NHR-NORD@GÖTTINGEN

Verarbeiten von Gesundheitsdaten - Regulatorische Anforderungen und technische Vorschläge @ GWDG

Hendrik Nolte **Sebastian Krey** Lars Quentin Robin Strahl Julian Kunkel
hendrik.nolte@gwdg.de

9. September 2025

hpc@gwdg.de
GWDG – Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Hintergrund

NHR-NORD@GÖTTINGEN

- EU Data Governance Act
  - Ziel wiederverwendung von (öffentlichen) Daten
    - Sowohl personen- als auch nicht personenbezogen
  - Definiert Datenmittler als neutrale Dritte
  - Definiert insbesondere eine „sichere Verarbeitungsumgebung“(Art. 5)
- EU Data Act
  - Ergänzt den DGA bzgl. Zugang und Nutzung der Daten
  - Anwendung z.B. bei IoT Geräten
    - Sichert Interoperabilität
    - Rechte von Betroffene
  - Definiert insbesondere „Datenräume“(Art. 44)

Regulatorische Aspekte ausgearbeitet von Robin Strahl

EHDS: Überblick

Ziel: Erleichterung des Zugangs zu elektronischen Gesundheitsdaten für die Primär- und Sekundärnutzung

- Ist die erste Instanziierung eines Datenraumes
- präzisiert und ergänzt die DSGVO [...] in Bezug [auf] [...] elektronische Gesundheitsdaten (Art. 1 Abs. 2 lit. a EHDS)
- Auf der Grundlage sollen technische und organisatorische Standards entstehen
  - Systeme interoperabler elektronischer Gesundheits- bzw. Patientenakten (*Electronic Health Records*; EHR bzw. EHR-Systeme)
  - MyHealth@EU
  - HealthData@EU (und entsprechende Zugangsstellen)

EHDS: Nutzungsarten

- Primärnutzung (Art. 3 bis 24 EHDS)
  - Verarbeitung von Gesundheitsdaten für die Gesundheitsversorgung
- Sekundärnutzung (Art. 50 bis 81 EHDS)
  - Verarbeitung der Gesundheitsdaten für in VO genannte Zwecke
  - u.a. wissenschaftliche Forschung im Bereich des Gesundheitswesens oder des Pflegesektors (Art. 53 Abs. 1 lit. e EHDS)
  - erfolgt in sicheren Verarbeitungsräumen

EHDS: elektronische Gesundheitsdaten

- personenbezogene elektronische Gesundheitsdaten
  - [elektronische] Gesundheitsdaten, sowie genetische Daten
  - Art. 6 DSGVO (Rechtmäßigkeit der Verarbeitung)
  - Art. 9 DSGVO (Verarbeitung besonderer Kategorien personenbezogener Daten)

- nicht personenbezogene elektronische Gesundheitsdaten
  - anonymisierte Gesundheitsdaten (kein Personenbezug)
  - Daten, die sich nie auf eine Person bezogen haben

EHDS: Rollen

Zugangsstelle für Gesundheitsdaten (Art. 55 EHDS)

- werden vom Mitgliedsstaat benannt
- entscheidet über Zugang zu Daten

Aufgaben und Pflichten in Art. 57 ff. EHDS:

- Verarbeiten der Gesundheitsdaten
  (im Rahmen der Aufgaben und Pflichten der EHDS VO)
- Bereitstellen der Gesundheitsdaten in sicherer Verarbeitungsumgebung

EHDS: Rollen

Gesundheitsdateninhaber (Art. 2 Abs. 2 lit. t EHDS)

- Stellung als (gemeinsam) Verantwortlicher bei Gesundheitsdatenverarbeitung von besonderer Bedeutung (Art. 2 Abs. 2 lit. t sublit. i EHDS)
- Fähigkeit, durch Kontrolle über technische Konzeption eines Produkts/Dienstes, Daten zur Verfügung zu stellen (Art. 2 Abs. 2 lit. t sublit. ii EHDS)

Pflicht:

