

**Konzeptionierung und Entwicklung eines Metadatenrepositoriums zur
Exploration des klinischen Datenbestandes in Forschung und Routine**

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Humanbiologie
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Eberhardt, Florian
aus Alsfeld

Gießen 2023

**Aus dem Institut für Medizinische Informatik (IMI)
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen**

Gutachter: Herr Prof. Dr. Keywan Sohrabi

Gutachter: Herr Prof. Dr. Marek Bartkuhn

Tag der Disputation: 19.12.2023

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

Ort, Datum

Unterschrift (Florian Eberhardt)

Inhaltsverzeichnis

1. Einleitung	1
1.1 Datenverarbeitung im Gesundheitswesen	1
1.2 Die Medizininformatik-Initiative und ihre Datenintegrationszentren	4
1.4 Metadaten und Analysen	8
1.5 Medizinische Klassifikationen und Terminologien	10
1.6 Stand der Wissenschaft und Technik	12
1.7 Rechtliche und ethische Aspekte	16
1.8 Motivation	18
2. Zielsetzung	20
3. Material und Methoden	22
3.1 PostgreSQL und JSON	23
3.2 Softwareentwicklung	25
3.2.1 Python und Flask	25
3.2.2 PyCharm	26
3.2.3 Python Erweiterungen	26
3.2.4 Node.js, Vue.js und Vuetify	27
3.2.5 Visual Studio Code	28
3.2.6 Plotly	29
3.2.7 Versionsverwaltung Git	30
3.3 Samplify.MDR	32
3.4 Docker	33
3.5 Airbyte	35
3.6 Keycloak	36
3.7 Datenintegrationsplattform	37
3.8 IT-Grundschutz	39
4. Ergebnisse	40
4.1 Projektplanung	40

4.2 Organisatorische und rechtliche Grundlagen	41
4.2.1 Datenzugriffs- und Nutzungsordnung	42
4.2.2 Geschäftsordnung des Use- & Access Committee	43
4.2.3 Geschäftsordnung des Datenintegrationszentrums	43
4.3 Kerndatensatz	44
4.4 Konzeptionelle Trennung der Informationen	45
4.5 Semantische Ausgliederung – Samply.MDR	46
4.6 eMDR	48
4.6.1 Konzeptionierung und interne Vorgaben	48
4.6.2 Backend	51
4.6.3 Frontend	55
4.6.4 Funktionsumfang	57
4.6.5 Integration in die Datenintegrationsplattform	61
4.7 Multiple Systeminstallationen	63
4.8 Containerbasierte Bereitstellung	66
4.9 Integration in das Gießener Datenintegrationszentrum	68
4.10 Umsetzung des IT-Grundschutzes	75
5. Diskussion	77
6. Schlussfolgerung	84
7. Zusammenfassung	87
8. Danksagung	89
9. Abbildungsverzeichnis	90
10. Tabellenverzeichnis	93
11. Abkürzungsverzeichnis	94
12. Literaturverzeichnis	96
13. Anhang	107
13.1 Daten-Zugriffs- und Nutzungsordnung	107
13.2 Geschäftsordnung des Use & Access Committee	117
13.3 Geschäftsordnung für das Datenintegrationszentrum	123

13.4 Technischer Ablauf des Datenintegrationszentrums	128
13.5 Organisatorischer Ablauf des Datenintegrationszentrums	129
13.6 Systemschicht des Datenintegrationszentrums	130
13.7 IT-Infrastruktur für das Datenintegrationszentrum	131
13.8 Datenschutzkonzept DIZ Gießen	136
13.9 Lastenheft	153
13.10 Pflichtenheft	157
13.11 Projektplan	182

1. Einleitung

1.1 Datenverarbeitung im Gesundheitswesen

Das deutsche Gesundheitswesen befindet sich weiterhin in einem starken digitalen Wandel. Die Gründe hierfür sind sekundär der technologische Wandel, primär jedoch die Sicht auf das enorme Potential im Kontext der Weiterentwicklung und Verbesserung der Patientenversorgung, welches sich durch diesen technologischen Fortschritt und die einhergehenden digitalen Prozesse ergeben. (Lux et al., 2017, p. 691) Das wichtigste Ziel ist stets die Patientenversorgung zu verbessern und zu optimieren, welches durch neue Techniken zur Diagnosefindung und Behandlung initial gewährleistet wird. Durch diese Verfahren und Techniken erfolgt die Generierung neuen Wissens, welches im Umkehrschluss dazu genutzt werden kann, die verwendeten Methoden anzureichern und kontinuierlich zu verbessern. (Behm and Klenk, 2021, p. 2)

Von der Digitalisierung im Gesundheitswesen wird auch eine Verbesserung der Versorgung aus infrastruktureller Sicht, insbesondere in unterversorgten ländlichen Umgebungen, erwartet. Der Fokus liegt hierbei auf dem Informationsaustausch zwischen Patienten und deren Ärzten beziehungsweise Leistungserbringern. Durch das E-Health-Gesetz (Gauck et al., 2015), welches Videosprechstunden ermöglicht und elektronische Patientenakten (ePA) sowie elektronische Rezepte (E-Rezept) verpflichtend vorschreibt, wurde bereits ein wichtiger Eckpfeiler für diesen digitalen medizinischen Datenaustausch geschaffen. (Bundesministerium für Gesundheit, 2022) In diesem Umfeld ist auch die Verwendung von mobilen Lösungen wie Apps angedacht, welche unter den Kontext mHealth fallen. mHealth stellt die Unterstützung medizinischer Bereiche, wie beispielsweise die Versorgung und Forschung dar, welche unter Zuhilfenahme mobiler Endgeräte wie Smartphones, Tablets oder sonstige Geräte zum mobilen Monitoring erfolgt. (World Health Organization, 2011, p. 6) Aus Sicht der allgemeinen Digitalisierung, der Verfügbarkeit und dem sozialen Verhalten in Bezug auf mobile Endgeräte, stellen Apps derzeit eine große Möglichkeit dar, um medizinische Informationen erfassen und verarbeiten zu können. Es gilt jedoch auch hierbei stets die Datenqualität und -Transparenz zu berücksichtigen, da durch die mittlerweile enorme Anzahl an Tracking-, Fitness- und Sport-, sowie Monitoring-Apps eine Vielzahl medizinischer Daten zwar generiert, jedoch teilweise nicht direkt auf ihre Qualität überprüft oder strukturiert werden. (Olla and Shimskey, 2015, p. 299) Weiter ist meist nicht ersichtlich, nach welchen Standards und Methoden die Daten erhoben beziehungsweise erfasst werden. Letzterer Punkt ist insbesondere bei Befragungen der

Fall, wodurch bei Apps nicht sichergestellt werden kann, ob die gestellten medizinischen Fragestellungen verstanden, richtig interpretiert, korrekt gestellt und von der richtigen Person beantwortet wurden. Wichtig für die Wissensgenerierung ist selbstverständlich eine große und umfangreiche Datenlage, jedoch muss die Herkunft und Validität stets gewährleistet sein. Je nach Geschäftsbedingungen der Lösung kann beispielsweise eine Weitergabe der Gesundheitsdaten an Dritte erfolgen, welches nicht nur aus Sicht des Datenschutzes, sondern auch aus Sicht der Ethik kritisch zu bewerten ist, da diese Daten so auch beispielsweise an Versicherungen und Pharmakonzerne gelangen könnten. (Behm and Klenk, 2021, pp. 2, 3) (Lux et al., 2017, p. 691) (Zingg et al., 2019, p. 113) (Bruland et al., 2018, p. 103) Ein Resultat daraus könnte konkret eine Benachteiligung oder Änderung im Versicherungsschutz darstellen. (Townend, 2010, pp. 477–479, 481–483)

Ein weiterer kritischer Aspekt stellt das mittlerweile hohe Vorkommen an mobilen Applikationen dar, wodurch immer mehr Apps durch Neue verdrängt und im späteren Verlauf nicht mehr betrieben werden. Eine durch Appfigures im Oktober 2020 durchgeführte Analyse konnte somit innerhalb des Apple App-Stores ein Vorkommen von 48.608 Apps im Bereich Medizin und 82.633 Apps im Bereich Gesundheit und Fitness, für Quartal drei, verdeutlichen. (Appfigures, 2020) Die Speicherung, Verarbeitung und Löschung der bereits innerhalb der Apps erhobenen Daten ist dann fraglich, ebenso wie die Migration der Daten in andere Systeme. (Pfanstiel et al., 2020, p. 22)

Die Digitalisierung im Gesundheitswesen bringt noch weitere Potenziale mit sich, wie beispielsweise die mögliche Qualitätsverbesserung und damit die Vermeidung von Fehlversorgungen. Durch eine bessere Dokumentation und einen digitalen Datenaustausch zwischen Ärzten und Kliniken werden beispielsweise Doppelbehandlungen vermieden und Behandlungswege sind deutlich transparenter. Insbesondere diese Transparenz ist wichtig, um Optimierungsmöglichkeiten zu erkennen und Teilprozesse dieser Wege mit neuen Erkenntnissen anreichern zu können. Aus Sicht der Krankenhäuser erfolgt so eine bessere Kosten-Nutzen-Relation und eine höhere Behandlungsqualität für die Patienten. Geprüft werden kann dieser Aspekt der Digitalisierung beispielsweise durch digitale Checklisten und Ablaufdokumentationen, welche dann innerhalb der Organisation in Echtzeit visualisiert und verglichen werden können. (Brönneke and Debatin, 2022, pp. 344, 345) (Lux et al., 2017, p. 693) Der administrative Hintergrund und Aufwand hierbei ist jedoch nicht zu unterschätzen und stellt Ärzte und Personal vor erhebliche Herausforderungen sowie Mehraufwände. Dies bezieht sich nicht nur auf die Beschaffung und Umsetzung von IT-

