 21.4 degrees Celsius, electrical conductivity 0.06 to 0.34 dS per metre, and soil pH at 8.0. The alkaline pH is notable: blueberry cultivars favour acidic substrates, and a pH of 8.0 represents an agronomic stress condition that the Agronomic Vector documents as part of the lot’s history. This is exactly the kind of contextual information that becomes auditable evidence when bound cryptographically to the genesis of the BDT.

8.3 Logistic Vector and Mass Flow

Of the seventeen baskets initiated in the pilot, fifteen completed the full origin to reception pipeline. Origin weights declared in the field had a mean of 3.20 kg per basket (standard deviation 1.08, sum 48.00 kg). Facility weights at plant intake had a mean of 3.55 kg (standard deviation 0.86, sum 53.27 kg). Transit times ranged from 3 to 837 minutes with a mean of 369 minutes, reflecting heterogeneity between baskets harvested near the plant and those transported over longer routes.

A systematic discrepancy is observable in the field versus facility weight comparison: the sum of facility weights exceeds the sum of origin weights by 5.27 kg, an apparent net gain of 11.0 percent during transit (Figure 5). This direction is physically implausible for biological transit, where mass typically decreases through transpiration and handling loss. The most plausible cause is that origin weights were operator estimates anchored to the nominal capacity of the basket (3 kg by basket design) rather

than physical measurements, while facility weights came from a calibrated scale. This is exactly the empirical signature the Tier 2 conservation invariant detects (Section 5.2).

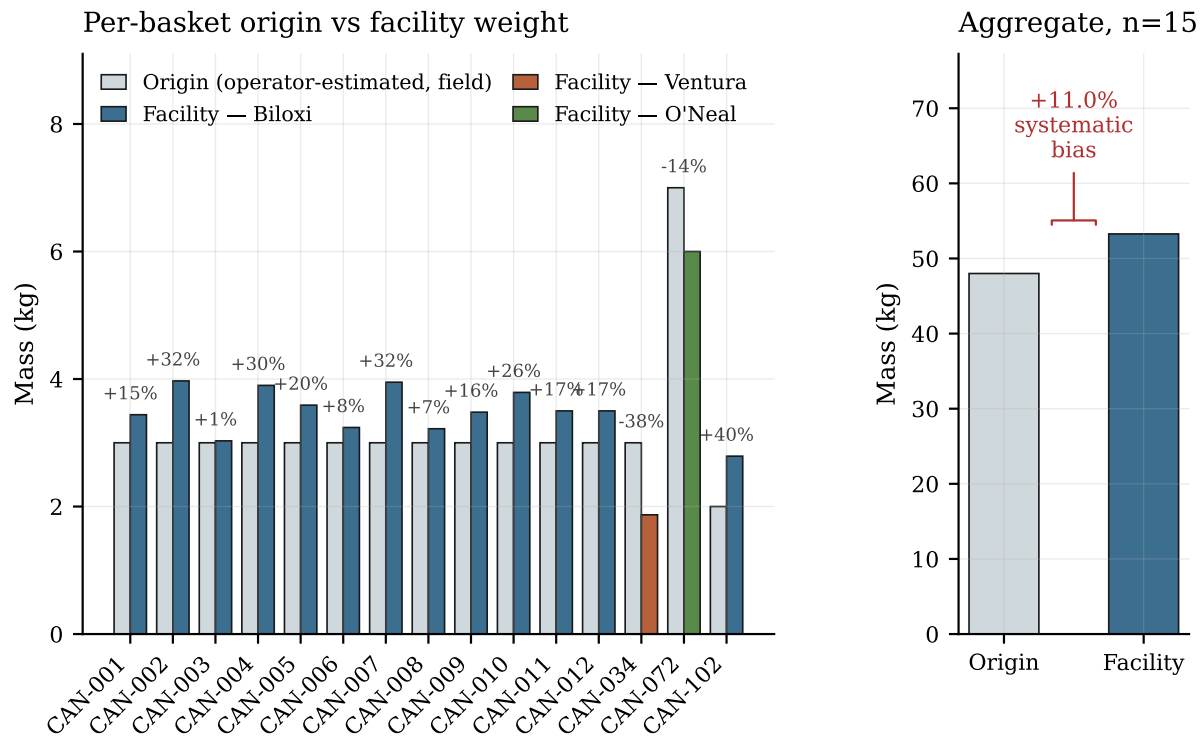


Figure 5: Tier 2 conservation invariant applied to the pilot’s Logistic Vector. Per basket, the operator declared origin weight (light grey) is plotted against the calibrated facility scale weight (coloured by cultivar). The aggregate panel sums the fifteen baskets that completed the full pipeline: the facility scale records 53.27 kg against 48.00 kg declared at origin, an 11 percent systematic positive bias that the conservation invariant flags on the first calibration cohort. The bias is concentrated in the Biloxi cultivar where origin weights cluster at the 3 kg nominal basket capacity, indicating an anchoring artefact rather than random measurement error.

