Technical and Organizational Measures (TOM)

pursuant to Art. 32 GDPR for the PII Pseudonymization Service

Last updated: 06.11.2026 · Version 1.6

Overview

This document describes the technical and organizational measures (TOM) pursuant to Art. 32 GDPR that the PII pseudonymization service by Michael SCHILLER - Organisation. Digital. has implemented to ensure a level of protection appropriate to the risk.

1. Confidentiality (Art. 32(1)(b) GDPR)

1.1 Physical Access Control

Protection against unauthorized access to data processing facilities

- Rented servers of dataforest GmbH in Frankfurt am Main
- Physical access protection by the data center operator
- No physical access by Processor employees required (fully remotely managed)

1.2 System Access Control

Protection against unauthorized use of data processing systems

- SSH access only with a key, password login switched off
- No root access, only dedicated user accounts
- Rate limiting per IP address: 60 requests/minute for pseudonymizing, 30 for restoring, 10 for schema requests
- Fail2Ban for SSH brute force protection

1.3 Data Access Control

Protection against unauthorized reading, copying, modification, or removal

- API endpoints are publicly accessible (no authentication required) – this is intentional as no personal data is permanently stored
- Mapping tables can only be fetched with the session ID. Without an API key the service issues and signs the session ID and accepts no other, with an API key it is bound to that key
- No administration interface for PII data
- No ability to retrospectively view processed texts, except for calls with a key released for the nightly check (see 2.1)

1.4 Separation Control

Separate processing for different purposes

- Without an API key every call gets its own signed session ID

- Session data is completely isolated from each other
- No shared use of mapping tables between sessions
- The staging environment runs on the same server as production, in its own containers and with its own database on the same database server. Both environments use the same Redis instance

2. Integrity (Art. 32(1)(b) GDPR)

2.1 Transfer Control

Protection during data transmission

- All API communication encrypted via HTTPS with TLS 1.2 or 1.3
- Caddy reverse proxy with automatic certificate renewal (Let's Encrypt)
- No unencrypted transmission channels
- No data transfer to third countries
- Sub-processors are the server provider dataforest GmbH and, for the nightly check only, IONOS SE (see DPA § 5)
- Nightly check: only calls with a key released by the operator. Their pseudonymized text is stored encrypted for up to 48 hours (up to 30 days after a reported finding) and checked on rented GPU servers in the EU or EEA. Calls without a key or with a key that is not released are excluded

2.2 Input Control

Traceability of data processing

- Log without plain text: time, a hashed session ID, the number of detected values and checksums of them made with a secret key (HMAC-SHA256) are logged. Retention: calls without a released key 7 days, otherwise by check result 30 days (clean), 90 days (errors, false alarms) or 3 years (missed values); operational entries 90 days, deletion records 3 years
- The application log replaces IP addresses with a placeholder before writing. The web server keeps no access log for the API. For rate limiting the IP address is held briefly as a counter
- Mapping tables are automatically deleted 30 minutes after their last use, at the latest 8 hours after their first use

3. Availability and Resilience (Art. 32(1)(b) GDPR)

3.1 Availability Control

Protection against destruction or loss

- Automatic container restarts on failures (Kamal/Docker)
- Monitoring with Grafana dashboard
- Daily database backup on the server itself, unencrypted. 14 daily, 8 weekly and 12 monthly copies are kept
- Original texts are not stored permanently. Checksums of recognised values and hashed session ids remain in the log for up to 3 years (see 2.2)

- In case of system failure, at most the temporary mapping tables are lost

4. Recoverability (Art. 32(1)(c) GDPR)

Since no original texts are permanently stored, recovery is limited to system availability:

- Container-based deployment architecture (Kamal) enables re-provisioning
- Automatic health checks and restart mechanisms
- Restoring the database from the latest backup was tested successfully on 17.10.2026

5. Regular Review Procedures (Art. 32(1)(d) GDPR)

5.1 Data Protection Management

- Regular review of technical measures
- Before every release an automatic check of the Ruby libraries (bundler-audit) and JavaScript packages (npm security database) in use for known security vulnerabilities; in addition an hourly check of the running systems for known vulnerabilities, every finding is filed as a task

5.2 Privacy by Design & Default

- Original texts are not stored. The cache (Redis) runs with no copy on disk
- Mapping tables with automatic TTL: 30 minutes after their last use, at the latest 8 hours after their first use, every use restarts it
- No plain-text logs, only checksums made with a secret key
- Sessions are bound to the caller: signed session ID or API key
- Several parallel detection methods (see 6.)

