

Technisch-organisatorische Maßnahmen (TOM)

gemäß Art. 32 DSGVO für den PII-Pseudonymisierungsdienst

Stand: 02.10.2026 · Version 1.3

Übersicht

Dieses Dokument beschreibt die technisch-organisatorischen Maßnahmen (TOM) gemäß Art. 32 DSGVO, die der PII-Pseudonymisierungsdienst von Michael SCHILLER - Organisation. Digital. implementiert hat, um ein dem Risiko angemessenes Schutzniveau zu gewährleisten.

1. Vertraulichkeit (Art. 32 Abs. 1 lit. b DSGVO)

1.1 Zutrittskontrolle

Schutz vor unbefugtem Zutritt zu Datenverarbeitungsanlagen

- Gemietete Server der dataforest GmbH in Rechenzentren im Raum Frankfurt am Main
- Physischer Zugangsschutz durch den Rechenzentrumsbetreiber (ISO 27001 zertifiziert)
- Kein physischer Zugang durch Mitarbeiter des Auftragsverarbeiters erforderlich (vollständig remote verwaltet)

1.2 Zugangskontrolle

Schutz vor unbefugter Nutzung der Datenverarbeitungssysteme

- SSH-Zugang ausschließlich über schlüsselbasierte Authentifizierung (ED25519)
- Kein Root-Zugang, nur dedizierte Benutzerkonten
- Anubis Bot-Schutz vor der Website. Die API-Adresse (api.schiller-partners.de) läuft ohne Anubis, weil Programme die Prüfaufgabe nicht lösen können; dort greift das Rate Limiting
- Rate Limiting: 60 Anfragen/Minute pro IP-Adresse
- Fail2Ban für SSH-Brute-Force-Schutz

1.3 Zugriffskontrolle

Schutz vor unbefugtem Lesen, Kopieren, Verändern oder Entfernen

- API-Endpunkte sind öffentlich zugänglich (keine Authentifizierung erforderlich) – dies ist beabsichtigt, da keine personenbezogenen Daten dauerhaft gespeichert werden
- Zuordnungstabellen sind nur mit der Session-ID abrufbar. Ohne API-Schlüssel vergibt und signiert der Dienst die Session-ID und nimmt keine fremde an, mit API-Schlüssel ist sie an den Schlüssel gebunden
- Keine Administrations-Oberfläche für PII-Daten

- Keine Möglichkeit, verarbeitete Texte nachträglich einzusehen

1.4 Trennungskontrolle

Getrennte Verarbeitung für unterschiedliche Zwecke

- Ohne API-Schlüssel erhält jeder Aufruf eine eigene, signierte Session-ID
- Session-Daten sind vollständig voneinander isoliert
- Keine gemeinsame Nutzung von Zuordnungstabellen zwischen Sessions
- Die Staging-Umgebung läuft auf demselben Server wie die Produktion, in eigenen Containern und mit eigener Datenbank auf demselben Datenbankserver. Beide Umgebungen nutzen dieselbe Redis-Instanz

2. Integrität (Art. 32 Abs. 1 lit. b DSGVO)

2.1 Weitergabekontrolle

Schutz bei Datenübertragung

- Verschlüsselung aller API-Kommunikation per HTTPS mit TLS 1.2 oder 1.3
- Caddy Reverse Proxy mit automatischer Zertifikatserneuerung (Let's Encrypt)
- Keine unverschlüsselten Übertragungswege
- Keine Datenübermittlung an Drittländer
- Einziger Unterauftragsverarbeiter ist der Serveranbieter dataforest GmbH (siehe AVV § 5)

2.2 Eingabekontrolle

Nachvollziehbarkeit der Datenverarbeitung

- Protokoll ohne Klartext: protokolliert werden Zeitpunkt, eine gehashte Sitzungskennung, die Zahl der erkannten Angaben und Prüfsummen mit geheimem Schlüssel (HMAC-SHA256) der erkannten Werte, 7 Tage lang
- Das Anwendungsprotokoll ersetzt IP-Adressen vor dem Schreiben durch einen Platzhalter. Der Webserver führt für die API kein Zugriffsprotokoll. Für das Rate Limiting wird die IP-Adresse kurzzeitig als Zähler gehalten
- Zuordnungstabellen werden 30 Minuten nach der letzten Nutzung, spätestens 8 Stunden nach der ersten Nutzung, automatisch gelöscht

3. Verfügbarkeit und Belastbarkeit (Art. 32 Abs. 1 lit. b DSGVO)

3.1 Verfügbarkeitskontrolle

Schutz gegen Zerstörung oder Verlust

- Automatische Container-Neustarts bei Ausfällen (Kamal/Docker)
- Monitoring mit Grafana-Dashboard
- Der Hoster dataforest sichert den Server. Zusätzlich lädt unser Wartungsassistent diese Sicherungen auf eigene Geräte in Deutschland herunter
- Keine dauerhafte Speicherung personenbezogener Daten → kein Datenverlust möglich

- Bei Systemausfall gehen höchstens die temporären Zuordnungstabellen verloren

4. Wiederherstellbarkeit (Art. 32 Abs. 1 lit. c DSGVO)

Da keine personenbezogenen Daten dauerhaft gespeichert werden, beschränkt sich die Wiederherstellung auf die Systemverfügbarkeit:

- Container-basierte Deployment-Architektur (Kamal) ermöglicht schnelle Wiederherstellung
- Automatische Health-Checks und Neustart-Mechanismen
- Disaster Recovery durch Image-basierte Deployments: Vollständige Neubereitstellung in unter 10 Minuten