8.4 Process Vector

The freeze drying cycle ran for 43.5 hours across seven sequentially numbered stages (Figure 6). Cumulative process energy at the end of the cycle was 87.61 kWh. Mean instantaneous power consumption was 2,009 W (peak 5,594 W, standard deviation 898 W). Stage 2 (primary sublimation) was the longest single phase at 600 minutes, with chamber pressure mean 491 mTorr and chamber temperature mean -28.6 degrees Celsius. Stage 7 (secondary drying) was the second longest at 966 minutes, with chamber pressure mean 1,706 mTorr (with anomalous excursions discussed below) and temperature mean -34.2 degrees Celsius. Heater temperatures ranged from -31 to 55 degrees Celsius depending on heater position and stage; one heater (heater 3) maintained substantially lower temperatures throughout the cycle, consistent with a condenser role.

8.5 Tier 2 Detector Demonstrations

The pilot data exhibits three specific anomalies that the Tier 2 detectors defined in Section 5.2 identify, plus one sanity check that passes within established benchmarks. Each is reported both as evidence

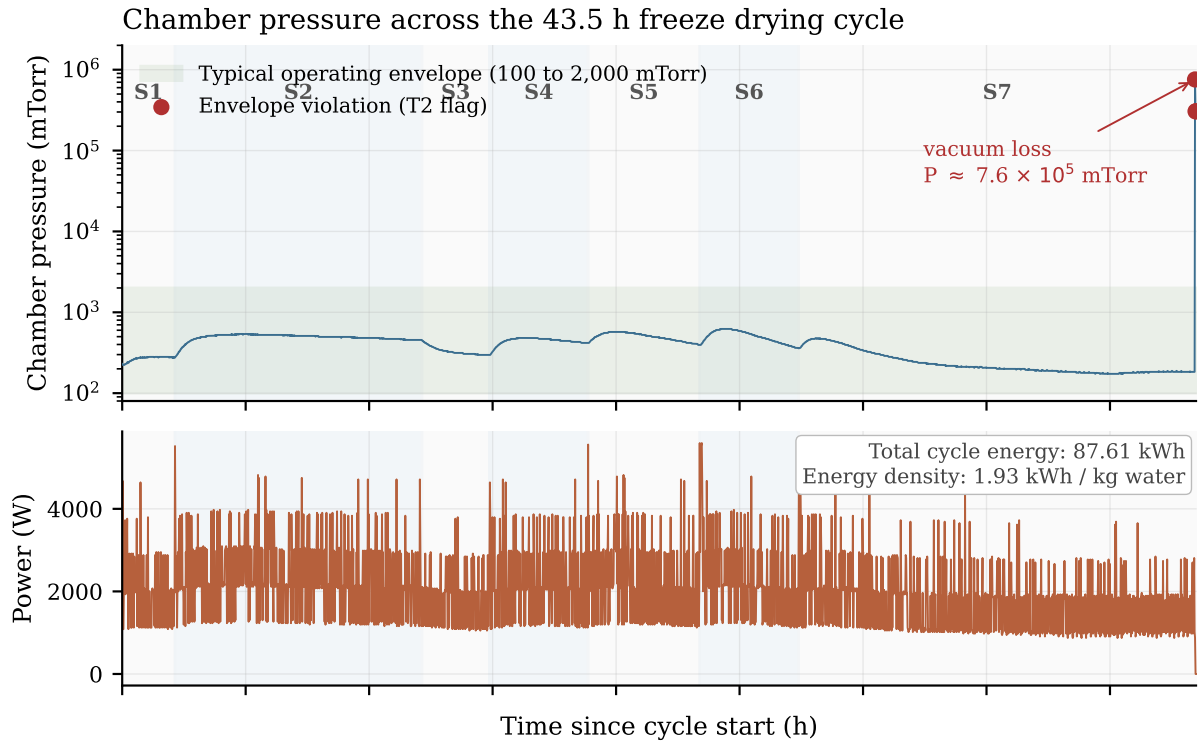


Figure 6: Process Vector telemetry from the pilot freeze drying cycle. Top panel: chamber pressure on logarithmic scale across the 43.5 hour cycle, with stage shading and the typical operating envelope (100 to 2,000 mTorr) highlighted. Three readings during stage 7 register pressures above 5×10^4 mTorr, with a peak at 7.6×10^5 mTorr approaching atmospheric pressure. The Tier 2 envelope test for chamber pressure conditional on stage flags these as immediate violations. Bottom panel: instantaneous power consumption over the same cycle, integrating to 87.61 kWh and yielding the 1.93 kWh per kg energy density reported below.

that the detection layer is empirically operational and as evidence of the trust gaps that the Hardening Pathways (Section 10.2) close.

