BONAFIDE™ SPECIFICATION V1.0 — PART 4

PII Protection & Privacy

Proxy identity, privacy scoring, canary detection, forensic watermarking, data minimization, right to erasure, data portability, and privacy-by-design regulatory alignment

1. Purpose

This part defines how Bonafide protects personally identifiable information (PII) throughout its lifecycle: from collection through storage, access, sharing, and deletion. It specifies the proxy identity system, the privacy scoring mechanism, canary detection for unauthorized data use, forensic watermarking, and the data minimization framework that makes ghost quanta the default for institutional interactions.

2. Proxy Identity

2.1 The Problem with Direct Identifiers

Every time a user provides their real email address, phone number, or mailing address to an institution, they create a correlation point. An attacker, data broker, or the institution itself can use these identifiers to link the user's activity across services, build profiles, and track behavior. Even within Bonafide's encrypted vault architecture, if the user gives Chase and Mount Sinai the same email address, those institutions can correlate the user outside the vault.

2.2 Proxy Addresses

Bonafide's relay network (Part 8) provides proxy addresses for every communication channel:

Channel	Proxy Format	Example	Routing
Email	opaque@relay.bonafide.id	a8f3x@relay.bonafide.id	Relay forwards to user's real email or vault inbox
Phone	Relay-assigned number	+1-555-BF-XXXXX	Relay forwards calls/SMS to user's real phone
Mailing address	Relay-operated PO box or forwarding	Bonafide Relay, Box 8F3X, City, State	Relay forwards physical mail

Each institutional peer receives a unique proxy address. Chase gets one email proxy; Mount Sinai gets a different one. No two institutions receive the same proxy. The relay operator routes inbound messages to the user's real address (or vault inbox) without revealing it to the sender.

2.3 Proxy Rotation

Proxy addresses can be rotated at any time. If a proxy is compromised (spam, data broker acquisition, breach), the user generates a new proxy for that institution and the old one is deactivated. Inbound messages to the old proxy are rejected or optionally held in a quarantine queue for review.

2.4 Unlinkability

Proxy addresses are cryptographically unlinkable. There is no derivation relationship between the proxies assigned to different institutions. An attacker who obtains proxy addresses from both Chase and Mount Sinai cannot determine that they belong to the same user. The relay operator knows the mapping (proxy → real address) but the relay operator does not know which institution the proxy was assigned to—the relay sees only inbound messages to an opaque address.

3. Privacy Scoring

3.1 Purpose

Every institution in the Bonafide network has a privacy score: a numeric assessment (0–100) of how well the institution protects user data. The score is visible to users during peer authorization, in vault management, and through the public ecosystem directory. It is the primary mechanism for market-driven privacy improvement—institutions with low scores lose users to competitors with higher scores.

3.2 Score Components

Component	Weight	Source	What It Measures
Privacy classification	40%	Hardware attestation + audit (Part 11)	Structural protection level (S/A/B/C)
Access pattern behavior	20%	Ledger analysis	Ghost quanta usage rate, plaintext request frequency, access pattern consistency with operational need
Revocation compliance	15%	Ledger + runtime telemetry	Speed and completeness of revocation processing when users revoke access
Override frequency	10%	Ledger analysis	How often the institution uses quantum-level overrides vs. standard access
Breach history	10%	Public records + canary detection	Number and severity of data breaches, canary appearances outside authorized context
Certification status	5%	Certification authority	Current certification tier and audit results

3.3 Score Baseline from Classification

The privacy classification provides the structural baseline: Classification S starts at 90/100, A at 75, B at 55, C at 35, U at 0. Behavioral components adjust the score up or down from this baseline. A Classification A institution with perfect behavior might score 95. A Classification A institution with poor operational history might score 60.

3.4 Score Computation

Scores are computed from ledger data and attestation records—not self-reported. Any Bonafide participant can independently verify an institution’s score by examining the same ledger data. The computation algorithm is published as part of the specification to ensure transparency and reproducibility.