Lösungen zur Digitalisierung, sondern insbesondere auf die Schulung des Personals und generell die Bereitstellung der zeitlichen Personalressourcen, zur Nutzung und Betreuung der digitalen Lösungen, im Versorgungsalltag. (Berufsverband der Deutschen Urologen e.V., 2021, p. 542) (Meißner, 2018, p. 32)

Im Zuge des Krankenhaus-Reportes 2019 wurden, bei einer Erhebung im Jahre 2017, Einrichtungen zum Stand der Digitalisierung in konkreten Bereichen befragt, wobei ersichtlich wurde, dass beispielsweise im Bereich der „Entscheidungsunterstützung für Diagnostik, Therapie und Pflege“ noch 72,3% der 213 befragten Einrichtungen keine digitale Lösung vorweisen können, wobei 43,2% davon auch keine Umsetzung vorgesehen haben. (Hübner et al., 2018, p. 30) (Klauber et al., 2019, p. 40)

Wichtig hierbei sind prinzipiell die Herkunft, der Umfang, die Qualität und die Validität der Daten, um ein solches entscheidungsunterstützendes System zum produktiven Einsatz adäquat und sicher antrainieren zu können. Dazu kommt die Problematik, dass aufgrund der unterschiedlichen Fachdisziplinen die Daten unterschiedliche Strukturen und Anforderungen vorweisen, wodurch in der Regel Insellösungen in den einzelnen klinischen Abteilungen entstehen, welche eine zusätzliche personelle Betreuung benötigen. Ziel der klinischen Entscheidungsunterstützung soll jedoch die Unterstützung und Entlastung des medizinischen Personals darstellen, wodurch entweder die Einführung zentralisierter Systeme oder die Einbindung weiterer IT-Fachkräfte angestrebt werden sollte. (Marx et al., 2021, pp. 133, 230)

Die Verarbeitung der medizinischen Daten zur Generierung neuen Wissens erfordert somit eine große Datenbasis, welche konkrete Eigenschaften vorweisen müssen. Hierbei ist jedoch aus Sicht der Herkunft der Daten noch wichtig zwischen Daten der klinischen Routine, der klinischen Forschung und freie mobile Applikationen zu unterscheiden.

Die innerhalb der klinischen Routine vorhandene Systeminhomogenität, aufgrund der historischen Entwicklung und erzeugter IT-Insellösungen, stellt den Bereich der Medizininformatik vor große Herausforderungen im Bereich der Datennutzung, aus Sicht der gesamtheitlichen Datenlandschaft. Der Bereich der medizinischen Forschung sieht diesen Problemen ebenfalls entgegen und birgt das Problem eines, je nach Abteilung und Institution vorgegebenen, freieren Datenmanagements. Dies beinhaltet auch die Datenhaltung auf externen mobilen Medien, Cloud-Umgebungen oder abgeschottet und teilweise ohne Netzanbindung auf dienstlichen Endgeräten. (Eggert, 2011, p. 1) (Sunyaev et al., 2010, pp. 223–224) (Waldschmidt-Dietz and Krippes, 2016, pp. 9–10)

Aufgrund des hohen Stellenwertes der Daten werden, im Hinblick diverser Förderungen und Fokusgruppen, die Bereiche der klinischen Routine und Forschung besonders betrachtet. Die eigentlich benötigte kontextweite Datenverarbeitung kann nur dann erfolgen, wenn eine Interoperabilität und Homogenität der Daten gewährleistet werden kann, wodurch aus Sicht der medizinischen Informatik ein hoher Verarbeitungs- und Anpassungsaufwand an den Daten entsteht.

Das nachfolgende Kapitel behandelt die Folge dieser resultierenden Problemstellungen im Sinne der Förderung einer Medizininformatik-Initiative, welche konkrete Ziele zur Etablierung, Optimierung und Anpassung des ganzheitlichen Datenmanagements verfolgt.

1.2 Die Medizininformatik-Initiative und ihre Datenintegrationszentren

Das Bundesministerium für Bildung und Forschung (BMBF) hat im November 2015 die Medizininformatik-Initiative (MII) ausgeschrieben. Das Ziel der Initiative stellt primär die Zusammenführung und intelligente Nutzung von medizinischen Versorgungs- und Forschungsdaten dar. Dazu zählt nicht nur die Nutzung von aktuellen IT-Innovationen im Sektor der Gesundheitsversorgung, sondern auch die Generierung neuer Lösungen und die Digitalisierung bestehender Prozesse. Daraus resultierend soll eine bessere datenbasierte Gesundheitsversorgung entstehen und das Potential dieser Datenbasen im vollen Umfang zur Generierung neuen Wissens genutzt werden. (H.-U. Prokosch, 2022) (Gehring and Eulenfeld, 2018, p. e46)

Im Zuge der Initiative bildeten sich deutschlandweit mehrere Konsortien, welche sich aus unterschiedlichen Universitätskliniken, Hochschulen, Forschungseinrichtungen, privatisierten Kliniken und Industriepartnern zusammensetzen (Abbildung 1).

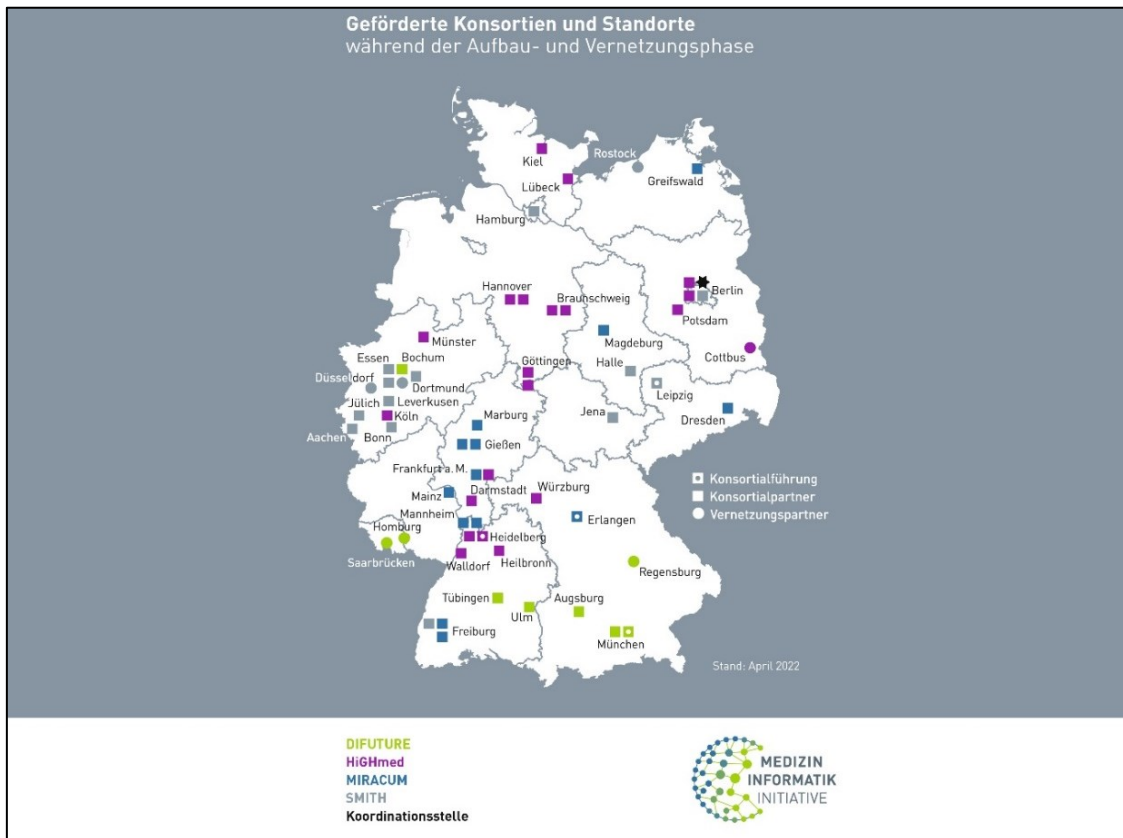


Abbildung 1: Standorte der MII-Konsortien (TMF e.V., 2020a)

Die Initiative selbst gliedert sich aus zeitlicher und inhaltlicher Sicht in drei unterschiedliche Phasen, welche in Abbildung 2 ersichtlich sind.



Abbildung 2: Förderphasen der MII (TMF e.V., 2020b)

Die Konzeptphase (2016-2017) förderte initial sieben dieser Konsortien (*DIFUTURE*, *HD4CR*, *HiGHmed*, *MIRACUM*, *share-it!*, *SMITH*) mit dem Ziel, Konzepte zur interoperablen Datennutzung zu erstellen und konkreter Anwendungsfälle, sogenannte Use-Cases, zu definieren, um diese Konzepte bestätigen zu können. Die nächste Phase stellt die Aufbau- und Vernetzungsphase (2018-2022) dar, welche nur noch vier der

Konsortien fördert, mit dem konkreten Ziel pro Versorgungsstandort ein lokales Datenintegrationszentrum, ein sogenanntes DIZ, aufzubauen. Dieses DIZ entwickelt die entsprechenden Lösungen für die Use-Cases und verfolgt primär das Ziel Forschungsdaten und Routinedaten in einer gemeinsamen Basis zu nutzen und einen Mehrwert für die Gesundheitsversorgung zu generieren. Die damit einhergehenden organisatorischen, rechtlichen und technischen Prozesse werden einem extern durchgeführten Audit unterzogen, um eine Bestätigung der Ziele und eine Nachvollziehbarkeit der Lösungen zu gewährleisten.