- Gesundheitsdateninhaber *müssen* der Zugangsstelle elektronische Gesundheitsdaten nach Art. 60, 51 Abs. 1 EHDS für Mindestkategorien bereitstellen

nat. Personen (einschließlich einzelner Forscher) und Kleinstunternehmen sind von dieser Pflicht befreit (Art. 50 Abs. 1 EHDS)

vertrauenswürdiger Gesundheitsdateninhaber (Art. 72 EHDS)

- Benennung als vertrauenswürdiger Gesundheitsdateninhaber durch Zugangsstelle für Gesundheitsdaten
- Bedingungen
  - sichere Verarbeitungsumgebung
  - Fachwissen um Anträge zu prüfen
  - erforderliche Garantien
- zusätzliche Aufgaben/Pflichten:
  - Zugangsstelle kann Anträge auf Zugang direkt weiterleiten
  - Prüfung des Antrags und Übermittlung eines Vorschlags für die Entscheidung
  - Gewährt Zugang in (eigener) sicheren Verarbeitungsumgebung

EHDS: Rollen

Gesundheitsdatennutzer (Art. 2 Abs. 2 lit. u EHDS)

- nat. o. jur. Personen, die rechtmäßig Zugang zu elektronischen Gesundheitsdaten erhalten haben

EHDS: Zugang zu Gesundheitsdaten in Sekundärnutzung

- grds. gilt Datenminimierung und Zweckbindung auch nach EHDS (Art. 66 EHDS)

- Antrag auf Zugang zu Gesundheitsdaten nach Art. 67 EHDS
- Zugangsstelle prüft, ob Kriterien für Datengenehmigung nach Art. 68 EHDS erfüllt sind
  - **JA:** Zugangsstelle gewährt Zugang zu pseudonymisierten Daten (Art. 66 Abs. 3 EHDS) in sicherer Verarbeitungsumgebung
  - **NEIN:** kein Zugang oder nur über Gesundheitsdatenanfrage

- Gesundheitsdatenanfrage (Art. 69 EHDS)
  - ausschließlich anonymisiertes statistisches Format

Rechtliche Definition von TREs

Art. 2 Nr. 20 DGA:

„sichere Verarbeitungsumgebung“ sind physische oder virtuelle Umgebung und die organisatorischen Mittel, mit denen die Einhaltung der Anforderungen des Unionsrechts, wie der [DSGVO], insbesondere im Hinblick auf die Rechte der betroffenen Personen, der Rechte des geistigen Eigentums und der geschäftlichen und statistischen Vertraulichkeit, der Integrität und der Verfügbarkeit, sowie des geltenden Unionsrechts und des nationalen Rechts gewährleistet wird und die es der Einrichtung, die die sichere Verarbeitungsumgebung bereitstellt, ermöglichen, alle Datenverarbeitungsvorgänge zu bestimmen und zu beaufsichtigen, darunter auch das Anzeigen, Speichern, Herunterladen und Exportieren von Daten und das Berechnen abgeleiteter Daten mithilfe von Rechenalgorithmen;

rechtliche Definition von TREs

Art. 2 Nr. 20 DGA:

1. physische oder virtuelle Umgebung **und** organisatorische Mittel
2. Einhaltung des Unionsrechts (inkl. DSGVO, etc.) und nationalen Rechts
3. Bestimmung und Aufsicht über alle Datenverarbeitungsvorgänge

EHDS: Definition TRE

Sicherheitsmaßnahmen (Art. 73 EHDS)

- Begrenzung des Zugangs
- Minimieren des Risikos unbefugten Verarbeitens
- Beschränkung der Eingabe
- Sicherstellung, dass Gesundheitsdatennutzer ausschließlich auf die von ihrere Datengenehmigung erfassten Daten zugreifen können
- Führung identifizierbarer Protokolle
- Sicherstellung der Befolgung und Überwachung dieser Sicherheitsmaßnahmen

Aber: Bis zum 26. März 2027 legt die Kommission technische und organisatorische Anforderungen sowie die Anforderungen an die Informationssicherheit, die Vertraulichkeit, den Datenschutz und die Interoperabilität fest.

