



UNIVERSITÄTS**medizin.**
MAINZ

DOKUMENTATION

clinical staging Data Warehouse (csDWH)



25. Februar 2022

Acronyms

| | | |
|----------------|---|---|
| csDWH | clinical staging Data Warehouse..... | 1 |
| DB | Datenbank..... | 4 |
| DIZ | Datenintegrationszentrum..... | 4 |
| ETL | Extraction Tranformation Load..... | 7 |
| GHH | Gutenberg Health Hub | |
| GTDS | Gißener Tumordokumentationssystem..... | 8 |
| IMBEI | Institut für Medizinische Biometrie; Epidemiologie und Informatik | |
| KHEntgG | Krankenhausentgeltgesetz..... | 7 |
| KIS | Krankenhausinformationssystem..... | 7 |
| LTS | Long-term support..... | 5 |
| PDMS | Patientendatenmanagementsystem..... | 8 |
| SC 6 | Servicecenter 6 IT | |
| UAC | Use & Access Committee..... | 9 |
| UM | Unimedizin..... | 9 |
| VM | Virtual Maschine..... | 6 |

Inhaltsverzeichnis

| | | |
|----------|--|----------|
| 1 | Dokumentationsinformationen | 3 |
| 1.1 | Beteiligte Personen und Status | 3 |
| 1.2 | Änderungshistorie | 3 |
| 2 | Einführung | 4 |
| 3 | Konfiguration der Instanz des csDWH | 5 |
| 4 | Sicherung des csDWH | 6 |
| 5 | Schemas | 7 |
| 5.1 | p21 | 7 |
| 5.2 | kis | 7 |
| 5.3 | copra | 8 |
| 5.4 | gtds | 8 |
| 5.5 | centrallab | 8 |
| 5.6 | imagic | 8 |
| 5.7 | metadata_repository | 8 |
| 6 | Benutzer & Zugang | 9 |

Kapitel 1

Dokumentationsinformationen

1.1 Beteiligte Personen und Status

| An Dokument beteiligte Person(en) | Status | Anmerkung zum Status |
|--|---|----------------------|
| Abel Hodelin Hernandez, SC 6, GHH [Verfasser] Sami Habib SC 6, GHH Daniel Schmitz, IMBEI | <input checked="" type="checkbox"/> in Bearbeitung <input type="checkbox"/> Aktiviert <input type="checkbox"/> Deaktiviert <input type="checkbox"/> Sonstig: | |

1.2 Änderungshistorie

| Wann? | Wer? | Was? |
|------------------|------------------------|--|
| 18. Februar 2022 | Abel Hodelin Hernandez | Version 1.1 |
| 25. Februar 2022 | Abel Hodelin Hernandez | Große Überarbeitung mit weniger technischen Aspekten |

Kapitel 2

Einführung

Im Datenintegrationszentrum (DIZ) werden Daten aus verschiedenen Fachabteilungen und Systemen zusammengeführt. Ein zentrales Puzzleteil für die Zwischenspeicherung der Information dieser Systemen ist das clinical staging Data Warehouse (csDWH). In dieser Datenbank (DB) werden alle relevanten klinischen Systeme abgebildet. Diese Daten werden im Rahmen des Datenschutzes sowie der Datenqualität aufbereitet und anschließend an weitere Komponenten des DIZ übertragen.

Kapitel 3

Konfiguration der Instanz des csDWH

Das csDWH, welches die Forschungsdaten beinhaltet, wird sich in einem Linux Server im klinik-Netz mit der letzten, stabilen Ubuntu Version (Long-term support (LTS)) befinden. Diese DB wird in PostgreSQL implementiert und administriert.

Auf dem Server wird ein Benutzer ohne administrative rechte auf dem Betriebssystem angelegt. Dieser ist für die Administration der DB-Instanz vorgesehen.

Das csDWH besitzt zwei strukturell gleiche DB, **staging** für die Produktion und **staging_test** zum testen. Die DB sind in verschiedenen Schemata geteilt, jede davon entspricht eine Quelle, Zusammenfassung von Systemen oder Anwendung. Die Information der Schemata liegt in Kapitel 5.

Kapitel 4

Sicherung des csDWH

Um das System zu sichern wird eine Sicherung der kompletten csDWH-Instanz täglich in der Nacht gemacht. Dieser Prozess verläuft in zwei Prozeduren, erst verläuft die Sicherung der csDWH-Instanz und direkt danach wird diese verschlüsselt komprimiert. Die komprimierte Sicherung wird auf dem Server und auf einer extra-Virtual Maschine (VM) gespeichert. Die Speicherung einer Sicherung wird von fünf Tagen beibehalten.