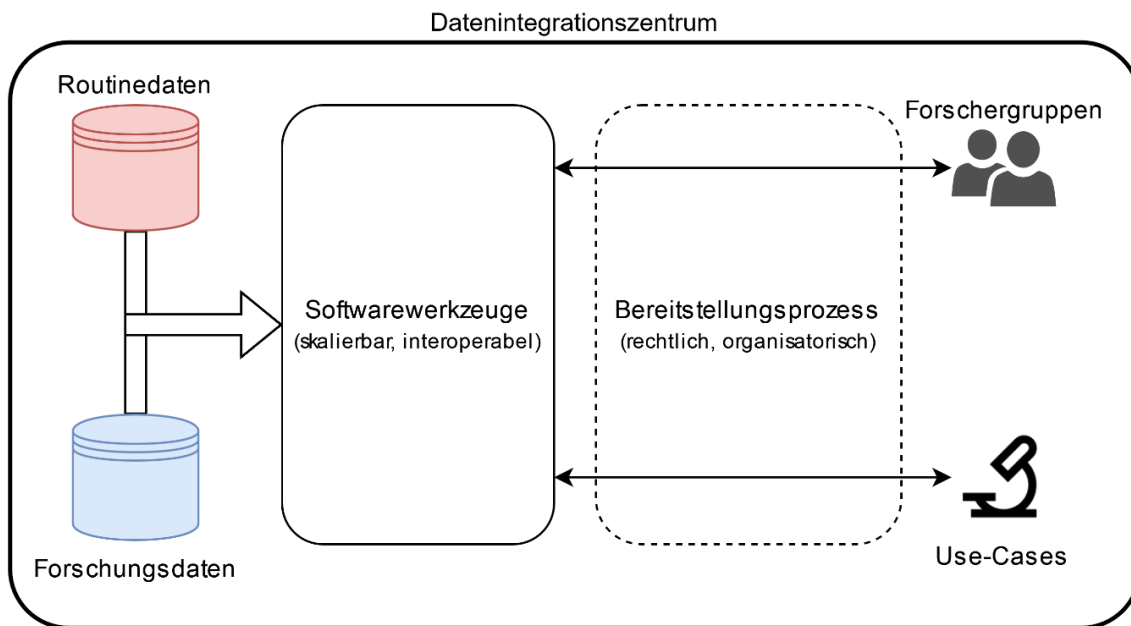


Abbildung 3: Schematischer Aufbau eines Datenintegrationszentrums (Eigene Darstellung)

Ein schematischer Ablauf dieses Vorhabens ist in Abbildung 3 verdeutlicht und stellt die einzelnen zu betrachtenden Zwischenschritte von der Zusammenführung der Daten, über die Verwendung skalierbarer und interoperabler Softwarelösungen, bis zur Bereitstellung der Daten für die Use-Cases und Forschergruppen, unter Berücksichtigung der rechtlich organisatorischen Bereitstellungsprozesse, dar.

Ein besonderes Augenmerk der beiden Phasen liegt auch auf dem datenschutzkonformen Austausch von Wissen zwischen den unterschiedlichen Datenintegrationszentren, wodurch dieses Wissen effizienter genutzt und eventuell durch weitere Daten angereichert werden kann. Die letzte Phase stellt die Ausbau- und Erweiterungsphase (2023-2026) dar. Sollte die zweite Phase erfolgreich sein und entsprechend nachvollziehbare sowie nutzbare Ergebnisse erzielt haben, so sollen diese auf weitere Kliniken und Einrichtungen übertragen werden, um einen flächendeckenden Nutzen zu ermöglichen. (TMF e.V., 2020b) (Semler et al., 2018)

Die vier Konsortien DIFUTURE, HiGHmed, MIRACUM und SMITH sind in die zweite Phase übergegangen und widmen sich dem Aufbau der Datenintegrationszentren. Zu dem Konsortium MIRACUM zählt neben den Standorten Erlangen, Freiburg, Mannheim, Mainz, Frankfurt am Main, Marburg, Magdeburg, Dresden und Greifswald auch der Standort Gießen mit dem Universitätsklinikum Gießen und Marburg GmbH Standort Gießen (UGKM-Gi), der Justus-Liebig-Universität Gießen (JLU) und die Technische Hochschule Mittelhessen (THM). (Gehring and Eulenfeld, 2018) Hierbei stellt das UKGM und die JLU einen der Standorte eines Datenintegrationszentrums dar, wohingegen die THM Tätigkeiten im Bereich der Fort- und Weiterbildung wahrnimmt. Das DIZ am Standort Gießen übernimmt ebenfalls die primäre Aufgabe, Daten der Forschung und Daten aus der Routineversorgung in eine gemeinsame Datenbasis zu integrieren und diese Daten intelligent nutzbar zu machen. Hierzu verfolgt das Konsortium MIRACUM drei Use-Cases („Alerting in Care“, „From Data to Knowledge“ und „From Knowledge to Action“), um die weiteren Ziele des Zentrums nachweisbar umzusetzen. (Cuggia and Combes, 2019) Darunter fallen aus technischer Sicht die Integration von Daten aus unterschiedlichsten klinischen Routine- und Forschungs-Systemen am Standort sowie der standortübergreifende Datenaustausch mit den anderen Datenintegrationszentren. Letzterer Aspekt hat einen großen Einfluss auf die lokal benötigte Interoperabilität und Standardisierung der Datenstrukturen, welche auch für die Wiederverwendbarkeit der Daten essenziell sind. (Prof. Dr. H.-U. Prokosch, 2022) (Prokosch et al., 2018)

Grundsätzlich stellt aus Perspektive des Standortes das DIZ auch den primären Einstiegspunkt für Forschende im medizinischen Bereich dar. Das DIZ übernimmt für die Forschenden eine Beratungsfunktion im Kontext der Digitalisierung, der optimalen Nutzung bestehender IT-Lösungen, zum Beispiel im Bereich von Datenauswertungen und die organisatorischen Wege bis zum Erhalt der benötigten Forschungsdatenextrakte. Das DIZ verfügt über Systeme, um sämtliche organisatorische Abläufe für die Forschenden bei einer gewünschten Datennutzung abzubilden und letztlich die ausgeleiteten Routine- und Forschungsdaten datenschutzkonform und zentralisiert bereitzustellen. (Michel-Backofen, 2021)

Die größtenteils sehr umfangreiche und komplexe Bereitstellung von Datenextrakten erfordert aus interner Sicht des Datenintegrationszentrums einen großen Grad an Strukturierung der relevanten Daten. Diese Strukturierung erfolgt mittels der Verwendung von Metadaten, welche im nachfolgenden Kapitel erläutert werden.

1.4 Metadaten und Analysen

Metadaten werden bereits seit mehreren Jahrhunderten und -tausenden und somit weit vor dem digitalen Zeitalter für unterschiedlichste Bereiche genutzt, so auch in der antiken Bibliothek von Alexandria zur Katalogisierung von Schriftstücken. (Kennedy, 2008, pp. 136–137) Grundsätzlich beschreiben Metadaten eine Menge an Informationen und ermöglichen somit eine Strukturierung und Ordnung. (Gartner, 2016, pp. 8, 15) Diese Art der Datenbeschreibung erfolgt nicht nur in wissenschaftlichen Bereichen, in denen eine enorme Menge und Vielfalt an Daten heterogen vorliegen, sondern beispielsweise auch in sämtlichen social-media-Bereichen, welche auch unter den Aspekt des Big-Data fallen. (Riley and National Information Standards Organization (U.S.), 2017, p. 2) Big-Data selbst beschreibt eine Datengrundlage, welche sich durch ihr Volumen, innere Strukturvielfalt (Heterogenität) und ihrer Geschwindigkeit, insbesondere im Wachstum und der eventuellen Änderung vorhandener Daten, auszeichnet. Eine solche Grundlage lässt sich nicht über gängige Verfahren wie simple Datenbankabfragen auswerten und benötigt somit spezielle, gegebenenfalls angepasste Verfahren und eine grundlegende Strukturierung beziehungsweise Organisation der Daten. (Fasel and Meier, 2016, p. V) (Sagiroglu and Sinanc, 2013, pp. 42–43) Dies unterstützend stellen Metadaten deskriptive Informationen über konkrete Bereiche und einzelne Daten bereit. Der Grundgedanke von Metadaten ist es somit Daten mit zusätzlichen Daten / Informationen zu beschreiben, welches insbesondere im Bereich der Wissenschaft für eine bessere Analysefähigkeit sorgt. (Duval, 2001, pp. 591–592) So können große heterogene Datenmengen durch einzelne Metadaten strukturiert werden. Ein Datenstrom mit Patienteninformationen kann demzufolge beispielsweise in die einzelnen Metadaten-Attribute *Fallnummer*, *Name*, *Vorname*, *Geburtsdatum*, *Diagnose* und *Prozedur* unterteilt sowie strukturiert werden. Die Metadaten liefern somit auch einen konkreten Kontext zu zunächst undefinierten Datenströmen.

Zur Verwaltung der verwendeten Metadaten, beziehungsweise der Metadaten-Attribute, werden sogenannte Metadaten-Repositoryn eingesetzt. (Bargmeyer and Gillman, 2020, p. 1) (Hippen and Greene, 2021, pp. 277, 278) Die Repositoryn können nicht nur die verfügbaren Attribute selbst, sondern auch Beziehungen unter diesen darstellen. Ebenso erfordert eine Vielzahl an Attributen die Möglichkeit Beschreibungen hinzuzufügen, um somit die deskriptiven Daten und ihren Nutzen für die weitere Verwendung offen zu legen. (Zeng et al., 2005, p. 339) Metadaten-Repositoryn dienen zwar somit primär der Verwaltung der Attribute, können und sollen jedoch auch im Bereich von Analysen direkt unterstützen.