Conservation invariant detects systematic bias in field weights. The mass balance check between origin and facility weights flags a systematic positive bias of 11 percent across the fifteen baskets. The bias is not random noise: every basket in the most populated cultivar (Biloxi, $n = 13$) shows the same direction of discrepancy, with origin weight near the basket nominal capacity and facility weight at the actual scale measurement. The Tier 2 detector correctly identifies this as a calibration or measurement provenance issue. Without the detector, this discrepancy would propagate undetected through the rest of the supply chain and into any downstream claim. With it, the discrepancy is flagged on the first calibration cohort and the deployment can either deploy field weighing infrastructure to close the gap, relabel the field operator’s role as estimation rather than measurement, or deploy Proplet compliant scales at the basket level.

Envelope test detects pressure excursion in stage 7. Of the 2,522 chamber pressure readings recorded during the cycle, three readings during stage 7 register pressures above 100,000 mTorr, two orders of magnitude above the expected envelope for that stage (typical secondary drying pressure 100 to 600 mTorr). Pressures of this magnitude approach atmospheric pressure (760,000 mTorr) and

indicate either a transient loss of vacuum or a sensor failure. The envelope test for chamber pressure as a function of stage flags these excursions immediately as T_2 anomalies. The deployment can then reconstruct the cause from the surrounding telemetry and document the process incident with cryptographic timestamps as part of the audit trail.

Cross vector coherence detects individual outliers. Two baskets exhibit per-basket weight changes inconsistent with their reported transit times. Basket CAN-034 (Ventura) shows a weight delta of 1.13 kg over a reported transit of 5 minutes; basket CAN-072 (O’Neal) shows a delta of 1.0 kg over 837 minutes. The first is biologically implausible for the duration; the second is at the upper limit of plausible dehydration. Both are flagged by the cross vector coherence test that compares logistic latency Δt_f with the observed weight delta.

Sanity pass: energy density per kg of water removed. Total cycle energy of 87.61 kWh divided by the estimated water mass (45.28 kg under the assumption of 85 percent water content for blueberry) yields 1.93 kWh per kg of water removed. This falls within the industrial benchmark range of 1.0 to 2.0 kWh per kg reported in the freeze drying literature [Pikal and Shah, 1990]. The sanity check passes with the result close to the upper bound of efficiency, indicating a cycle that is operating but with optimisation room.

8.6 Limitations and Ongoing Work

The pilot reported here is a first production deployment and the empirical scope is bounded. Of the seventeen baskets initiated, fifteen completed the full origin to reception pipeline. Only one lyophilisation cycle is included in the dataset. Statistical significance for T_3 classifier training requires substantially larger cohorts, typically at least fifty baskets per cultivar across three vintages. No T_4 laboratory analysis was performed on this pilot batch, so the reference base required for forensic verification has not been constructed yet.

A 25 day gap between the final harvest event and the start of the lyophilisation cycle indicates an intermediate storage phase that this pilot did not instrument. In a fully populated deployment, a Storage Vector or an extension of the Logistic Vector should capture the conditions of this period (typically frozen storage, with its own temperature and integrity logs).

The genesis attestation in this pilot was performed by human operators with no hardware root of trust at the field. The systematic mass bias documented in Section 5.2 is a direct empirical illustration of the trust assumption that Hardening Pathways (Section 10.2) addresses. Subsequent pilots will progressively deploy Proplet compliant sensors at the field, transport, and plant intake to close this gap and shift the genesis attestation from human operator to hardware attested measurement.

The Process Vector data from this pilot is rich enough to seed a per-cultivar envelope model for the T_2 detectors. Calibration of these envelopes against further cycles, and tuning of the anomaly thresholds in collaboration with plant operators, is the next workstream.

9 Applications

The protocol is general purpose infrastructure for verifiable provenance of transformed bioassets. Different applications populate different tiers of the attestation model, according to the operative property each application requires evidence for. We discuss six classes of application below. The first five are the immediate addressable cases for our pilot deployments. The sixth opens the protocol to verticals beyond freeze drying and to evidence types beyond geographic origin.