Kapitel 5

Schemas

Die Schemata speichern die "rohe" **pseudonymisierte** Information der ursprünglichen Systeme oder die Metadaten. Diese Daten werden in Tabellen gespeichert. Wichtige Hinweis ist, dass die Daten in dem csDWH unverändert bleiben sollen. Deswegen werden (materialisierten) Sichten für weitere Uses Cases bereitgestellt.

5.1 p21

Dieses Schema speichert die jährliche Information der §21, die von Medizincontrolling generiert wird.

Der jährliche Rhythmus ist zu groß, als dass die Daten bspw. zur Rekrutierung von Patienten für Studien aber auch zu Forschung genutzt werden können. Auf diesem Grund wird diese Information in näherer Zukunft nicht mehr von Medizincontrolling genommen, sondern direkt aus dem Krankenhausinformationssystem (KIS).

In diesem Schema befinden sich auch die Sichten für Extraction Transformation Load (ETL)-Prozessen die, solche Information aus §21 benötigen. Der Inhalt dieser Sichten entspricht die Formatierung der Krankenhausentgeltgesetz (KHEntgG) §21 Übermittlung und Nutzung der Daten.

5.2 kis

Hier werden die tagesaktuellen extrahierten Daten zu Patienten, Fällen, Bewegungen, Diagnosen und Prozeduren direkt aus dem Quellsystem KIS gespeichert. Mit Hilfe diesem Schema lassen sich viele der Abbildungen für weitere Projekte realisieren.

5.3 copra

Hier wird die tagesaktuelle Information aus Patientendatenmanagementsystem (PDMS), in unserm Fall das COPRA-System, gespeichert. Dieses Schema beinhaltet Biosignalsparameter, Befunde, ärztliche Anweisungen und Überblick über Behandlungsschritte.

5.4 gtds

Dieses Schema speichert die Daten der mainzenen Instanz des Gißener Tumordokumentationssystem (GTDS) und somit die Erfassung und Verarbeitung der Daten der revidierten Basisdokumentation klinischen Krebsregistern.

5.5 centrallab

Dieses Schema dient die Speicherung der Daten aus dem Zentrallabor (Institut für Klinische Chemie und Laboratoriumsmedizin) gespeichert.

5.6 imagic

Hier wird die Information aus dem IMAGIC-System gespeichert. Dieses Schema beinhaltet Information aus der Hautklinik. Davon Metadaten der Bilder sowie Befunde anhand der Bilder.

5.7 metadata_repository

Dieses Schema speichert die Information der Metadaten aller anderen Schemata.

Kapitel 6

Benutzer & Zugang

Die PostgreSQL-Instanz des csDWH hat einen Administrator, der keine administrative Rechte auf dem Server besitzt. Die Instanz hat verschiedene Benutzer mit unterschiedlichen Aufgaben. Dadurch das System unterschiedliche Schemas und zwei DB besitzt, nämlich für Testen und Produktion, werden die Benutzer von jeder DB getrennt. Das bedeutet, dass die Benutzer beider Systeme nicht dieselbe sind.

Als Benutzernamen in dem csDWH werden dieselbe Kürzel wie an der Unimedizin (UM) in dem produktiven System und mit dem Suffix `_test` in dem Testsystem benutzt. Die Benutzernamen in dem csDWH werden wie folgt definiert:

- Produktives System: Kürzel. z.B. `her11a`
- Testsystem: Kürzel_test. z.B. `her11a_test`

Die Datenlieferanten haben nur Zugriffsrechte zu ihren Schemas. Dazu können sie in dem Testsystem Tabellen und Sichten kreieren, ändern und löschen, und Daten einfügen, lesen, ändern und löschen. In der Zeit der Umsetzung eines Schemas auf dem produktiven System dürfen die Datenlieferanten auch Tabellen oder Sichten kreieren, ändern und löschen. Nach dieser Zeit dürfen die Datenlieferanten keine strukturelle Änderungen mehr in dem zugewiesenen Schemas übernehmen. In dem produktiven Systemen dürfen die Datenlieferanten weiter Daten einfügen und lesen.

Der Zugang zu dem csDWH und dessen Verschlüsselung werden technisch auf die höchste Stufe gesetzt.

Um mit der Daten des csDWH zu arbeiten, sollten die Personen oder Anwendungen via Use & Access Committee (UAC) eine Genehmigung für Lesezugriff beantragen.