Im Falle von großen inhomogenen Datenmengen, welche zusätzlich noch auf unterschiedlichste Standorte verteilt sind, ist eine gemeinsame Analysemöglichkeit nur durch gemeinsam definierte Strukturen möglich. Die Notwendigkeit der Repositorien ist, aufgrund der wiederum steigenden Anzahl an Metadaten und neuen Strukturierungsideen, unerlässlich und benötigt auch in diesem abstrakten Bereich klar definierte Strukturen. Diese sind beispielsweise über die Norm ISO/IEC 11179-3:2013 und konkret aus medizinisch wissenschaftlicher Sicht auch über die FAIR-Prinzipien als Richtlinie definiert. (ISO/IEC JTC 1/SC 32 Data management and interchange, 2013) (Kapsner et al., 2021, pp. 828–832) Die Norm selbst gibt bereits konkrete Namensschemata, den Detailgrad der Informationen und zu verwendende Datentypen, basierend auf referenzierten Standards, wie beispielsweise Datumsformate vor, wodurch normbasierte Repositorien selbst bereits interoperabel verwendet werden können. (ISO/IEC JTC 1/SC 32 Data management and interchange, 2013)

Sobald Metadaten verwendet werden, können die darunterliegenden Daten einheitlicher und vereinfachter für Analysen verwendet werden. Sofern Datenströme und Datenbasen an unterschiedlichen Standorten trotz ihrer Inhomogenität gleiche Metadaten-Attribute zur Strukturierung erhalten, könnten diese abstrahierten beschreibenden Daten für gemeinsame standortübergreifende Analysen genutzt werden. Die Verwendung von Metadaten setzt jedoch nicht automatisch eine ausreichende Datenqualität der darunterliegenden Daten voraus und benötigt meistens eine umfangreiche Aufbereitung der Datenbasis. (Gonçalves and Musen, 2019, p. 2) (Canakoglu et al., 2019, p. 19)

Ein wichtiger Aspekt ist die Unterscheidung von Metadaten in drei Kategorien, welche im Zuge der Anwendung im medizinischen Bereich eine tragende Rolle spielen. Zunächst kann eine Unterteilung in *beschreibende* Metadaten erfolgen, welche Informationen zum korrespondierenden Datensatz im Kontext der Herkunft, dem Format, dem Inhalt und dem Ersteller liefern. Diese erfüllt den klassischen Gedanken von Metadaten und dienen als primärer Informationsgeber. Eine weitere Kategorie wird durch die *administrativen* Metadaten repräsentiert, welche insbesondere im medizinischen Bereich im Kontext des Datenschutzes und der Datensicherheit von hoher Bedeutung sind. Diese Art der Metadaten definiert vordergründig Zugriffsrechte auf die zu Grunde liegende Datenbasis, beziehungsweise den referenzierten Datensätzen und kann weitere Informationen zu den technischen Gegebenheiten abbilden. Letzterer Aspekt ist für ein zukunftsorientiertes Datenmanagement unerlässlich, da sowohl die Hardware, als auch die Software, welche die Daten erzeugt haben, erfasst werden müssen. Somit können auch rückwirkend bei neuen Erkenntnissen, beispielsweise bei vergangenen Software-Fehlern, die betroffenen

Datensätze identifiziert und aus der gesamten Datenbasis extrahiert und gegebenenfalls korrigiert werden. Die letzte Kategorie stellt die *strukturellen* Metadaten dar, welche die Verknüpfung von Daten ermöglicht. Durch diese Fähigkeit können sämtliche medizinische Daten mit einem konkreten Patienten verknüpft und die Erzeugung von bestimmten Abläufen und Hierarchien innerhalb des Behandlungsablaufes durchgeführt werden. Der Umfang in der Abbildung der einzelnen Kategorien basiert auf dem wünschenswerten Grad an Strukturierung und der korrespondierenden Datenqualität. (Gartner, 2016, pp. 6–8) (Leipzig et al., 2021, p. 4)

Aufgrund der Abstrahierung und Strukturierung von Daten an unterschiedlichen Standorten erfolgen derzeit ein sich weiter erhöhendes Aufkommen an Analysen auf Basis von Metadaten. Ein Vorteil ist, dass eine definierte Suchabfrage auf ein Metadaten-Set direkt an unterschiedlichen Standorten, unabhängig von Kenntnissen der darunterliegenden Datenbasis und Struktur, ausgeführt und im Anschluss aggregiert werden kann. Solche Auswertungen erfolgen beispielsweise im Kontext der Medizininformatik-Initiative übergreifend in den Datenintegrationszentren und auch im Kontext der globalen SARS-CoV-2 Pandemie. (Warin, 2021) (Deutsche Gesellschaft für Urologie, 2019, p. 711) (Kapsner et al., 2021, p. 827)

1.5 Medizinische Klassifikationen und Terminologien

Im Zuge der Verwendung von Metadaten sowie der Strukturierung und Ordnung von Daten, nehmen medizinische Klassifikationssysteme und Terminologien eine vordergründige Rolle ein. Diese Terminologien repräsentieren das Vokabular zur Benennung medizinischer Gegebenheiten unter anderem in Bezug auf deren Inhalt, Lokalität und Struktur. Viele Begrifflichkeiten erhielten bereits eine erste Erwähnung im 5. bis 2. Jahrhundert v. Chr. in der griechischen Antike, genauer im Corpus Hippocraticum und entwickelten sich im Laufe der Zeit und Entwicklung der Medizin stets weiter, woraus die heutigen Terminologien hervorgingen. Eine wichtige Bedeutung und Notwendigkeit erhielten diese Systeme, aus zeitlicher Sicht erst nach dem zweiten Weltkrieg, aufgrund der steigenden Datenauswertungen im maschinellen Bereich, wodurch konkrete vergleichbare Fachbegriffe benötigt wurden. (Rienhoff and Semler, 2015, p. 29) (Fangerau et al., 2017, p. 13) (Steger, 2019, pp. 7–12) (Golder, 2007, pp. 9–17)

Die daraus resultierenden Klassifikationssysteme dienen primär der einheitlichen Kodierung im medizinischen Alltag und stellen die Grundlage von Datenanalysen und Datenintegrationen dar. Die aktuellen Standards, welche seitens der klinischen Routine

primär durch ICD-10 und OPS-301 und seitens der klinischen Forschung durch SNOMED CT und LOINC ergänzt und repräsentiert werden, finden bereits einen großen Einsatz und bieten eine ideale Grundlage zum Austausch und zur Analyse medizinischer Fragestellungen, aufgrund der standardisierten gemeinsamen Struktur, über sämtliche Informationssysteme hinweg. (Bodenreider et al., 2018, p. 129,138) (Richesson et al., 2019, p. 413)

ICD-10 (International Classification of Diseases, in der 10. Revision) ist eine medizinische Klassifikation, die zur Kodierung von Diagnosen und Krankheiten verwendet wird. ICD-10 wird in der klinischen Praxis zur Dokumentation von Diagnosen, zur Abrechnung von Leistungen und zur epidemiologischen Forschung eingesetzt. (Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023a) Ergänzend hierzu bildet OPS-301 (Operationen- und Prozedurenschlüssel), ebenfalls als eine medizinische Klassifikation, die Kodierung von medizinischen Verfahren und Operationen ab. Die Klassifikation wird in der klinischen Praxis zur Dokumentation von durchgeführten Verfahren und Operationen, zur Abrechnung von Leistungen und zur Qualitätssicherung eingesetzt. (Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023b) (DIMDI, 2003) LOINC (Logical Observation Identifiers Names and Codes) stellt eine Terminologie zur standardisierten Darstellung von Untersuchungen und klinischen Betrachtungen im Kontext der Labormedizin dar. Die Terminologie umfasst weiter Dokumententypen, Fragebögen und konkrete Untersuchungen aus dem Laborumfeld. (Regenstrief Institute, Inc, 2023) LOINC, welches ursprünglich vom Regenstrief Institut entwickelt wurde, wird derzeit im Zuge des Paragraf 355 Abs. 7 SGB V vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) in die deutsche Sprache übersetzt und an die Gegebenheiten, insbesondere die der klinischen Forschung und Routine, angepasst. (Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023c) (Bundestag, 1989) Die derzeit umfassendste Terminologie wird durch SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) repräsentiert, welche multilingual und bereits in über achtzig Ländern eingesetzt wird. SNOMED CT bietet eine Abbildung und Referenzen unter verschiedenen internationalen Standards, weshalb sie auch als Ontologie einen wichtigen Bestandteil in der Digitalisierung, dem Austausch und dem Zusammenführen von medizinischen Informationen darstellt. (SNOMED International, 2023) (Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023d)

Auf der Grundlage dieser bereits vorhandenen Kodierung, auf Basis standardisierter Klassifikationen und Terminologien, setzen diverse medizininformatische Handlungsfelder auf, so auch die Datenintegration im klinischen Bereich. Die

Zusammenführung und Einspeisung von Daten kann idealerweise Anhand definierter Kriterien erfolgen, welches durch die Verwendung einer Klassifikation erleichtert wird. Im Zuge dessen können beispielsweise Daten mit der ICD-10-Kodierung *U07.1!* verglichen, zusammen ausgewertet oder in separate Systeme integriert werden. Dies erfolgt rein anhand der Kodierung ohne Berücksichtigung sonstiger Dokumentationen, welche in der Regel unstrukturiert und ohne Ordnung vorliegen. Zu diesem Zweck existieren Systeme zum Abgleich der Terminologien und zur Validierung der Kodierungen, um eine möglichst hohe Qualität der Daten gewährleisten zu können. (Rector et al., 1995, pp. 147–151) (Rienhoff and Semler, 2015, pp. 115–117) (Bodenreider, 2008, pp. 5–6)