EHDS/DSGVO: Rollen bei der Nutzung einer TRE

Art. 74 EHDS

	EHDS	DSGVO
Bereitstellung	Gesundheitsdateninhaber	Verantwortlicher
Verarbeitung (i.R.d. Aufgaben der EHDS)	Zugangsstelle	Verantwortliche
Verarbeitung in TRE	Zugangsstelle Gesundheitsdateninhaber	Auftragsverarbeiter Verantwortlicher

EHDS: deutsche Umsetzung

Die EHDS wird in DE durch das Gesundheitsdatennutzungsgesetz (GDNG) auf nat. Ebene umgesetzt und in Teilen ergänzt

- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM)
  - wird Datenzugangsstelle **und** Datenkoordinierungsstelle
- Das BfArM bekommt im wesentlichen die Aufgaben aus der EHDS zugeteilt
- zusätzlich kommt das Erstellen von Konzepten ...
  - zur Nutzung von sicheren Verarbeitungsumgebungen
  - zur Weiterentwicklung der zentralen Datenzugangs- und Koordinierungsstelle
  - zur Verknüpfung und gemeinsamen Verarbeitung von pseudonymisierten Gesundheitsdaten verschiedener datenhaltender Stellen

EHDS: Ausblick

- Viele Anforderungen müssen erst noch gemacht werden
- Stichtag 26. März 2027
  - Anforderungen zur Erfassung personenbezogener elektronischer Gesundheitsdaten (Art. 13)
  - europäisches Austauschformat für EHR (Art. 15)
  - Maßnahmen zur Entwicklung von MyHealth@EU (Art. 23)
  - Spezifikationen von EHR-Systemen (Art. 36)
  - Muster für Antrag auf Zugang zu Gesundheitsdaten (Art. 70)
  - technische und organisatorische Anforderungen an sicherer Verarbeitungsumgebungen (Art. 73)
- Vorbild sind Länder wie Finnland, die eine ähnliche Struktur bereits nutzen

TRE PoC Hintergründe

NHR-NORD@GÖTTINGEN

- Plötzliches Aufkommen der „sicheren Verarbeitungsumgebung“und EHR
- Weckt(e) Bestrebungen zu einer einheitlichen, kollaborativen Plattform
  - in Anlehung an die UK-Biobank
- Weckt(e) Verunsicherung/Neugierde bzgl. der technischen und organisatorischen Unklarheiten
  - Protoyp von GWDG mit 2 DIZen

Technische Aspekte ausgearbeitet von Lars Quentin

- Entwicklung eines eigenen TREs
- Möglichst Bare-Metal auf HPC
- Anwendungsfall: Medizinische Daten
- Vorbild: 5-Safe Framework
  - Initial UK ONS, HDR UK Green Paper (1 2)
- Zugriff: Linux VDI via Browser
- Verschlüsselung + Key Management
- Multi-User Support
- Balance: Sicherheit ggü Produktivität

5 Safe

Safe Data: Daten wurden pre-ingest pseudonymisiert.