6. Detection Systems

The PII pseudonymization service runs these detection methods in parallel:

Pattern Detection – Regex-based detection of structured data (email, phone, IBAN, credit cards, EU-VAT)	GLiNER NER – PII-specialized Named Entity Recognition (local, no cloud service, ONNX-optimized)
Address Detection: address patterns for 29 countries	Crypto Wallet Detection – 12 cryptocurrencies (Bitcoin, Ethereum, Litecoin, Ripple etc.) via pattern matching

Abbreviations in capital letters (for example ACME) are currently not replaced

All detection systems run locally on our rented servers in Germany. Detection transmits no data to external services. Only the nightly check (see 2.1) uses rented GPU servers.

7. Pseudonymization Modes

After detection, personal values are pseudonymized in one of four modes depending on the data type and target system:

Tagged mode – HMAC-based markers in the format [TYPE:8hex], deterministic per session. Example: [PERSON:a1b2c3d4]. Used on external surfaces (MCP, public API, logs). The receiving system clearly recognizes that a placeholder stands here.

Faker mode – Selection from German faker pools (~30,000 entries; names, addresses, companies). HMAC-indexed with deterministic linear probing, so the same input value within a session always maps to the same pseudonym. Used on internal AI pipelines (AI chat, voice, mailer). Language models handle plausible pseudonyms better than bracket markers.

Reserved-Range mode – For phone numbers and IP addresses we use officially reserved drama/documentation ranges: BNetzA reserved mobile blocks (Mitteilung 148/2021), FCC 555-0100..0199, Ofcom 07700 900xxx, RFC 5737 (IPv4 documentation), RFC 3849 (IPv6 documentation). HMAC-indexed like the Faker mode, but every pseudonym comes from a range that, by regulator allocation, cannot reach a real subscriber or network host. Advantage over Sentinel: the language model can disambiguate multiple phones in one conversation without any risk that a real number gets called.

Strict-Pool mode – For highly sensitive financial and identifier values (IBAN, credit-card numbers, BIC, German tax/social/health-insurance IDs, Swiss AHV, Dutch BSN, French INSEE, Belgian Rijksregister, Spanish DNI/NIE, German ID card, Austrian social-security number) we generate format-preserving pseudonyms from reserved test ranges (ISO-3166-non-allocated IBAN prefixes, Stripe-test PAN 4242, leading-zero IDs, T-prefix markers). Each pseudonym passes the format check but deliberately FAILS its own algorithm validator (re-detection guard) — structurally impossible to hit a real identifier. The language model sees a plausible shape but cannot trigger a real bank or government action.

Detected original values do not leave our processing boundary in plain text. Values that were not detected stay unchanged. Before delivery to end users, pseudonyms are replaced with the real values on the server.

8. Infrastructure Overview

Server

- Rented servers of dataforest GmbH in Frankfurt am Main, Germany
- Debian GNU/Linux 13 (hardened)
- Kamal (Docker-based)

Services

- Rails 8 (Ruby 4.0) – API and web interface
- GLiNER NER Service – Local AI container (PII-specialized)
- Crypto & EU-ID Pattern Matching – Integrated in Pattern Detector
- Caddy: reverse proxy with TLS 1.2 or 1.3
- Redis – Session cache with TTL

9. Contact

For questions about the technical and organizational measures:

Michael Schiller (Managing Director)
Michael SCHILLER - Organisation. Digital.
info@schiller-partners.de