9.1 Verifiable Provenance for Premium Markets

Buyers of premium freeze dried products, of denomination of origin (DOP, IGP) goods, and of high value gourmet ingredients want defensible evidence that the product they purchase comes from where the producer claims and was processed as the producer claims. The operative property is parcel level origin authentication. The relevant tiers are T_3 and T_4 , with T_1 and T_2 as the structural floor.

A buyer scans the QR code of a retail package, resolves to the package's BDT, and through it to the full lineage of the parent processed lot. The producer's claim becomes auditable rather than declarative. A reference deployment for this application class typically runs T_3 continuously on every batch through a benchtop NIR or hyperspectral imaging system with a classifier trained against a multivintage reference base, and reaches into T_4 annually on a representative subset to maintain the base and provide forensic backing in cases of dispute. The implementation roadmap for this configuration is detailed in Section 10.

9.2 Carbon Markets and dMRV for Recovered Food Carbon

The voluntary carbon market has bifurcated. The high integrity tier, around 35 percent of credits, now commands roughly four times the price of paper credits. The remaining 65 percent is being repriced or repriced out. The defining property of the surviving tier is digital monitoring, reporting and verification (dMRV): credits backed by hardware grade evidence rather than paperwork.

The methodology landscape for food has a structural gap. Verra's existing methodologies (notably ACM0003, VM0042) cover feedstock production at the agricultural plot level. Gold Standard's Climate and Land Use framework extends similarly. Neither covers industrial transformation of food assets at the recipe and process level. The closest precedent is the Sustainable Aviation Fuel chain, where Verra VM0085 and ISO 18752 audit the transformation of feedstock with process level energy balance certification. The same rigour is technically feasible for freeze drying, drying, fermentation, roasting, and other transformations of recovered or surplus food, and is currently a methodological white space (Figure 7).

BDF is positioned as the verification primitive on which a methodology for recovered food carbon credits can be built. The operative properties for carbon credit issuance are the verifiable energy balance of the transformation, the verifiable mass and water balance, the verifiable origin (at the regional or country level, sufficient for jurisdictional carbon accounting), and the verifiable absence of double counting through the protocol's burn semantics. None of these requires Tier 4. Tier 1 (T_1) provides cryptographic integrity. Tier 2 (T_2) provides anomaly detection over the signed energy and mass telemetry.

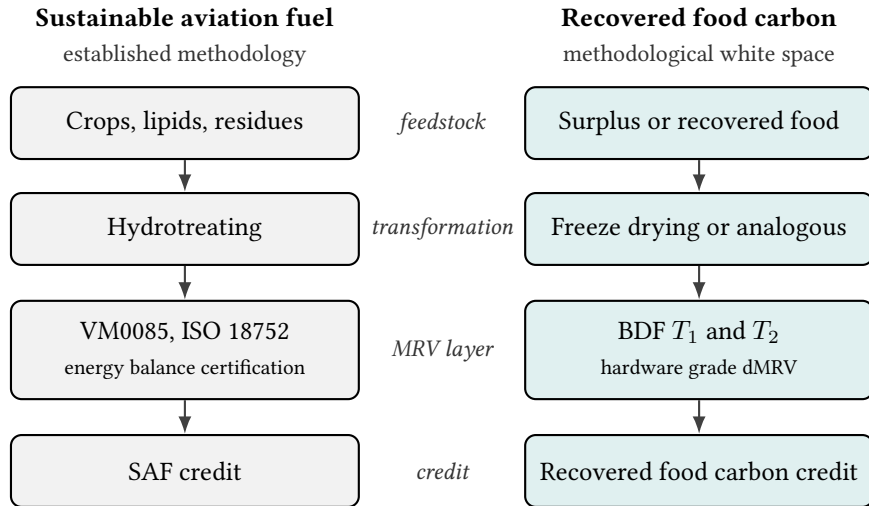


Figure 7: BDF positions itself in the methodological architecture for recovered food carbon credits analogously to how Verra VM0085 and ISO 18752 certify the transformation step in Sustainable Aviation Fuel. The structural pattern (feedstock, transformation, MRV layer, issued credit) is shared. The vertical and the methodology are distinct. The white space on the right is not occupied by any incumbent registry today.

The dMRV layer that a carbon credit registry needs is precisely $T_1 \cup T_2$, with optional T_3 for added resolution on origin claims when the credit class requires it.

We refer to this class of credit as *recovered food carbon*: emissions avoided through the recovery and processing of food that would otherwise have been wasted, with measured, attestable energy use and provenance at every step. The sub category is novel and has no incumbent infrastructure. BDF defines it operationally and provides the verification rail on which it can be issued, audited and traded.