3.5 Score Display

During peer authorization, the user sees:

- The institution’s numeric score (0–100)

- The letter classification (S, A, B, C, U)
- A plain-language summary (“This institution uses hardware-encrypted storage and processes most requests without seeing your raw data”)
- Comparison to the user’s other peers (“Higher than 4 of your 6 current peers”)
- Trend indicator (score improving, stable, or declining over the past 6 months)

4. Canary Detection

4.1 What Canaries Are

Canary quanta are synthetic data elements embedded within real data releases, unique to each institutional peer. The institution cannot distinguish canaries from real data—they carry the same encryption, the same metadata format, and the same access characteristics. If a canary appears outside the authorized context (in a data broker’s database, on a dark web marketplace, in an unauthorized third party’s system), it identifies the source institution with certainty.

4.2 Canary Types

- **Identity canaries:** Synthetic names, addresses, or phone numbers that are unique per institution but plausible in format. If “John M. Canfield at 742 Evergreen Terrace” appears in a leaked dataset, and that specific combination was embedded only in Chase’s branch, the leak came from Chase.
- **Financial canaries:** Synthetic transaction records with unique amounts or merchant names. A $47.23 charge at “Verde Coffee” that exists only in the insurance company’s view of the user’s financial data.
- **Medical canaries:** Synthetic but plausible medical entries (a minor diagnosis code, a supplement prescription) unique to each healthcare peer.
- **Document canaries:** Invisible modifications to shared documents—whitespace variations, Unicode homoglyphs, or steganographic markers that are imperceptible to humans but uniquely identify the recipient.

4.3 Canary Lifecycle

- Canaries are generated automatically by the Bonafide runtime when data is shared with an institution.
- The user’s vault distinguishes canaries from real data. The institution’s view does not.
- Canaries age and evolve alongside real data to remain plausible over time.
- Canary detection is passive: the Bonafide network periodically scans public data sources for canary signatures. Active detection can be triggered by user report or automated monitoring services.
- When a canary is detected, the event is recorded in the ledger, the institution’s privacy score is impacted, and the user is notified.

5. Forensic Watermarking

5.1 Purpose

Every time plaintext data is rendered or transmitted from the Bonafide runtime, it carries an invisible forensic watermark. The watermark is session-unique: it encodes the institution, the session ID, the timestamp, the accessing employee (if available), and the rendering context. If the data appears outside the authorized context, the watermark traces it to the specific access event.

5.2 Watermark Properties

- **Invisible:** The watermark is imperceptible to human viewers. Text watermarks use Unicode variation selectors, zero-width characters, and whitespace patterns. Image watermarks use steganographic embedding in the spatial or frequency domain. Document watermarks use formatting micro-variations.
- **Robust:** The watermark survives common transformations: screenshot, print-scan, copy-paste, format conversion, compression. Different watermark techniques are layered so that even if one is stripped, others survive.
- **Unique:** Each access event produces a different watermark. Two accesses to the same quantum by the same institution produce distinct watermarks that trace to their respective sessions.
- **Non-repudiable:** The watermark's content is signed with the runtime's attestation key. The institution cannot claim the watermark was fabricated.

5.3 Watermark in Share Links

When data is shared through a share link (Part 12, Section 5), the Bonafide viewer applies a viewer-specific watermark in addition to the session watermark. The viewer watermark encodes the share link ID and the viewing session. If a share link recipient screenshots the data and redistributes it, the watermark traces back to the specific share link and viewing session.

6. Data Minimization

6.1 Ghost-First Principle

The specification's default posture is that institutions should receive ghost quanta (zero-knowledge proofs, redacted views, hashed confirmations) rather than raw data. Plaintext release is the exception, justified only when the institution's core function genuinely requires it.

The ghost-first principle is enforced through multiple mechanisms:

- **Default access policies:** When a user peers with an institution, the default access policy grants ghost-level access. The institution must specifically request plaintext access for operations that require it, and the user must approve.
- **Classification incentive:** Ghost quanta usage rate is a component of the privacy score. Institutions that maximize ghost usage improve their score. Institutions that request plaintext when ghost would suffice are penalized.
- **Verification rules:** Institution-to-institution verifications (Part 12, Section 6) use ghost quanta exclusively by default. The user must explicitly override this for plaintext transfer.