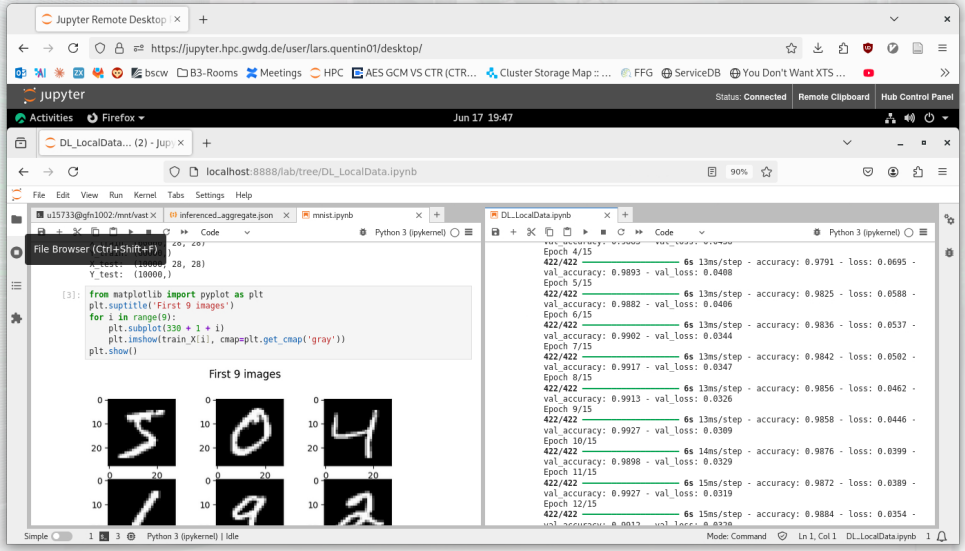
Safe Projects: Nur zugelassene Projekte dürfen Datenzugriff kriegen.

Safe People:

- KYC (Person + Institution)
- Nutzungsvertrag
- Absolviertes Training

Safe Settings: Sichere Datenumgebung

Safe Outputs: Pre-Publikation Statistical Disclosure Control (SDC).


Jupyter Remote Desktop
https://jupyter.hpc.gwdg.de/user/lars.quentin01/desktop/
jupyter
Status: Connected
Remote Clipboard
Hub Control Panel
Activities
Firefox
Jun 17 19:47
DL_LocalData... (2) - Jupy
localhost:8888/lab/tree/DL_LocalData.ipynb
File Edit View Run Kernel Tabs Settings Help
u15733@gfn1002:/mnt/vast
inferenced_aggregate.json
mnist.ipynb
DL_LocalData.ipynb
Code
Python 3 (ipykernel)
File Browser (Ctrl+Shift+F)
Y_test: (10000,)
[3]:
from matplotlib import pyplot as plt
plt.suptitle('First 9 images')
for i in range(9):
plt.subplot(330 + 1 + i)
plt.imshow(train_X[i], cmap=plt.get_cmap('gray'))
plt.show()
First 9 images
Epoch 4/15
422/422 6s 13ms/step - accuracy: 0.9791 - loss: 0.0695 -
val_accuracy: 0.9893 - val_loss: 0.0408
Epoch 5/15
422/422 6s 13ms/step - accuracy: 0.9825 - loss: 0.0588 -
val_accuracy: 0.9882 - val_loss: 0.0406
Epoch 6/15
422/422 6s 13ms/step - accuracy: 0.9836 - loss: 0.0537 -
val_accuracy: 0.9902 - val_loss: 0.0344
Epoch 7/15
422/422 6s 13ms/step - accuracy: 0.9842 - loss: 0.0502 -
val_accuracy: 0.9917 - val_loss: 0.0347
Epoch 8/15
422/422 6s 13ms/step - accuracy: 0.9856 - loss: 0.0462 -
val_accuracy: 0.9913 - val_loss: 0.0326
Epoch 9/15
422/422 6s 13ms/step - accuracy: 0.9858 - loss: 0.0446 -
val_accuracy: 0.9927 - val_loss: 0.0309
Epoch 10/15
422/422 6s 14ms/step - accuracy: 0.9876 - loss: 0.0399 -
val_accuracy: 0.9898 - val_loss: 0.0329
Epoch 11/15
422/422 6s 15ms/step - accuracy: 0.9872 - loss: 0.0389 -
val_accuracy: 0.9927 - val_loss: 0.0319
Epoch 12/15
422/422 6s 15ms/step - accuracy: 0.9884 - loss: 0.0354 -
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
DL_LocalData.ipynb