9.3 ESG Reporting and Scope 3 Disclosure

Corporate sustainability reporting under CSRD in the European Union and analogous frameworks elsewhere obliges firms to disclose Scope 3 emissions across their supply chain. The CBAM regime adds a parallel obligation at the import boundary. Both regimes accept disclosure today on the basis of estimated factors and aggregated supplier data. Both are converging towards stricter evidence requirements as the regulatory infrastructure matures.

A firm sourcing freeze dried ingredients can use the BDF BDT of each lot as the evidentiary basis for its Scope 3 disclosure. The energy density per kilogram of water removed is signed in the Process Vector. The logistics distance and cold chain integrity are signed in the Logistic Vector. The agronomic origin is signed in the Agronomic Vector. The disclosed Scope 3 number for the lot is the deterministic aggregate of these signed quantities under the firm's chosen accounting methodology. The chain from physical events to disclosed number is auditable end to end.

The relevant tiers for ESG disclosure are typically $T_1 \cup T_2$. The disclosure does not generally require parcel level authentication. It requires that the numbers reported can be retraced to signed observations on the public record.

9.4 Quality Assurance for Producers

A producer running freeze drying or analogous transformations operates under quality constraints: residual moisture, water activity, color, texture, retention of polyphenols and antioxidants, and homogeneity within batch. The producer's internal quality control already measures these quantities at line. With BDF, those measurements become signed components of the Process Vector and feed the T_2 detectors automatically.

A T_2 detector flagging a batch before shipment is a quality intervention. The producer can rework, downgrade, or reject the batch internally rather than ship a non conforming lot to a buyer. The cost saved is the cost of a downstream quality dispute or product recall. The relevant tier is T_2 . The marginal cost of the application is software, since the data is already collected and signed.

9.5 Fraud Deterrence

Food fraud takes two operative forms in the matrices BDF addresses. Substitution fraud, where the actual physical input differs from the declared input. Adulteration fraud, where the input is mixed with materials of inferior quality, different origin, or different cultivar. The protocol detects both forms at different tiers.

T_2 detects opportunistic substitution and adulteration through cross vector inconsistency. A lot declared as fresh that shows aged tissue in the Process Vector is flagged. A homogeneous lot that shows heterogeneous telemetry across parallel trays is flagged. A lot with logistics latency inconsistent with the declared freshness is flagged. These detections are continuous and automatic.

T_3 extends fraud detection to cultivar substitution and gross origin substitution at screening grade. T_4 extends to parcel level substitution and adulteration at forensic grade.

A complete fraud deterrence deployment runs T_1 and T_2 continuously on every batch, T_3 as ongoing screening when equipment is available, and T_4 on disputed batches or audit cycles. The aggregate cost structure makes the application economically rational at any scale of producer that already has price exposure to fraud claims.

The honest limitation is that the protocol does not detect fully coordinated collusion across all signers. Section 7 states this explicitly. Section 10.2 discusses pathways to close the residual risk through hardware and cryptoeconomic means.

9.6 Extensibility to Other Verticals and Evidence Types

The transformation agnosticism of the four attestation vectors (Section 3) opens the protocol to verticals beyond freeze drying: roasting, conching, fermentation, brewing, extrusion, and analogous processes. The External Attestation mechanism (Section 4.4) opens it to evidence types beyond geographic origin: organic certification, no deforestation declaration, labor compliance audits, allergen control, halal and kosher certifications. Each vertical and each evidence type is added through a typed schema and a recognised attesting party, without modification of the protocol's core. The protocol becomes the public verification rail for a growing set of evidence classes about the same physical product.

10 Roadmap and Future Work

This section combines the implementation roadmap for production deployments with the open research and engineering directions that strengthen the protocol’s guarantees and extend its scope.

10.1 Implementation Roadmap

A production deployment of the protocol involves a sequence of workstreams that progressively populate the tiers, accumulate the reference base required for upper tier detection, and align with certification milestones.

Per tier workstreams. T_1 is the structural floor and is operational from the moment the protocol is instantiated (beeson canonicalisation, content addressed storage, ownership chain, SOC feeds). T_2 requires detector calibration: conservation tests, envelope tests, and cross vector coherence tests calibrated against an initial cohort with threshold tuning for false positive and false negative rates. T_3 requires spectroscopic equipment procurement and classifier training against the T_4 reference base, with biannual recalibration. T_4 requires laboratory analysis of a representative cohort to construct the reference base, signed and anchored as External Attestation with periodic refresh.

Operational profile. Once trained, T_3 runs on every batch automatically through benchtop NIR or hyperspectral imaging on conveyor. Per batch operation requires no operator, laboratory, or sample shipment. Human and laboratory effort is concentrated in initial reference base construction (one time per matrix, of order four to eight hundred thousand euros analytical cost) and periodic recalibration (semestral, in the order of a few thousand euros).

