Architectuur, gegevensstromen, soevereiniteit en overheidsverzoeken

Status	Geldend leveranciersdocument
Versie / uitgiftedatum	1.2 / 5 november 2026
Documenteigenaar	Summa Labs B.V., handelend onder de naam Summa
Classificatie	Publiek / due diligence

1. Soevereiniteit als controlemodel

Summa gebruikt het woord soevereiniteit uitsluitend voor aantoonbare controle over locatie, jurisdictie, toegang, afhankelijkheden en exit. EU-dataresidentie alleen is niet hetzelfde als digitale soevereiniteit. Daarom maakt dit document uitzonderingen zichtbaar in plaats van een absoluut "alles blijft in Nederland"-statement.

2. Actuele hoofdarchitectuur

Laag	Component / locatie	Vertrouwensgrens
Publieke toegang	Leafcloud LB + ingress-nginx, Nederland	TLS-ingang en routering
Webapp	Next.js, Leafcloud Nederland	Geen directe database/modeltoegang
Beleidslaag	FastAPI, Leafcloud Nederland	Authenticatie, tenantcontext, autorisatie, quota, toolbeleid
Identity	Zelfgehost ZITADEL, Leafcloud Nederland	OIDC/identiteit
Data	PostgreSQL/pgvector, Redis, private MinIO, Nederland; dagelijkse databasebackups naar Leafcloud-objectopslag	Tenantdata, files, cache
Tools	First-party workers/services, Nederland	Attachment, artefacten, Python, web broker, connectors
Modelgateway	Bifrost, Leafcloud Nederland	Geauthenticeerde interne modelroute
Actieve inference	Leafcloud, Nederland; NVIDIA A30 GPU-servers	Zelfgehoste vLLM-inference binnen de Nederlandse Leafcloud-omgeving

3. Normale chatgegevensstroom

1. Gebruiker authenticatie via ZITADEL.
2. Browser stuurt request naar Next.js/backend.
3. Backend valideert identity, tenant, room/projectrechten en gebruikslimieten.
4. Backend stelt alleen de voor de run benodigde context/toolresultaten samen.
5. Via de interne Bifrost-route wordt modelcontext naar de zelfgehoste vLLM-inference op NVIDIA A30 GPU-servers van Leafcloud in Nederland gestuurd.
6. Gegenereerde tokens keren via dezelfde route terug.
7. Het resultaat wordt volgens tenantretentie in Nederland opgeslagen.

Er is geen gewone browser-to-modelverbinding en geen automatische fallback naar een commerciële externe model-API in de beschreven route.

4. Bestanden en bedrijfskennis

Uploads worden in Nederland opgeslagen, malwaregescand en verwerkt via Tika/OCR. Grote documenten kunnen als retrieval/searchbron worden geïndexeerd in de eigen dataomgeving. Bedrijfskennis is tenantgebonden en alleen beschikbaar volgens tenant-/access policies.

Bij een externe bron wordt uitsluitend de benodigde brondata geïmporteerd of live gelezen volgens de connector. Na import valt de kopie onder Summa-retentie, tenzij de specifieke functie anders is beschreven.

5. Webresearch

Webzoekvragen gaan van backend naar de in-cluster web-retrievalbroker en vervolgens naar Linkup. Openen van een URL gebeurt bij voorkeur rechtstreeks vanuit de broker met SSRF-beveiliging; bij bepaalde publisher-statuscodes kan Linkup Fetch worden gebruikt. Webcontent is onbetrouwbare externe input en krijgt geen systeemautoriteit.

6. Meetings

Wanneer Teams is toegestaan, synchroniseert een achtergrondtaak kalender-metadata binnen een beperkt tijdvenster. Transcript, attendance en optionele chat worden niet puur door kalenderdetectie ingelezen; verwerking vereist een relevante gebruikers-/autoprocessactie en providerbevoegdheid. Een transcript kan uitspraken van externe deelnemers bevatten. Ruwe bronnen zijn standaard 90 dagen bewaard; afgeleide notities vallen onder tenantretentie.

Wanneer een gebruiker Summa Work koppelt, houdt een achtergrondtaak via Microsoft Graph-delta een kopie bij van de mail in de door de gebruiker ingeschakelde mappen en van de agenda in een beperkt venster. Bijlagen worden alleen als metadata vastgelegd. De kopie staat in Nederland en volgt de mailbox: wat in Outlook wordt verwijderd, wordt ook in Summa verwijderd, en een uitgezette map of verbroken koppeling verwijdert de kopie. Afgeleide inzichten (personen, onderwerpen, toezeggingen) worden door de zelfgehoste inference op Leafcloud gemaakt. Concepten en verzending gaan alleen via Microsoft na een expliciete actie van de gebruiker en een aparte toestemming (Mail.ReadWrite, Mail.Send).

7. Persoonlijk geheugen en samenwerken

Persoonlijk geheugen is een afzonderlijke duurzame gegevensklasse, versleuteld en bedoeld voor de betreffende gebruiker; het is geen algemene beheerderkennisbron. Gedeelde rooms kunnen deelnemers uit andere tenants bevatten via expliciete membership/invitation; de hosttenant blijft gegevenshouder van de room en guesttoegang wordt expliciet gere-scoped na membershipcontrole.

8. Locatie versus rechtsmacht

Aspect	Huidige positie
Kernopslag	Nederland
Identity	Nederland
Kernapplicatie	Nederland
Actieve inference	Nederland (Leafcloud, NVIDIA A30); nergens anders
Transactionele mail	Mailjet, verwerking binnen de EU
Betaling	Stripe: EER, VS en andere Stripe-locaties onder adequaatheidsbesluiten en/of standaardcontractbepalingen
Websearch	Linkup Frankrijk/EU voor search; websites kunnen wereldwijd staan
Klantgekozen connectors	Google, Microsoft, Salesforce en andere door de gebruiker gekozen diensten: geen subverwerkers, eigen voorwaarden, mogelijk buiten de EER

Daarom claimt Summa op uitgiftedatum geen volledig Nederlands dataverwerkingspad en geen absolute EU-only verwerking voor alle ondersteunende functies.

9. Buitenlandse overheidsverzoeken

Voor klantgegevens geldt het volgende interne beleid:

- toegang wordt niet vrijwillig verstrekt zonder geldige rechtsgrond;
- identiteit/bevoegdheid van de verzoekende instantie en territoriale/juridische basis worden gecontroleerd;
- reikwijdte wordt geminimaliseerd tot wat juridisch verplicht is;
- ongeldige of disproportionele verzoeken worden waar redelijk/juridisch mogelijk bestreden;
- klant wordt geïnformeerd tenzij wet dit verbiedt;
- verzoeken en besluitvorming worden geregistreerd;
- contractuele/providermaatregelen worden gebruikt om conflict met EU-recht te beperken.

Dit ondersteunt tevens de transparantie over internationale toegang die voor data-processing services onder de Data Act relevant is.

10. Striktere gemeentelijke deploymentprofielen

Een gemeente kan een schriftelijk restrictieprofiel overeenkomen, bijvoorbeeld:

- alleen EER-subprocessors in het klantcontentpad;
- transactionele mails zonder inhoud;
- connectorallowlist per tenant;
- Teams/chatcapture uit;
- webresearch uit;

- geen consumer cross-tenant sharing;
- specifieke retentie/legal hold;
- SSO/MFA verplicht.

Een dergelijk profiel is pas een contractuele garantie wanneer de technische configuratie aantoonbaar overeenkomt en in de nadere overeenkomst is opgenomen.