5. Verfahren zur regelmäßigen Überprüfung (Art. 32 Abs. 1 lit. d DSGVO)

5.1 Datenschutz-Management

- Regelmäßige Überprüfung der technischen Maßnahmen
- Jeden Montag automatische Prüfung der eingesetzten Ruby-Bibliotheken auf bekannte Sicherheitslücken (bundler-audit); Funde werden als Aufgabe angelegt

5.2 Privacy by Design & Default

- Originaltexte werden nicht gespeichert. Der Zwischenspeicher (Redis) läuft ohne Kopie auf der Festplatte
- Zuordnungstabellen mit automatischem TTL: 30 Minuten nach der letzten Nutzung, spätestens 8 Stunden nach der ersten Nutzung, jede Nutzung startet die Frist neu
- Keine Klartext-Logs, nur Prüfsummen mit geheimem Schlüssel
- Sessions sind an den Aufrufer gebunden: signierte Session-ID oder API-Schlüssel
- 5 parallele Erkennungsmethoden für maximale Abdeckung bei minimaler Datenhaltung

6. Erkennungssysteme

Der PII-Pseudonymisierungsdienst setzt diese Erkennungsmethoden parallel ein:

Pattern Detection – Regex-basierte Erkennung strukturierter Daten (E-Mail, Telefon, IBAN, Kreditkarten, EU-VAT)

GLiNER NER – PII-spezialisierte Named Entity Recognition (lokal, kein Cloud-Service, ONNX-optimiert)

Address Detection: Adressmuster für 29 Länder

Crypto-Wallet-Erkennung – 12 Kryptowährungen (Bitcoin, Ethereum, Litecoin, Ripple u.a.) per Pattern-Matching

Abkürzungen aus Großbuchstaben
(zum Beispiel ACME) ersetzt der
Dienst derzeit nicht

Alle Erkennungssysteme laufen lokal auf unseren gemieteten Servern in Deutschland. Es werden keine Daten an externe Dienste übermittelt.

7. Pseudonymisierungs-Modi

Nach der Erkennung werden personenbezogene Werte je nach Datentyp und Zielsystem in einem von vier Modi pseudonymisiert:

Tagged-Modus – HMAC-basierte Marker im Format [TYPE:8hex], deterministisch pro Session. Beispiel: [PERSON:a1b2c3d4]. Eingesetzt an externen Schnittstellen (MCP, öffentliche API, Logs). Das empfangende System erkennt klar, dass hier ein Platzhalter steht.

Faker-Modus – Auswahl aus deutschen Faker-Pools (~30.000 Einträge, Namen, Adressen, Firmen). HMAC-indiziert mit deterministischer linearer Probing, sodass derselbe Eingabewert in einer Session immer dasselbe Pseudonym erhält. Eingesetzt an internen KI-Pipelines (AI-Chat, Voice, Mailer). Sprachmodelle verarbeiten plausible Pseudonyme besser als Bracket-Marker.

Reserved-Range-Modus – Für Telefonnummern und IP-Adressen werden offiziell reservierte Drama-/Documentation-Ranges genutzt: BNetzA-Mobilfunk-Reserve (Mitteilung 148/2021), FCC 555-0100..0199, Ofcom 07700 900xxx, RFC 5737 (IPv4-Dokumentation), RFC 3849 (IPv6-Dokumentation). HMAC-indiziert wie der Faker-Modus, aber alle Pseudonyme stammen aus Ranges, die per Behörden-Vergabe keine realen Anschlüsse oder Hosts erreichen können. Vorteil gegenüber Sentinel: das Sprachmodell kann mehrere Telefonnummern in einer Konversation unterscheiden, ohne dass eine echte Nummer angerufen werden könnte.

Strict-Pool-Modus – Für hochsensible Finanz- und Identifizier-Werte (IBAN, Kreditkartennummern, BIC, deutsche Steuer-ID, Sozialversicherungs- und Krankenversicherungsnummern, Schweizer AHV, niederländische BSN, französische INSEE, belgische Rijksregister, spanische DNI/NIE, deutscher Personalausweis, österreichische SV-Nummer) werden formaltreue Pseudonyme aus reservierten Test-Ranges erzeugt (ISO-3166-non-allocated IBAN-Präfixe, Stripe-Test-PAN 4242, Leading-Zero-IDs, T-Prefix-Marker). Jedes Pseudonym besteht den Format-Check, FAILT aber bewusst den eigenen Algorithmus-Validator (Re-Detection-Guard) — strukturell unmöglich eine reale Identifizier-Nummer zu treffen. Das Sprachmodell sieht ein plausibles Format, kann aber keine echte Behörden- oder Bank-Aktion auslösen.

Erkannte Original-Werte verlassen unsere Verarbeitungsgrenze nicht im Klartext. Nicht erkannte Werte bleiben unverändert. Vor der Auslieferung an Endnutzer werden die Pseudonyme serverseitig wieder durch die echten Werte ersetzt.

8. Infrastruktur-Übersicht

Server

- Gemietete Server der dataforest GmbH, Rechenzentren im Raum Frankfurt am Main, Deutschland
- Ubuntu Linux (gehärtet)
- Kamal (Docker-basiert)

Dienste

- Rails 8 (Ruby 4.0) – API und Web-Oberfläche
- GLiNER NER Service – Lokaler KI-Container (PII-spezialisiert)
- Crypto & EU-ID Pattern-Matching – Integriert in Pattern-Detector
- Caddy: Reverse Proxy mit TLS 1.2 oder 1.3
- Redis – Session-Cache mit TTL

9. Ansprechpartner

Bei Fragen zu den technisch-organisatorischen Maßnahmen:

Michael Schiller (Geschäftsführer)

Michael SCHILLER - Organisation. Digital.
info@schiller-partners.de