Certification milestones. Deployments targeting carbon markets organise themselves against International Carbon Registry (ICR) certification for the recovered food carbon credit class, with a subsequent methodology submission to Verra or Gold Standard paralleling the energy balance certification of VM0085 and ISO 18752 for sustainable aviation fuels. ISO 17025 accreditation for the laboratory partner conditions the juridical weight of the T_4 reference base. The methodology itself is the subject of a companion paper [Fair Food Data, 2026a].

10.2 Hardening Pathways

The adversarial model leaves a residual risk in fully coordinated collusion across all attestors. Three complementary directions progressively close it.

Proplet compliant sensor architecture. Following Szabo [Szabo, 2001], IoT sensors and gateways can be re-engineered as tamper resistant devices with hardware root of trust, remote attestation, GPS based self geofencing, and on chain ownership registry. Under this architecture, the trust assumption shifts from operator to hardware: the device itself is the cryptographic principal that signs attestations. This includes the genesis attestation, where a Proplet field sensor autonomously mints the BDT at the

moment it physically witnesses the harvest event, removing the human operator from the genesis loop. Components are commercially available today (secure elements, tamper resistant enclosures, TPM 2.0 with remote attestation).

Cryptoeconomic mechanism design. Each attester posts a stake released after a challenge window; a successful fraud proof slashes the stake. Optimistic challenges allow any verifier to produce a fraud proof using the Tier 2 to 4 detectors and earn a bounty if the challenge prospers. Reputation onchain accumulates with honest attestation and is destroyed by proven fraud. Insurance pools collect premiums from attestors and pay claims when fraud is proven. Calibrating stakes above the marginal benefit of fraud converts coordinated collusion into expected economic loss.

Zero knowledge proofs for selective verification. ZK enables a verifier to confirm a property of a payload without seeing the payload: that GPS coordinates fall inside a declared denomination of origin polygon without revealing the parcel, that energy density is below a regulatory threshold without revealing the recipe. ZK is currently practical for simple circuits and impractical for complex multivariate forensic computations; adoption is incremental.

The three directions reinforce each other. Slashing is operationally robust only when the attester is identifiable hardware with persistent on chain identity. Proplets without economic incentives are tamper resistant but do not punish coordinated collusion. ZK proofs enable verification without revelation but do not erase data. Together, they close the residual risk of the adversarial model to a quantifiable economic parameter.

10.3 Open Research and Extensions

Open conjecture: biophysical causality of sublimation telemetry. A research conjecture motivating the original BDF design, but not a precondition of the protocol, is that agronomic conditions of a crop shape its cellular structure and that cellular structure dictates its thermodynamic and chemical behaviour during sublimation [Roos, 1995, Pikal and Shah, 1990, Murphy and Koop, 2005]. If the conjecture holds quantitatively, the Process Vector contains recoverable information about the Agronomic Vector. Validation would extend T_2 detection with cross vector consistency tests of higher specificity. The pilots described in Section 8 provide initial paired data; formal investigation is open.

BDF Standards Stack. A companion paper [Fair Food Data, 2026c] documents the open standards composed by the protocol (SenML, GS1 EPCIS 2.0, FoodOn, SAREF4AGRI, PROV-O, W3C VC 2.0, JCS, beeson, Swarm) and the canonical mapping from each attestation vector to its standards based payload, intended as implementation reference for new deployments.

Multi vertical expansion. The transformation agnosticism documented in Section 3 will be operationalised through vertical specific schemas and detector calibrations for roasting, conching, fermentation, brewing, and extrusion. Biological transformations on month long timescales (cured cheeses, aged spirits, fermented vegetables) require adaptation of operational tooling and Process Vector sampling cadence.

Methodology development for carbon markets. The recovered food carbon application of Section 9 opens a path for methodology submission to a major carbon credit registry. The companion paper [Fair Food Data, 2026a] defines the methodology in a form compatible with Verra or Gold Standard, paralleling VM0085 and ISO 18752 for sustainable aviation fuels.

11 Conclusion

The protocol reframes food traceability as a problem of verifiable state transitions between typed digital objects. The contribution of this paper is the Tiered Attestation Model, a structural decomposition of the evidence supply problem into four progressively rigorous tiers, organised by source of evidence, marginal cost, and probative value, and additive in their detection capabilities. A deployment selects the depth of evidence its use case requires; the protocol is identical at every depth.