1.6 Stand der Wissenschaft und Technik

Die Datenerfassung in den Bereichen der klinischen Routine und der Forschung unterscheiden sich sowohl standortintern, als auch übergreifend enorm. Dies ist unter anderem auf die unterschiedlichen Infrastrukturen aus kommerziellen und frei verfügbaren Produkten zurückzuführen und hat hohe Auswirkungen auf den Stand der Wissenschaft und Technik für die Forschenden. Hierbei steht primär die Datenerfassung im Vordergrund, welche die Grundlage für die Datenbasis und damit verbundene Metadaten, Analysen und Resultate bildet. Der große Bereich der elektronischen Datenerfassungssysteme (EDC), welcher sich in herstellereigene und kommerzielle sowie open-source Erfassungssysteme unterteilt, kann durch eine bereits initiale Strukturierung der Daten und hintergründige Metadatenextraktion, zur Vorbereitung der Daten für eine folgende einheitlichere Verarbeitung, einen hohen Beitrag für die Wissenschaft liefern. (Shah et al., 2010)

Einen Ansatz für eine strukturierte metadatenbasierte Erfassung und somit die Funktion als EDC System übernimmt beispielsweise REDcap, welches insbesondere im Forschungskontext und dem Bereich der Erstellung von Case-Report-Forms (CRF), mit seinen über 6.377 aktiven Partnern, eine der führenden Rollen übernimmt. (Vanderbilt University, 2023) Aufgrund der initialen Metadaten-Erstellung können die erfassten Daten grundlegend strukturiert und für interne sowie standortübergreifende Analysen bereitgestellt werden. Basierend auf der elektronischen Erfassung werden Eingaben direkt validiert und es können semantische Fehler bei der Erfassung vermieden werden. (Harris et al., 2009) Ebenfalls führend in diesem Bereich verfügt das webbasierte open-source Produkt OpenClinica, insbesondere im Forschungsumfeld, über einen hohen Bekanntheitsgrad und zeichnet sich ebenfalls durch eine strukturierte Datenerfassung aus. OpenClinica selbst basiert primär auf eCRFs (electronic case report form) und

organisiert darüber seine interne Struktur und Funktionalität. Die Erstellung von benutzerdefinierten Formularen und späterer Analyse, Erzeugung von Reports sowie Darstellung der darüber erfassten Daten, kann ebenfalls metadatenbasiert erfolgen. (OpenClinica, LLC, 2023) (Cavelaars et al., 2015) Diesen beiden offenen Lösungen gegenübergestellt stehen beispielsweise noch die beiden kommerziellen Lösungen ResearchManager und trialkit. Das Produkt ResearchManager basiert, wie auch OpenClinica, auf eCRF und bietet einen ähnlichen Funktionsumfang. Ergänzend dazu wird auch der Bereich klinischer Studien und dessen ganzheitlichem Daten- und Ablaufmanagement abgedeckt. (ResearchManager, 2021) Die Software trialkit fokussiert sich primär auf Lösungen im Bereich von Studien und stellt neben einem EDC-System auch ein System für das ganzheitliche Datenmanagement dar. Weiter können sämtliche Prozesse, welche im Zuge einer Studiendurchführung anfallen, wie der Consent, die Randomisierung und die Integration von Geräten, mittels der Plattform abgebildet werden. Aus Gründen der Interoperabilität können die Daten der eCRFs im Anschluss exportiert und in andere Systeme integriert werden. (Crucial Data Solutions, 2023)

Zwar existieren eine enorme Vielzahl an Systemen, welche Metadaten und Metadatenrepositorien einsetzen, verarbeiten und freigeben, wichtig ist jedoch die Grundlage auf welcher diese beiden Themenbereiche basieren. Metadaten selbst können unterschiedlich strukturiert sein und differenzieren sich stark in ihrem Detailgrad. Aus diesem Grund existieren Konventionen zur Beschreibung von Objekten und Sachverhalten, woraus sich die konkreten Metadatenattribute ableiten lassen. Der primäre Standard in diesem Umfeld wird durch Dublin Core repräsentiert, welcher ein Schema zur Beschreibung elektronischer Ressourcen darstellt und konkrete Aussagen über diese Ressourcen trifft. Auf Basis dieser definierten Richtlinien und Hilfestellungen können elektronische Konstrukte und Datenstrukturen detailliert beschrieben werden, wodurch Metadaten generiert werden. (Baker, 2000, pp. 2–4) (Coyle and Baker, 2009)

Die Verwendung von Repositorien zur Verwaltung dieser beschreibenden Informationen von Daten unterliegen zwar keinen verpflichtenden Regularien, jedoch ist ihre interne Struktur durch die ISO/IEC 11179-3:2013 geregelt und definiert konkrete Vorgaben zu dem internen Aufbau. Somit kann eine strukturierte und standardisierte Verwaltung der Informationen innerhalb der Repositorien erfolgen. Das Metadatenrepositorium Sampil.MDR basiert als eines der wenigen MDR-Systeme auf dem genannten Standard und wird im weiteren Verlauf dieser Dissertation detaillierter behandelt. (ISO/IEC JTC 1/SC 32 Data management and interchange, 2013) (Kadioglu et al., 2018, p. 51) Ein weiteres im medizinischen Kontext eingesetztes Metadatenrepositorium, welches auf

dem Standard basiert, stellt caCSR dar. Das System wurde im Rahmen des National Cancer Institute entwickelt und erfasst, verwaltet und harmonisiert Metadaten im Bereich der Krebsforschung. Das System selbst ist Teil des Common Ontologie Representation Environment – caCORE, welches eine interoperable, modellbasierte Infrastruktur darstellt. (Komatsoulis et al., 2008, pp. 1–4) (National Cancer Institute and Warzel, 2023) Ein weiteres etabliertes Werkzeug in diesem Bereich ist die CEDAR Workbench des Center for Expanded Data Annotation and Retrieval, welche seit dem Jahr 2014 ein komplettes Ökosystem für biomedizinische Metadaten repräsentiert und fertige Vorlagen für Datensets und Datenelemente bereitstellt. Das darin enthaltende MDR ist streng nach ISO/IEC 11179 organisiert und bietet somit eine optimale Grundlage für das standardisierte und interoperable Datenmanagement. (National Institutes of Health, 2023) (O'Connor et al., 2020) Nicht zu vernachlässigen sind jedoch auch bereits etablierte Lösungen von Metadatenrepositorien im Bereich der medizinischen Forschung und Versorgung. Aufgrund ihrer gegebenenfalls bereits seit längerer Zeit vorgehaltenen und integrierten Daten, stellen sie einen bedeutenden Baustein der Infrastrukturen dar und werden, trotz fehlendem Standard, auch langfristig verwendet. Hierbei gilt es tranSMART zu betrachten, welches eine Data-Warehouse-Lösung darstellt und im medizinischen Umfeld auch als MDR genutzt werden kann. Ein Vorteil dieser Lösung ist die große Nutzerzahl und open-source Basis, welche das System unter anderem zu einem Kern-Werkzeug, auch im Kontext des MIRACUM-Projektes, deklariert. (Athey et al., 2013) (Fyler, 2023) Dem angebunden gilt es auch i2b2 zu berücksichtigen, welches ebenfalls aus der i2b2 tranSMART Foundation stammt und eine Analyseplattform inklusive klinischem Data-Warehouse bereitstellt. Der Fokus kann hierbei aufgrund der Systemstruktur ebenfalls metadatenzentriert gelegt werden, wodurch die Softwarelösung im Forschungsumfeld weit verbreitet ist und in unterschiedlichsten Formen, wie auch als MDR-Lösung, genutzt wird. (The i2b2 Community et al., 2023) (Murphy et al., 2010, pp. 124–125)

Aus globaler Sicht erfolgt die Verwendung von Metadatenrepositorien und ihren Metadaten in einer weiterhin steigenden Vielzahl von Bereichen, insbesondere auch im Finanzwesen und der Automobilindustrie, wo ein hoher Sicherheitsstandard und klare Strukturen ebenso unerlässlich sind, wie im Gesundheitswesen und der medizinischen Forschung. Die erfassten Daten müssen vor einer Analyse klar strukturiert werden, welches insbesondere im Big-Data Kontext eine große Rolle spielt. Somit finden Metadaten auch im Bereich der Europäischen Zentralbank eine konkrete Anwendung, wie in (Bholat, 2016, pp. 4–6, 23) ersichtlich ist. Komplexe und globale Sachverhalte werden strukturiert und vereinheitlicht. Aus der Perspektive einer industriellen Seite wie

die der Automobilindustrie stehen insbesondere Themen wie die Fertigung und autonomes Fahren im Vordergrund, wodurch eine weitgefaste Datenbasis generiert wird, welche auf unterschiedlichen Strukturen und Informationen basiert. In einem solchen Fall können Metadaten genutzt werden, um beispielsweise Bauteilinformationen genauer zu beschreiben und somit Analysen über die Qualität, Verschleiß und Wartungen zentral durchführen zu können. (Rui et al., 2020, pp. 1–4) Dies lässt sich auch konkret auf das Gesundheitswesen und die Medizininformatik in Kombination mit der Medizintechnik beziehen.

