

Open Anonymizer

<https://github.com/ConductionNL/OpenAnonymiser/>

Mark Westerweel – markwesterweel@conduction.nl

Razo van Berkel – 06 20263365 - razo.van.berkel@centric.eu

Functionaliteit today:

- Basis anonimisering goed op orde. (meer een check)
 - nu moeten we entities meegeven. Liefste standaard geen en alles analyseren/anonymiseren?
 - Plannen opzetten voor GitHub issues en community (Mark) -> eigen organisatie?
 - Docker container en Helm deployment in cluster (Mark)
 - Database voor metadata + auditlogging
- Razo datamodel specialist?

API architectuur

BASE URL: `/api/v1/`

`/documents/`

Interne datastructuur

Encryptie met AES – GCM; <https://www.pycryptodome.org/src/cipher/modern#gcm-mode>

Idee Haarlem:

Toelichting

- `Data-encrypted-value` : de versleutelde PII (AES-GCM base64)
- `Data-key-id`: Document- of GDPR-niveau-specifieke sleutel
- `Data-fingerprint`: Hash van gebruikte sleutel voor audit
- `Content`: RDFa-style machinewaarde
- `[PERSOON]`: placeholder voor gebruiker zonder rechten

Probleem: AES-GCM is Hazmat in de cryptography lib.

<https://cryptography.io/en/latest/hazmat/primitives/aead/>

Database model:

Table DocumentsUploaded

- ID

- Filename
- Datetime_uploaded
- Datetime_analysis_started NULLABLE
- Datetime_analysis_completed NULLABLE
- Datetime_anonymization_started NULLABLE
- Datetime_anonymization_completed NULLABLE
- Analysis_type (enum: Null,
- Anonymization_type (enum: Null, TextMask, Encryption)
- UserID
- (toekomst: splits anonymization and analysis into different tables)

Table Users (toekomst)

Table Anonymizations (toekomst)

Table Analysis (toekomst)

Roadmap:

- Meerlaagse Swiss-cheese model:

Anonymy overnemen -> k-anonimiteit, L-diversiteit en T-closeness

- uitzoeken wat we kunnen gebruiken; anonymy is tabular only, niet voor text.

-