The protocol guarantees structural integrity of the audit trail under explicit network assumptions and supports composable verification verdicts. It does not guarantee the truthfulness of physical measurements nor protect against fully coordinated collusion across all data sources. We sketch in Section 10.2 two pathways (Proplet compliant hardware and cryptoeconomic mechanism design) that close the residual risk to a calibrated economic parameter. What the protocol offers, in steady state, is reproducibility: any verifier with access to the public record and the relevant reference materials can independently retrace the chain of evidence and arrive at the same verdict.

The bioasset carries its history but cannot speak. The Bio-Digital Twin gives it a voice that anyone can verify. Traceability built on evidence, not trust.

A Mathematical Formalisation

A.1 Canonical Object

Let O denote a BDT aggregating its core attestation vectors and any external attestations:

$$O = (\vec{A}, \vec{L}, \vec{P}, \vec{R}, \mathcal{E}, \tau_O).$$

Deterministic canonical serialisation under beeson yields a unique identifier:

$$H_O = \mathcal{H}_{\text{BMT}}(\text{Canonical}(O)).$$

H_O uniquely identifies the object and is the basis for content addressed retrieval and verification.

A.2 Attestation Vector

Each core vector $\vec{V} \in \{\vec{A}, \vec{L}, \vec{P}, \vec{R}\}$ is a tuple

$$\vec{V} = (a, \tau, \vec{x}, t, \sigma)$$

with attestation party a , schema τ , payload $\vec{x}$, temporal anchor t , and signature

$$\sigma = \text{Sign}_a(\mathcal{H}_{\text{BMT}}(\vec{x}, \tau, t)).$$

External Attestations $\vec{E} \in \mathcal{E}$ follow the same shape with an additional reference r_e to the BDT they attest.

A.3 Ownership Chain

Let the ownership sequence of object O be

$$(A_0, A_1, \dots, A_n)$$

with signatures

$$\sigma_i = \text{Sign}_{A_i}(H_O \parallel A_{i+1}).$$

The chain is valid if each σ_i verifies against A_i and no conflicting signature exists from the same A_i . Conflicting signatures constitute publicly provable double signing.

A.4 Burn and Distribution Conditions

An object is *consumed* (burned) if its ownership chain terminates at a designated transformation address:

$$A_n = \text{TransformAddr}.$$

No further valid transfers of the consumed object are accepted. A consumed object may participate in two kinds of derivation.

Transformation. The object is referenced as a burned input by exactly one TransformationEvent that produces an output object cryptographically linked to it. Multiple input objects may converge into a single output (fan-in). Each input is referenced by at most one output.

Distribution. A finalised object O_p is consumed by a single atomic distribution event that instantiates a set of $M \geq 1$ output package objects $\{O_{p,1}, \dots, O_{p,M}\}$ and an optional remainder twin R , each cryptographically linked to O_p and instantiated through a fork of the parent's content addressed feed. The schema of O_p must declare distribution as admissible. The event preserves a quantitative conservation invariant over the metadata declared in the schema:

$$\sum_{k=1}^M q(O_{p,k}) + q(R) = q(O_p),$$

with $q(R) = 0$ when no remainder is produced. The remainder R is itself a finalised object eligible to undergo its own distribution event later, producing a chain of distributions until the lineage is exhausted. Each output object remains a unique non fungible token; conservation operates on metadata, not on token fungibility.

A.5 Provenance Graph

Objects form a directed acyclic graph $G = (V, E)$ where vertices are BDTs and edges encode derivation relations of two kinds. *Transformation edges* $O_i \rightarrow O_j$ indicate that O_j references H_{O_i} as a burned input. *Distribution edges* $O_p \rightarrow O_{p,k}$ indicate that $O_{p,k}$ is a package derivative of the finalised parent O_p .

Structural validity requires:

- All referenced inputs and parents exist in the public record.
- All referenced inputs (transformation) are burned and all referenced parents (distribution) are finalised.
- For transformation, no input is referenced by more than one output.
- For distribution, the parent is consumed by exactly one distribution event whose outputs (packages plus optional remainder) conserve the parent’s declared metadata quantity. The remainder, if produced, is itself a finalised object that may undergo its own distribution event.

This defines structural consistency independent of runtime contract validation. Detection at T_1 (Section 5.1) is precisely the predicate set induced by the validity conditions above.

B Feature Inventory

This appendix lists illustrative variables used to instrument the four core attestation vectors in a reference deployment. The lists are not exhaustive: production deployments may extend or restrict them based on the matrix, the regulatory regime, and the available sensor inventory.