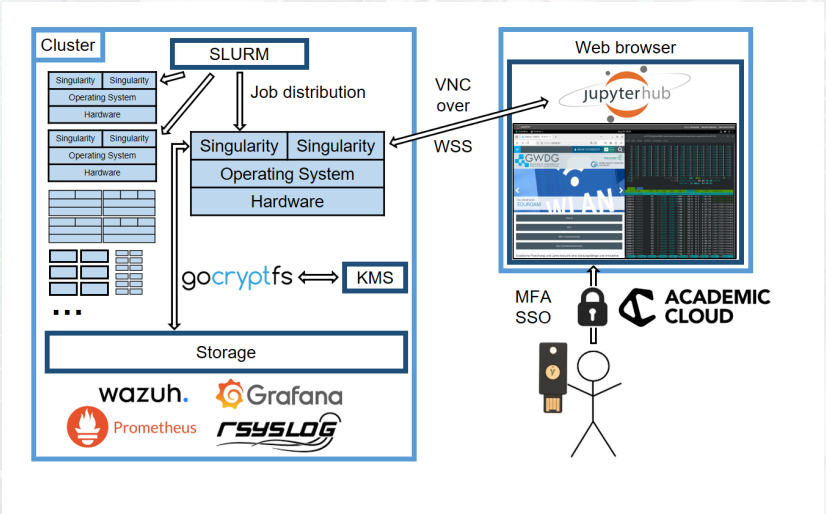
Features

- HPC-Hardware, inkl. GPUs
- Kollaboratives Arbeiten
- End-to-end Verschlüsselung mit Envelope Encryption
- Nutzbar ohne extra Software
- Terrabyte-Scale Uploader mit E2E-Verschlüsselung
- Two-Layer Permissions:
  1. POSIX
  2. Datei-Verschlüsselung
- Audit-Trail aller Dateizugriffe

Philosophie

- **Zero-Trust:** Compute Nodes können nicht als sicher angesehen werden
- **Permanente Verschlüsselung:** Daten sind von Upload zu Datenexport für Publikation nie entschlüsselt
- **Integration in existierende Prozesse:**
  - Datenanfragen über Data Owner
    - Impliziert KYC
  - Pseudonymisierung via DIZ/FDPG
  - SDC über Data Owner
- **Nutzung existierender HPC-Infra**

TRE@GWDG: Architektur

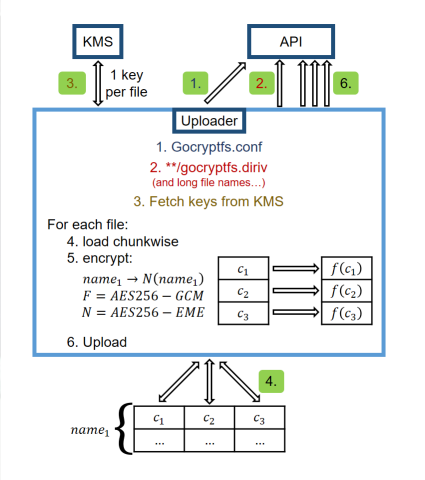


TRE@GWDG: Architektur

- JupyterHub-embedded Desktops
  - TigerVNC in Container, websockify TCP Socket ⇒ WSS, NoVNC JS Frontend
  - Somit via AcademicCloud SSO abgesichert
- VDI in Singularity/Apptainer Container, gestartet als SLURM-Job
- Internet deaktiviert, Software a-priori installiert
- Container read-only, mit Ausnahme von Gocryptfs mount.
  - Gocryptfs fork, per-file verschlüsselung via user-owned Key Encryption Key (KEK)
- Daten nicht exportierbar ohne manuelle Review
- Hauptprobleme:
  1. Sicherer Upload
  2. Sicheres Key-Management