Sofern eine Datengrundlage geschaffen wurde und diese um Metadaten angereichert wurde, kann bereits eine konkretere und verbesserte Verwendung der Daten erfolgen. An dieser Stelle unterscheidet sich jedoch das Datenmanagement der medizinischen Forschung zu den zuvor genannten globalen Bereichen, indem klare Prinzipien zu dem Datenmanagement in Form der FAIR-Prinzipien existieren. Zur Einhaltung der Prinzipien ist eine Anpassung der Datenbasen und somit auch der Repositorien nötig, wodurch ein entsprechender technischer und personeller Anteil benötigt wird. Die FAIR-Prinzipien stellen bei ihrer Verwendung den aktuellen Stand der Technik dar und ermöglichen eine offene strukturierte und interoperable System- und Datenlandschaft. (Jansen, 2016) (Wilkinson et al., 2016, pp. 4–5)

Der aktuelle Stand der Wissenschaft sieht im Bereich des medizinischen Datenmanagements, wie der Strukturierung von Daten und Generierung von Metadaten, den deutschlandweiten Aufbau von Datenintegrationszentren vor, wodurch diese Aufgaben nicht bereichsintern, sondern zentral an einem Standort mit medizinischer Versorgung und Forschung durchgeführt werden sollen. Die Datenintegrationszentren erfüllen bereits jetzt Vorgaben der Bundesministerien zur Datenintegration und standortübergreifenden Analyse, wofür Metadaten und ihre Repositorien zwingend notwendig sind. (Drepper et al., 2017)

Ein weiterer Meilenstein im Zuge der aktuellen Digitalisierung des Gesundheitswesens stellt das im Oktober 2020 erlassene Krankenhauszukunftsgesetz (KHZG) dar, welches die Digitalisierung innerhalb des Gesundheitswesens speziell bei Krankenhäusern stark vorantreiben soll und insbesondere bei der Festlegung förderfähiger Vorhaben mittels §19 Absatz 2 Satz 1 einen konkreten Fokus auf die Gewährleistung von einrichtungsinterner und -externer Interoperabilität legt. Die Grundlage dieser, durch das KHZG geforderten digitalen Vernetzung, dem Datenaustausch und der Datenanalyse, basieren auf festgelegten Standards, welche durch Metadaten angereichert werden. Der Bereich der Metadaten wird dabei zur Strukturierung, Nachvollziehbarkeit der

Datenherkunft, Qualitätssicherung und weiterer internen Datenverarbeitung und Datenaufbereitung herangezogen. (Bundestag, 2020) (Henke et al., 2022, pp. 223–225, 285–287)

1.7 Rechtliche und ethische Aspekte

Im Zuge dieser Themenkomplexe wird aus rechtlicher Sicht primär der Datenschutz und die damit verbundene Datenschutzgrundverordnung (DSGVO) sowie das Bundesdatenschutzgesetz (BDSG) betrachtet. Grundsätzlich gilt die Annahme, dass eine Analyse medizinischer Daten, welche durch eine Datenlieferung erhalten wurden, im Umkehrschluss auch ein Profit für die entsprechenden Datenlieferanten darstellt. Hierbei gilt es jedoch stets die Rechte der Datenlieferanten, die Patienten / Probanden, zu wahren. (Landrock and Gadatsch, 2018, p. 63) Insbesondere im Kontext von Metadaten ist die Referenz auf den Ursprung der Daten und somit auf eine reale Person für die Analyse und die Ergebniserückführung von großer Bedeutung. Die drei großen Vorgaben, welche in diesen Prozessen stets berücksichtigt werden müssen, stellen dar:

1. Die Patienteneinwilligung
2. Die Anonymisierung von Daten
3. Die Pseudonymisierung von Daten

Die Patienteneinwilligung spiegelt die Bereitschaft des Patienten zum Nutzen seiner personenbezogenen Daten wider und bedingt die grundsätzliche Verarbeitung seiner Daten. An dieser Stelle ist zu berücksichtigen, dass der Widerruf der Einwilligung jederzeit und im vollen Umfang erfolgen kann. (Das Europäische Parlament und der Rat der europäischen Union, Art. 17 Nr. 1) (Janda, 2019, pp. 381–382) Medizinische Informationssysteme müssen mit dieser Gegebenheit auf technischer Ebene umgehen können, was insbesondere im Bereich der medizinischen Informatik stets berücksichtigt wird und in der Regel durch Architekturen wie „Data protection by design & by default“ direkt bei der Softwareentwicklung umgesetzt werden kann. (Darms et al., 2019, pp. 160–161)

Zur Wahrung der Rechte einer Person wird primär versucht aus Sicht eines Forschungsdatenzentrums keine patientenidentifizierenden Merkmale herauszugeben. Bereits die Herausgabe der Daten selbst unterliegt rechtlichen Vorgaben und kann nicht grundsätzlich erfolgen. (dtv Verlagsgesellschaft, 2021, p. 928, SGB V §303e Absatz 3) Eine Herausgabe muss somit in einem berechtigten Falle anonymisiert und aggregiert erfolgen. Zur Anonymisierung selbst können unterschiedliche Verfahren angewandt

werden, wodurch beispielsweise durch K-Anonymität der Grad der daraus resultierenden Datenanonymisierung angepasst werden kann. (Landrock and Gadatsch, 2018, p. 65) Ziel der Anonymisierung ist es konkret, sämtliche Verweise auf patientenidentifizierende Merkmale zu entfernen und eine Rückführbarkeit unmöglich zu machen. (Kushida et al., 2012) Speziell dieser Aspekt ist jedoch im Rahmen der Forschung, beispielsweise im Kontext longitudinaler Datensätzen, welche eine Kombination an Datensätzen aus unterschiedlichen Datenquellen darstellt, ein schwieriges Vorhaben und bedingt den Einsatz spezieller Werkzeuge und Verfahren. (Tamersoy et al., 2012) Eine Umsetzung dieser Aspekte kann aus Sicht einzusetzender IT-Systeme einen undefinierten Umfang vorweisen. Dies ist primär durch den Umfang und die Komplexität der Daten bedingt. Grundsätzlich kann jedoch ein Forschungsdatenzentrum freier mit anonymisierten Datensätzen arbeiten, da lediglich die Anonymisierung gegeben sein muss und die Nutzungsbedingungen laut SGB V, §303e Absatz 2 erfüllt sein müssen. Im Falle einer gewünschten Rückführbarkeit und Nachvollziehbarkeit der Daten auf eine Person muss die Anonymisierung gegebenenfalls durch eine Pseudonymisierung ersetzt werden, wodurch Personen und ihre Daten einen eindeutigen Identifikator, das sogenannte Pseudonym, erhalten, welcher jedoch nur durch bestimmte Verfahren und unter definierten Bedingungen rückführbar sein darf.

Die Pseudonymisierung von Daten unterliegt nicht nur im Zuge der Datenverarbeitung und Datenweitergabe diversen gesetzlichen Regularien, sondern auch die Generierung und Herleitung des Pseudonyms selbst. Primär erfolgt eine Pseudonym-Generierung durch eine entsprechend definierte Vertrauensstelle, welche basierend auf Vorgaben und Regularien des Bundesamtes für Sicherheit in der Informationstechnik (BSI) spezielle Verfahren zur Generierung einsetzt. Hierbei darf insbesondere eine direkte Rückführung auf die Person hinter dem Pseudonym nicht realisierbar sein. (dtv Verlagsgesellschaft, 2021, p. 926, SGB V §303c Absatz 2) Diese Vorgaben betreffen jedoch nicht nur die interne Datenverarbeitung innerhalb der IT-Systeme, sondern auch die Heraus- und Weitergabe von konkreten Datensätzen. So darf ein Forschungsdatenzentrum, sofern ein nachvollziehbarer Nutzungszweck nach innerhalb des Sozialgesetzbuch V §303e Absatz 2 definierten Kriterien vorliegt, unter Berücksichtigung strenger Vorgaben, welche durch den Standort umzusetzen sind, eine Datenherausgabe einleiten. (dtv Verlagsgesellschaft, 2021, p. 928, SGB V §303e Absatz 4) Die Vorgaben beziehen sich insbesondere auf die Art und den Umfang der Datenbereitstellung im Hinblick auf die Einschränkung der Datenfreigabe für Dritte und Unbefugte im Sinne der Geheimhaltungspflicht. Weiter müssen konkrete

Nutzungszwecke definiert sein und die Nutzungsberechtigten dürfen, mittels der Daten, unter anderem keinen Bezug zu den real existierenden Personen hinter den Pseudonymen herstellen. (dtv Verlagsgesellschaft, 2021, p. 928, SGB V §303e Absatz 5)

Ein weiterer wichtiger Komplex wird durch die Ethik realisiert. Diese verfügt über einen großen Einfluss auf Forschungsdatenzentren beziehungsweise Datenintegrationszentren, da in der Regel nicht nur eine, sondern eine Vielzahl an komplexen Auswertungen über unterschiedlichste Datenbasen zum Ziel der Gewinnung unterschiedlichster Resultate erfolgt. Die dadurch generierte Menge an Auswertungs- und Interpretationsmöglichkeiten kann ein solches Datenzentrum nicht prinzipiell ethisch befürworten. Stattdessen muss jedes Projekt einer separaten ethischen Begutachtung unterzogen werden, was durch lokale Ethikkommissionen erfolgt. Grundsätzlich müssen die Ziele der Auswertung stets klar definiert und vor allem nachvollziehbar sein. Dazu zählt auch eine volle Transparenz des Projektvorhabens, der Patienteneinwilligung und die Zweckvereinbarkeit, welche noch einmal separat in der DSGVO Art. 5 I b rechtlich geregelt ist. (Friele et al., 2020, pp. 744–746) Speziell im Hinblick auf die Komplexität medizinischer Forschungsdaten gilt es die ethischen Aspekte auch stets im Gesamtkontext zu betrachten und die Folgen der Analysen abzuschätzen. So könnten fehlende ethische Bedenken, beispielsweise bei epigenetischen Auswertungen, im späteren Verlauf für eine konkrete Stigmatisierung oder Diskriminierung bestimmter Gruppen sorgen, was es strengstens zu vermeiden gilt. (Santaló and Berdasco, 2022)

Sowohl die rechtlichen als auch die ethischen Aspekte bedingen eine am Standort vorhandene organisatorisch-rechtliche Struktur, welche die beiden Themenbereiche für Forschende, Patienten und Probanden transparent abbilden und ein Forschungsdatenzentrum rechtlich, in Form einer Vertragsgrundlage, absichert.