B.1 Agronomic Vector

Sensor readings are captured at field IoT gateways and embedded as SenML records inside the canonical payload of $\vec{x}_A$. Representative variables include:

- Soil water potential (Ψ_s). Hydraulic stress indicator. Affects cellular structure and downstream sublimation behaviour.
- Bulk conductivity (σ_b). Proxy for soil mineral density and nutrient uptake. Influences the biochemical profile of the fruit.
- Evapotranspiration (ET_0). Metabolic activity indicator in the days leading up to harvest.
- Solar radiation (I_s) and ambient temperature (T_{air}). Energy inputs affecting sugar content, moisture levels, and tissue density.

Sensor data is queried from the IoT API over a time window before being aggregated and signed. An example query payload:

```
POST /api/Datapoints/getDatapoints
```

```
-
```

```
"fromTime": "2026-01-10T00:00:00Z",
"toTime":    "2026-01-12T23:59:59Z",
"type":      "soilwaterpotential"
```

Figure 8: IoT API query for an Agronomic Vector component over a defined time window.

B.2 Logistic Vector

- Geolocation (L_{gps}). High precision coordinates at harvest and at plant intake, establishing geographic provenance.
- Harvest timestamp (t_h). When the lot was harvested.
- Logistics latency (Δt_f). Time elapsed from harvest to plant intake. Key degradation variable for T_2 cross vector coherence tests.
- Custody metadata (μ_{log}). Transportation conditions, handling events, cold chain readings where available.

B.3 Process Vector

- Chamber vacuum pressure (P_{vac}). Must remain stable during primary drying for proper sublimation.
- Sublimation gradient (Γ_{sub}). Rate of temperature change versus heat input during the drying phases.
- Energy density (E_ρ). Kilowatt hours consumed per kilogram of water removed. Varies with the fruit's initial moisture and cellular structure.
- Phase duration (Δt). Time to reach critical transition points (triple point, secondary drying completion).
- Final water activity (a_w). Residual moisture in the finished product. Determines shelf stability.

B.4 Recipe Vector

The Recipe Vector encodes the operator selectable parameters of the cycle: programmed shelf temperature schedule, target chamber pressure, freezing protocol, annealing time, tray geometry, particle size, chamber load, tray distribution, and equipment calibration metadata. The recipe payload is signed by the plant operator before the cycle begins. The detection capability of T_2 (Section 5.2) compares the committed recipe against the observed Process Vector to flag execution drift.

A future extension of the model may introduce an explicit operational vector $\vec{O}$ that decomposes the recipe into deterministic and stochastic components, tightening the predicted envelopes used by the T_2 detectors as more pilot data becomes available.

Document History

Versions follow semantic versioning: a major increment changes protocol semantics or normative requirements, a minor increment adds material without breaking existing conformance, a patch corrects errors. The current version is stated on the title page and in the PDF metadata.

Version	Date	Changes
0.1, 0.2	February 2026	Preliminary single-document drafts.
0.3	May 2026	Rewritten as the entry document of a four-document corpus; methodology, security analysis and standards stack split into companion documents.
draft	June 2026	Revised draft; reference copy published on Swarm.
1.0	24 September 2026	First public release. Version and date fixed, licensed under CC BY 4.0, document history added. No technical changes from the June draft.

References

- Fair Food Data. BDF methodology for recovered food carbon. 2026a. Companion paper, version 1.0, September 2026. <https://fairfooddata.org>.
- Fair Food Data. BDF security analysis: Threat model, trust assumptions, and formal reductions. 2026b. Companion paper, version 1.0, September 2026. <https://fairfooddata.org>.
- Fair Food Data. BDF standards stack: Implementation reference. 2026c. Companion paper, version 1.0, September 2026. <https://fairfooddata.org>.
- D. M. Murphy and T. Koop. Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarterly Journal of the Royal Meteorological Society*, 131(608):1539–1565, 2005.
- M. J. Pikal and S. Shah. The collapse temperature in freeze drying: Dependence on measurement methodology and rate of water removal from the glassy phase. *International Journal of Pharmaceutics*, 62(2–3):165–186, 1990.
- Yrjö H. Roos. *Phase Transitions in Foods*. Academic Press, San Diego, 1995.
- Nick Szabo. Proplets: Devices for controlling property. Online, 2001. URL <https://nakamotoinstitute.org/library/proplets/>. Accessed via the Satoshi Nakamoto Institute Library.
- Viktor Trón. *The Book of Swarm: Storage and Communication Infrastructure for a Self-Sovereign Digital Society*. Self-published, 2024. ISBN 978-615-01-9983-2. Available at <https://www.ethswarm.org/the-book-of-swarm.pdf>.
- Viktor Trón and Henning Diedrich. A dream come true: Deletable content in immutable storage. *ACM (manuscript)*, 2025. Manuscript submitted to ACM, September 2025.
- Viktor Tron et al. Beeson: A canonical object notation for content-addressed data. 2026. Forthcoming.