TRE@GWDG: Upload

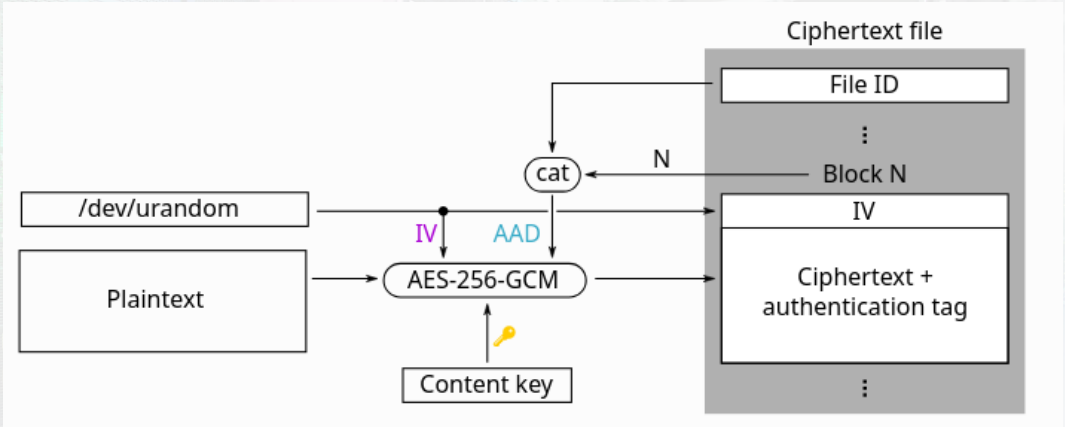
- Für TRE geschriebenen Uploader
  - Beliebig große Dateien
  - Multithreaded Upload und Verschlüsselung
  - Wiederaufnehmbar auf Dateiebene
- POSIX-Backend, keine ext. Dependencies (wie Redis, S3 etc)
- Nutzer live-verschlüsselt jeden chunk beim Upload
- Nutzt gocryptfs-kompatible Verschlüsselung
  - AES256-GCM für Inhalt
  - AES256-EME für Name



TRE@GWDG: Zero Trust

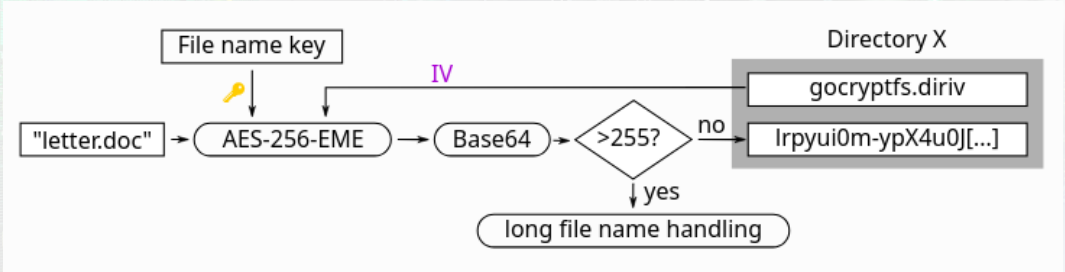
- Daten sind permanent verschlüsselt at rest.
  - Verschlüsselung für Nutzer transparent (gocryptfs)
- Jede Datei hat eigenen Schlüssel
  - Fine-Grained Encryption $\Rightarrow$ Principle of Least Privilege
- Daten können nicht durch Admins entschlüsselt werden
- Upload Node kriegt nie entschlüsselte Daten
- KMS speichert keine Keys unverschlüsselt
- Nutzergerät muss per Definition trusted sein
- Das TRE (die Containerlaufzeitumgebung) muss per Definition sicher sein

TRE@GWDG: Inhaltverschlüsselung (gocryptfs)



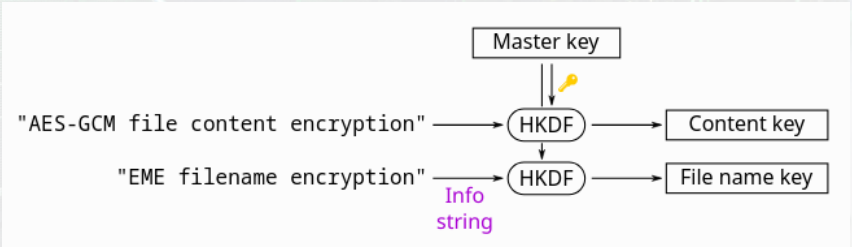
Quelle: GoCryptFS Manual

TRE@GWDG: Namensveschlüsselung (gocryptfs)



