SOMA

Security & Data Processing Overview

Version 2.0 · May 2026
somaintegral.io

This document summarises SOMA's data handling, security architecture, and GDPR compliance posture for review by data protection officers, IT security teams, or legal counsel ahead of a design partner engagement.

Data residency	Frankfurt, Germany — Google Cloud europe-west3 (EU)
GDPR role	SOMA is the data processor; your organisation is the data controller
Raw files leave network?	Never — on-premise connector extracts text locally
AI inference	Anthropic Claude API (EU DPA in place) + optional EU-only fallback providers
Embeddings	Voyage AI voyage-3-lite, 512-dim — text only, EU DPA in place
DPA	Available for review and signature before any data is connected
Deletion on exit	30 days, confirmed in writing

1. DATA RESIDENCY

- All data stored and processed in **Google Cloud europe-west3 (Frankfurt, Germany)**.
- Firestore, Cloud Functions, and Cloud Run services all pinned to europe-west3.
- No customer data is written outside the EU in the standard deployment.
- Firebase Hosting (CDN) serves static assets globally — no customer data is in the CDN layer.

2. WHAT LEAVES THE CUSTOMER NETWORK

SOMA uses an **on-premise connector model**. A lightweight connector is installed inside the customer's own infrastructure (Google Workspace, Microsoft 365, or local filesystem).

- **Raw files never leave** the customer network.
- The connector extracts text locally; only extracted text is transmitted to SOMA over HTTPS.
- Binary file contents (PDF, DOCX, PPTX, XLSX) are discarded after local extraction.
- Email content is stripped of personal data (email addresses, phone numbers, patient IDs, named individuals) before any cloud write — PII never reaches Firestore.
- The connector architecture is open to customer IT inspection.

3. GDPR COMPLIANCE

- SOMA acts as a **data processor**; the customer organisation is the data controller.
- Processing is limited to the purposes defined in the DPA (knowledge base population, AI query, content generation).
- Sub-processors: **Google Cloud** (EU data residency, Frankfurt); **Anthropic** (AI inference, EU DPA available); **Voyage AI** (embedding inference, EU DPA available); **SendGrid** (transactional email delivery only — no customer knowledge base data transmitted).
- Anthropic API terms: inputs are not used to train models; data is not retained beyond the API request lifecycle.
- Voyage AI: same terms — embeddings are ephemeral within the API call; no retention.
- For enterprise customers requiring model-level data residency: **private deployment on AWS Bedrock or Azure AI (EU regions)** eliminates all non-EU inference.
- Data processing agreements available for all sub-processors on request.

4. ACCESS CONTROL

- Role-based access enforced at **database level** via Firestore security rules — not only in the UI.
- Five function roles (Scientific, Regulatory, Medical Affairs, Marketing, Sales) with separate entitlements; users may hold multiple entitled roles but activate one at a time.
- SCIM 2.0 provisioning available for Azure AD / Okta — user lifecycle managed by the customer's IdP.
- Every document defaults to **regulatory_status: internal_only** on write — a hard server-side constraint, never auto-elevated.
- Disclosure reclassification requests logged to audit trail; admin approval required.
- Prompt injection detection: all AI calls screened server-side; detected attempts logged to an append-only `injection_attempts` collection (admin-read only, no client access).

5. AI MODEL USAGE

- Primary AI: **Anthropic Claude** (`claude-sonnet-4-20250514`) via server-side Cloud Functions — API key never exposed to the client.
- Embeddings: **Voyage AI** (`voyage-3-lite`, 512-dim) — text extracted client-side, embeddings computed server-side; raw document content never sent to Voyage.
- Dual-provider AI failover: primary retries 3× with exponential backoff; configurable fallback provider (Azure AI, Mistral, local/Ollama, AWS Bedrock) retries 3× before error is surfaced.
- Per-role AI system prompt overrides configurable by tenant admin — stored in `tenant_config`; no Anthropic-level prompt sharing across tenants.
- Prompt caching enabled on system prompts (Anthropic ephemeral cache) — reduces token costs; cached prompts are not shared across tenants.
- Cost and token usage logged per call to `usage_log` collection — accessible to admin for monitoring.

6. DATA RETENTION AND EXIT

- Customer data is **logically isolated per tenant** — no data is shared between organisations.
- Documents remain in the customer's own storage (Google Drive / SharePoint / OneDrive / local). SOMA stores only extracted text and embeddings.
- Design partners may request a **full data export** at any time: all documents, metadata, generated content, and audit logs in standard formats (JSON / CSV).
- On exit or non-renewal, all customer data (extracted text, embeddings, metadata, generated content, audit logs) is **deleted within 30 days**. Deletion confirmed in writing.
- Backups containing customer data are purged within the same 30-day window.

7. SECURITY PRACTICES

- HTTPS enforced on all endpoints. No unencrypted connections.
- Firestore data encrypted at rest (Google Cloud AES-256); Cloud Functions environment encrypted at rest.
- All API keys (Anthropic, Voyage AI, SendGrid, Google OAuth) stored in Google Secret Manager — never in client code or source control.
- Audit log maintained for: document approvals, disclosure changes, admin actions, retrieval exclusions, role switches, content piece status transitions.
- XSS hardening: all user-facing strings constructed via DOM APIs, not `innerHTML` interpolation.
- Retrieval exclusion trail: every document excluded from an AI query (by disclosure level or sensitivity marker) is logged to an append-only `document_audit_log` subcollection.

Data Processing Agreement

A standard DPA is available for review before any data is connected. Design partners are encouraged to share this document and the accompanying DPA with their DPO or legal team. To request the DPA, ask a data governance question, or discuss private deployment options:

Rosa Garcia-Verdugo, PhD · rosa@somaintegral.io · somaintegral.io