1.8 Motivation

Das sich im Aufbau befindliche Gießener Datenintegrationszentrum wird im Zuge der Medizininformatik-Initiative die zentrale Anlaufstelle für Informationen rund um Forschungsdaten und Routinedaten des medizinischen Bereiches am Standort Gießen. Im klassischen Sinne sollen Möglichkeiten zur Datenanalyse sowie die ganzheitliche Durchführung von Forschungsprojekten auf der Datenbasis innerhalb des DIZ, welche zukünftig den vollständigen Datenbestand des Standortes enthalten soll, durchgeführt werden. Dies bedingt grundsätzlich eine Integration von Routinedaten und Forschungsdaten in eine gemeinsame Datenbasis, jedoch wird ein Fokus auf die

Struktur und Organisation der Daten benötigt. An dieser Stelle spielen die Metadaten eine primäre Rolle, welche für die Organisation der Daten sorgen und die Suche nach spezifischen Vorkommen in der enorm großen Datenlandschaft unterstützen können.

Die Metadaten sollten in Form eines Metadatenrepositoriums gelistet, verwaltet und abgefragt werden. Die Verwaltung ist für die Aktualität der Daten von großer Rolle, welche für die medizinische Forschung unabdingbar ist. Die Abfragemöglichkeiten sollen im Hintergrund nicht nur eine Interoperabilität der Daten in der IT-Landschaft des Standortes gewährleisten, sondern auch direkt für die Forschenden im Kontext eines zuvor gestellten Nutzungsantrages für Machbarkeitsanalysen, im Zuge von Forschungsprojekten, bereitgestellt werden.

Ein am Standort zu etablierendes Metadatenrepositorium kann somit Forschende bereits in der Antragsstellung von Forschungsvorhaben unterstützen, indem konkrete Zahlen zu Vorkommen in der klinischen Versorgung bereitgestellt werden können. Einer der wichtigsten Nutzen hierbei wäre eine klare Zeitersparnis für Forschende bei einer initialen Machbarkeitsanalyse, da die Anfragen an die klinische EDV-Abteilung und die einzelnen Fachabteilungen entfallen. Weiter entfallen für die Forschenden die großen Themen des Datenschutzes und der Datensicherheit bis zu dem Punkt der Ergebnisübermittlung aus der Machbarkeitsabfrage, da das reine Datenmanagement die Aufgabe des Datenintegrationszentrums unter Berücksichtigung lokaler rechtlicher und organisatorischer Rahmenbedingungen darstellt.

Die Motivation dieser Dissertation begründet sich primär auf die Möglichkeit der Verbesserung und Optimierung der Datenexploration für Forschende am Fachbereich Medizin im Kontext von Machbarkeitsabfragen und der Erarbeitung von Förderanträgen für Forschungsvorhaben. Ebenfalls folgt daraus die einheitliche, nachvollziehbare und technisch belegbare Bereitstellung von Informationen für die beteiligten Akteure, der Ethikkommission und des Datenschutzes, wodurch die internen Arbeitsabläufe optimiert werden können. Weiter kann das System den internen Informationsfluss, unter den IT-Systemen des DIZ und sonstigen Systemen der klinischen Forschung, unterstützen. Es dient somit als Bindeglied, in der größtenteils inhomogenen Systemlandschaft und kann für die Organisation, der enorm großen Datenbereiche, einen Beitrag zur Strukturierung leisten.

2. Zielsetzung

Das Datenintegrationszentrum Gießen benötigt für die Durchführung der Machbarkeitsabfragen im Forschungskontext zunächst eine detaillierte Behandlung der rechtlichen Aspekte zur Wahrung des Datenschutzes sowie der Rechte der Patienten und Probanden in Anlehnung an die DSGVO und das Sozialgesetzbuch V (SGB V). Die Generierung und Etablierung eines rechtlichen fundamentalen Konstruktes, welches sowohl die organisatorischen Aspekte des Datenintegrationszentrums, die Begutachtung und Freigabe von Forschungsanfragen sowie die allgemeinen Richtlinien und Vorgaben zur Nutzung von Daten am Fachbereich Medizin im Kontext der medizinischen Informatik abdeckt, stellt eine große Herausforderung dar. Die Rahmenbedingungen gilt es in Form von Ordnungen am Standort zu etablieren.

Im Vorfeld erfolgte eine Analyse der am Standort zentral und für Forschungsvorhaben verfügbaren Metadatenrepositorien mit Fokus auf das Vorhandensein von Routine- und Forschungsdaten. Hierbei stand primär und ausschließlich das klinische Routinearchiv Synedra.AIM im Fokus, welches jedoch nicht über die benötigten Metadaten und Semantik im Kontext der Klassifikationen und Terminologien verfügt und somit nicht für das Vorhaben verwendet werden kann. Aus diesem Grund wird ein neues und zentrales System benötigt, welches diese Aspekte zunächst in einer Basisvariante abdecken kann.

Das zu entwickelnde metadatenbasierte System zur Datenaufbereitung und -verwaltung muss sich nahtlos in die bestehende Infrastruktur und die technischen sowie organisatorischen Abläufe des Datenintegrationszentrums einfügen. Hierbei spielt die Kompatibilität mit anderen Systemen der IT-Landschaft des Standortes eine wichtige Rolle. Es ist essenziell, dass das neue System über Kommunikationsschnittstellen verfügt, um eine nahtlose Integration in die Infrastruktur des Standortes zu ermöglichen und somit eine hart gekapselte Insellösung zu vermeiden.

Die Kommunikationsschnittstellen sollten dabei nach Industriestandards entwickelt werden, um eine maximale Interoperabilität mit anderen Systemen zu gewährleisten. Beispielsweise sollten die Schnittstellen auf offenen Protokollen und Standards wie REST, JSON und XML basieren, um eine einfache und standardisierte Integration mit anderen Systemen zu ermöglichen.

Es gilt sicherzustellen, dass das neue System sämtliche erforderliche Sicherheitsmaßnahmen umsetzt, um die Vertraulichkeit, Integrität und Verfügbarkeit der Daten zu gewährleisten. Dies muss auf Basis des IT-Grundschutzes des Bundesamtes für

Sicherheit in der Informationstechnik (BSI) basieren, um eine Evaluation, auch im Zuge späterer Prüfungen, gewährleisten zu können. Im gleichen Zuge muss der zu verwendende Technologie-Stack stets aktuell und für den Einsatz in der klinischen Infrastruktur, Routine und Forschung, geeignet sein.

Zusätzlich sollten organisatorische Aspekte berücksichtigt werden, um eine erfolgreiche Integration des neuen Systems zu befördern. Die Grundlage hierfür bieten Vorgaben, welche aus dem zu erstellenden rechtlichen Konstrukt resultieren. Hierbei sollten konkrete Abläufe, die Zuständigkeiten und Rollen der beteiligten Personen und Abteilungen, sowie Zugriffsrechte und Zeitpunkte der Zugriffe klar definiert werden können. Weiter sollen klare Prozesse und Verfahren erstellt werden, um den Betrieb und die Wartung des neuen Systems sicherzustellen. Des Weiteren muss das System stets den gesetzlichen und regulatorischen Anforderungen entsprechen.

Die Einbindung des Systems sowie das rechtliche Konstrukt muss einem, durch das BMBF beauftragten, Audit eines Wirtschaftsprüfungsunternehmens bestehen und weiterhin anpassbar für sämtliche Änderungen im Kontext des Datenschutzes, der Datensicherheit und den standortspezifischen rechtlich-organisatorischen Aspekten sein.

Zusammengefasst stellt das Ziel der Dissertation die Generierung rechtlich-organisatorischer Ordnungen am Standort Gießen dar, welche die Grundlage für die Arbeit des Datenintegrationszentrums und daran angelehnte Analysen ermöglicht. Weiter soll ein Metadatenrepositorium entwickelt werden, welches initiale Machbarkeitsanalysen auf den klinischen Datenbestand vollziehen, sich in die Betriebsabläufe und Prozesse des Datenintegrationszentrums eingliedert und als Informationsgeber, zwischen mehreren Systemen in der IT-Landschaft des Standortes in Form eines Bindegliedes, fungiert.

3. Material und Methoden

Nachfolgend werden die für das Vorhaben verwendeten Materialien und Methoden erläutert. Zu Beginn erfolgte eine Darstellung der initial identifizierten Lasten durch die Betreuer dieser Dissertation sowie der Leitung des Datenintegrationszentrum und des MIRACUM-Projektes am Standort Gießen. Auf dieser Basis gilt es im weiteren Verlauf ein Pflichtenheft zu erstellen, welches eine Identifikation der benötigten Materialien und Methoden zwingend voraussetzt. Aus dem Pflichtenheft resultiert weiter ein konkreter Projektplan, welcher sämtliche benötigten Prozesse mit den Materialien in Relation setzt und den zeitlichen Ablauf verdeutlicht. Die Dokumente sind Teil der Projektplanung und werden in Kapitel 4.1 im Detail behandelt.

Sämtliche im Zuge dieser Dissertation verwendeten Systeme wurden auf der Infrastruktur des Datenintegrationszentrums innerhalb des Gießener Forschungsnetzes (GIFONET) installiert und konfiguriert. Eine schnelle und konkrete Übersicht ist in Abbildung 4 ersichtlich.