TRE@GWDG: Verschlüsselung

- Jede Datei oder Verzeichnis hat eigenen Data Encryption Key (DEK) im KMS
  - Envelope Encryption ermöglicht schnelles und sicheres Teilen der Daten
- Schlüssel-Mapping basierend auf ID (16 byte), nicht Dateipfad
  - Datei: File ID in Gocryptfs-Header
  - Verzeichnis: `./goecryptfs.dirid` Datei
- Bevor DEK genutzt wird: HKDF in File name key und file content key
  - Grund: Kryptografisch GCM/EME-Interaktion unbewiesen



TRE@GWDG: Envelope Encryption

- Data Encryption Keys (DEKs) werden verschlüsselt im KMS gespeichert
- Hauptziel: Leak von KMS-Datenbank führt nicht zu kompromittierten DEKs
- Stattdessen besitzt User einen Key Encryption Key (KEK), der DEK entschlüsselt
- Verschlüsselungsalgorithmus hat Integritätscheck (z.B. AES-GCM)
  - Somit kann KEK als Authentifizierung genutzt werden

Funktionsweise anhand von Operationen

- **Bei Mount:** Der KEK wird an goecryptfs übergeben
- **Readdir:**
  - Verzeichnisidentifier `goecryptfs.dirid` wird gelesen.
  - Anfrage des DEK basierend auf KEK + dirid
  - HKDF zum file name key
  - Readdir-Operation auf Dateisystem
  - For jeden Eintrag: Ciphernamen entschlüsseln
- **Read:**
  - Analog Verzeichnis-DEK erhalten
  - Plain-Name verschlüsseln um Ciphernamen zu erhalten
  - Aus Cipherdatei File ID extrahieren, Datei-DEK von KMS anfragen
  - Nach HKDF können chunks gelesen und AES256-GCM entschlüsselt werden

TRE@GWDG: Sicherheits- und Diasteranalyse

- **Bekannte Sicherheitslücke:** CVE-Management bereits deployed
- **Datenleck via Zwischenablage:** Deaktiviert im VDE
- **Falsche Nutzung:** Container isoliert; Internet blockiert; Training verpflichtend
- **Falsche POSIX-Rechte:** Doppelte Berechtigungsebene via KMS
- **KMS-Datenbank geleaked:** Schlüssel nicht nutzbar ohne KEKs der Nutzer.
- **Nutzer verliert KEK:** “Tresor”-Funktion basierend auf Shamir’s Secret Sharing

TRE@GWDG: Akzeptierte Sicherheitsrisiken

- **Bösartige Forscher:** Keine Möglichkeit abzusichern da Remotezugriff erlaubt
  - Beispiel: Bildschirmaufnahme mit OCR
- **Unbemerktes Teilen von KEKs:** Da Zugriffe von überall erlaubt keine Anomalieerkennung bisher geplant.
- **Zero-Day Exploits:** Angreifer kann dann aus dem Arbeitsspeicher von goecryptfs theoretisch den KEK extrahieren.

TRE@GWDG: Aktueller Stand

- Software-Entwicklung abgeschlossen
- Nächster Schritt: Deployment
- Pilotprojekt: SHIP 2 Datensatz (Universitätsmedizin Greifswald)
  - Aktueller Stand: Warten auf Datenimport
- Noch fehlende Features:
  - SLURM-Job-Support
  - Symlinks / Hardlinks

Zusammenfassung

- Neue regulatorische Anforderungen und Möglichkeiten
  - Zum Zugriff auf (Gesundheits-)daten
  - Zum Prozessieren dieser Daten
- Details werden bis Q2 2027 auf sich warten lassen
- Genügend Beispiele im europäischen Kontext um erste Erfahrungen zu sammeln
- GWDG ist dabei einen solchen PoC zu deployen
  - PoC wird in Q4 von 2 Universitätskliniken evaluiert