6.2 Minimum Necessary Exposure

When plaintext is required, the specification requires that the institution access only the minimum data necessary for the operation:

- A credit check accesses the credit score quantum, not the full financial branch.
- An age verification accesses a ghost quantum proving age ≥ threshold, not the date of birth.
- A medical referral transfers specific quanta (the relevant diagnosis and treatment history), not the entire medical branch.
- An identity verification uses ghost quanta for name, address, and ID number confirmation without revealing the actual values.

6.3 Exposure Logging

Every plaintext exposure is logged in the ledger with the quantum ID, the accessing entity, the justification (which operation required plaintext), and the scope. Users can review their exposure history and identify institutions that request more plaintext than their service requires.

7. Right to Erasure

7.1 GDPR Article 17 Compliance

Users can exercise right-to-erasure at any granularity: per-quantum, per-channel, per-branch, or entire institutional relationship.

7.2 Erasure Mechanism

- User initiates erasure from any vault holder device.
- The target quanta’s DEKs are cryptographically destroyed (all wrappings—user, institutional, and third-party—are deleted).
- The encrypted payload becomes permanently unreadable. Even if the ciphertext persists on institutional storage, no key exists to decrypt it.
- The quantum’s Merkle leaf is replaced with a tombstone marker (Part 2, Section 7.3). The tree’s integrity is preserved.
- The erasure event is recorded in the ledger: timestamp, scope, confirmation of DEK destruction.
- The institution’s Bonafide runtime confirms erasure completion.

7.3 Erasure Limitations

Erasure destroys the keys. It does not retroactively delete plaintext that the institution may have copied outside the Bonafide layer during prior authorized access. This is the same limitation that applies to any erasure request under current law—GDPR requires reasonable efforts, not time travel. The anti-duplication framework (Part 12, Section 10) and canary detection provide traceability for unauthorized copies.

7.4 Erasure and Third-Party Access

If an active third-party authorization (legal access, share link) exists for a quantum targeted for erasure, the erasure request is held until the authorization expires or is revoked. A user cannot destroy evidence that is subject to an active court order. The hold is recorded in the ledger. Once the authorization ends, the erasure executes automatically.

8. Data Portability

8.1 GDPR Article 20 Compliance

Users can export their vault data at any time in a portable, machine-readable format. The export includes the encrypted quanta, the user’s key wrappings (enabling decryption with the user’s Bio Root on any compliant implementation), the Merkle tree for integrity verification, and the ledger history for audit trail continuity.

8.2 Export Format

The export is a self-contained vault file that can be imported into any Bonafide-compliant implementation. The user’s Bio Root decrypts the exported data on the new platform. Institutional wrappings are not included in the export—they are specific to the institutional relationship and are re-established during re-peering.

8.3 Migration Path

- User exports vault from Provider A.
- User imports vault into Provider B (different vault provider, different infrastructure).
- User authenticates biometrically on Provider B’s infrastructure.
- The Bio Root decrypts the exported data. The vault is operational on Provider B.
- User re-peers with institutions from Provider B. New institutional wrappings are created.
- Provider A’s copy can be erased.

9. Privacy by Design Summary

Bonafide's privacy architecture is not a feature bolted onto a data management system. It is the data management system. Every design decision is a privacy decision:

Principle	Implementation	Regulatory Alignment
Data minimization	Ghost-first principle, minimum necessary exposure, ZK proofs for verification	GDPR Art. 5(1)(c)
Purpose limitation	Quantum access policies encode permitted use, ledger documents deviations	GDPR Art. 5(1)(b)
Storage limitation	User-controlled retention, right to erasure with cryptographic destruction	GDPR Art. 5(1)(e), Art. 17
Integrity and confidentiality	AES-256-GCM per quantum, Merkle integrity, forensic watermarking	GDPR Art. 5(1)(f)
Data protection by design	Encryption is the architecture, not a feature. Proxy identity. Blind validation.	GDPR Art. 25
Accountability	Immutable ledger, privacy scoring, canary detection, classification audits	GDPR Art. 5(2)
Data portability	Self-contained vault export, Bio Root-decryptable on any compliant platform	GDPR Art. 20
Transparency	Privacy classification visible to users, scores verifiable from ledger data	GDPR Art. 12–14

Bonafide™ — Privacy by architecture, not by promise.

An open specification by Sly Technologies Inc. | bonafide.id | bonafideid.org
V1.0 — February 2026