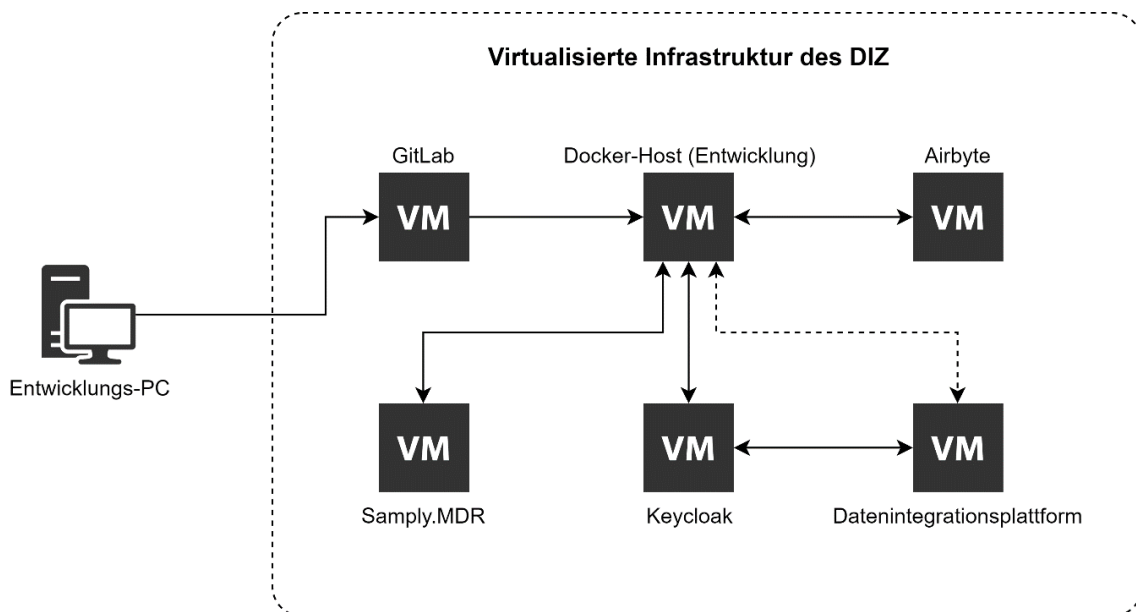


Abbildung 4: Systemstruktur innerhalb der DIZ-Infrastruktur (Eigene Darstellung)

Die beteiligten Systeme und Dienste werden in den nachfolgenden Unterkapiteln beschrieben und das konkrete Zusammenspiel sowie die Funktionsweisen anschließend im Kontext der Ergebnisse behandelt.

3.1 PostgreSQL und JSON

Bereits in den 1980er Jahren erfolgte im Zuge eines Projektes der University of California at Berkeley die Entwicklung eines Datenbank Management Systems (DBMS), welches auf dem zuvor 1975 bis 1977 ebenfalls hauseigen entwickelten INGRES DBMS aufsetzt. Dieses neue System wird durch PostgreSQL repräsentiert (POST inGRES), welches mittlerweile in der Version 15 vorliegt. (Stonebraker and Rowe, 1986, p. 340) (Fröhlich, 2018, p. 2)

Das DBMS steht als open-source System zur Verfügung und zählt seit vielen Jahren weltweit zu den beliebtesten Datenbanksystemen. Derzeit befindet sich das System laut DB-Engines auf Platz 4 der meist verwendeten DBMS Systeme. (solidIT, 2022) (Cubukcu et al., 2021, p. 2490) Hierbei ist auch der medizinische Bereich eingeschlossen, speziell aufgrund der vollständigen Unterstützung und Abbildbarkeit des HL7-Standards Version 3. (Havinga et al., 2010) (Austin et al., 2015, pp. 327–328)

PostgreSQL ist objektrelational organisiert und stellt somit logische Beziehungen und Verbindungen zwischen Tabellen und ihren Daten mittels Primär- und Fremdschlüsseln dar. Basierend auf einem Großteil des SQL Standards ISO/IEC 9075 bietet das System auch die Unterstützung von Transaktionen, welche insbesondere bei Systemen mit multiplen Verbindungen aus Sicht der Datenkonsistenz unerlässlich sind. Weiter besteht die Möglichkeit, abgesehen von alphanumerischen Werten innerhalb der Tabellen, auch JSON-Objekte hinterlegen zu können. (Stinson, 2001, pp. 36–37) (The PostgreSQL Global Development Group, 2022) (Douglas and Douglas, 2003, pp. 1–2)

JSON-Objekte (JavaScript Object Notation) stellen ein menschenlesbares und leichtgewichtiges Datenaustauschformat dar, welches auf den Konzepten des *key-value* Schemas und einer Hierarchie aufbauen. Letzteres kann durch unterschiedliche Modelle umgesetzt werden wie Arrays oder Listen. Die Objekte verfügen somit stets über eine geordnete Struktur und können somit auch automatisiert verarbeitet werden. (Ecma International, 2017)

Ein Beispiel für ein JSON-Schema ist in Abbildung 5 ersichtlich und verdeutlicht die Struktur eines Objektes.

```

{
  "patient": {
    "fallnummer": 2022010517,
    "name": "Doe",
    "vorname": "John",
    "geburtsdatum": "1990-01-01"
  }
}

```

Abbildung 5: Beispielhafte Struktur eines JSON-Objektes (Eigene Darstellung)

Durch unterschiedliche bereitgestellte Funktionen bietet PostgreSQL nativ auch die Möglichkeit relational integrierte Daten in JSON-Objekte zu konvertieren oder umgekehrt diese Objekte intern zu verarbeiten und relational zu hinterlegen. (The PostgreSQL Global Development Group, 2018)

Die Architektur von PostgreSQL unterstützt ebenfalls ein komplexes Rollen-, Benutzer-, Gruppen- und Rechtemanagement, wodurch Sichten auf Daten eingeschränkt, überwacht und protokolliert werden können. Weiter erfolgt die Bereitstellung sogenannter Views, welche die Möglichkeit besitzen, Daten aus Tabellen in einem Zwischenschritt zu abstrahieren. Views stellen virtuelle Tabellen dar, welche aus Datenbank-Abfragen erstellt wurden, um große Datenmenge im Vorfeld einzuschränken, zusammenzufassen oder besser zu strukturieren. (Studer, 2019, p. 121) Ein Beispiel für eine solche View ist in Abbildung 6 ersichtlich.

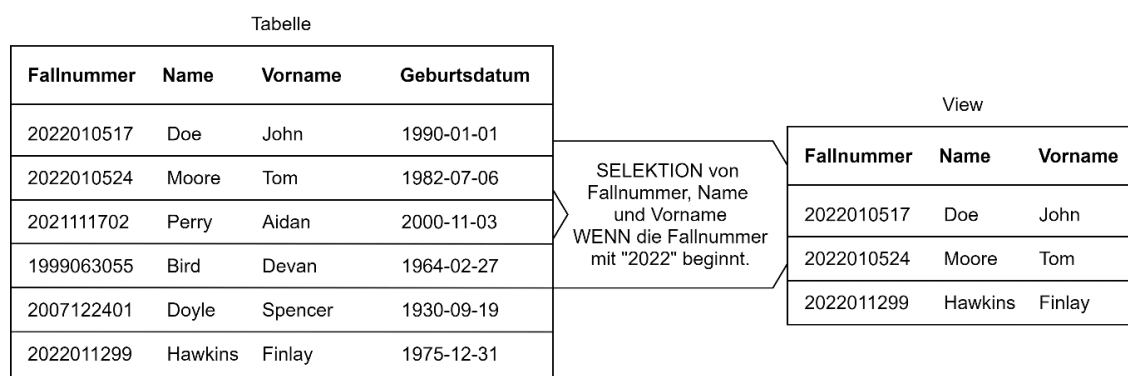


Abbildung 6: Darstellung der Erzeugung einer PostgreSQL-View (Eigene Darstellung)

Die Abbildung verdeutlicht, wie Daten einer Tabelle konkret eingeschränkt werden können, wodurch die Geschwindigkeit der Abfrage auf die View gesteigert werden kann. Im Zuge der Dissertation wurde PostgreSQL in der Version 12.11 verwendet.

3.2 Softwareentwicklung

Die im Zuge dieser Dissertation verwendeten Programmiersprachen, Entwicklungsumgebungen und Frameworks werden in den nachfolgenden Unterkapiteln detailliert behandelt.

3.2.1 Python und Flask

Python stellt seit der Entwicklung im Jahre 1989 eine interpretative und gehobene Programmiersprache dar, welche auch funktionale und objektorientierte Konzepte unterstützt. Der Python Interpreter verarbeitet die erstellten Python-Skripte, welche plattformunabhängig ausgeführt werden können. Die Sprache selbst bietet im Vergleich zu anderen Sprachen einige Vorteile, insbesondere in ihrer vereinfachten Syntax und der Möglichkeit Bibliotheken einzubinden. Insbesondere im Python-Umfeld existieren für eine Vielzahl an Anwendungsfällen kleinere Erweiterungen, welche zu einem großen Gesamtbild schnell zusammengefügt werden können. Insbesondere für konkrete Anwendungsfälle existieren etablierte Frameworks, welche bereits eine Struktur und Erweiterungen für die Python-Skripte vorgeben, wodurch eine schnelle Entwicklung erfolgen kann. (Weigend, 2020, pp. 13–14) (Kaminski, 2016, pp. 3–5)

Eines dieser (micro) Frameworks wird durch Flask repräsentiert, welches unterschiedliche Erweiterungen nutzt, um eine Web-Entwicklung zu ermöglichen und zu vereinfachen. Es können nicht nur umfangreiche Websites generiert, sondern auch reine API-Endpunkte entwickelt und performant genutzt werden. Das Framework selbst stellt sämtliche benötigten Erweiterungen bei der Installation, welche durch den Python-eigenen Paketmanager pip erfolgt, bereit. Dies bezieht sich zunächst auf die Jinja2 Template Engine und dem WSGI Toolkit Werkzeug. Alle weiteren benötigten Funktionalitäten können über Erweiterungen und den Paketmanager nachinstalliert werden. Damit ist Flask deutlich vereinfachter als das vollumfängliche Pendant im Kontext der Webentwicklung Django. (Singh et al., 2019, p. 24895) (Maia, 2015, pp. 1–2)

Durch die native Unterstützung von Unit-Tests, Secure-Cookies und der Dokumentierbarkeit des Systems bietet Flask nicht nur eine schnelle, sondern auch sichere und nachvollziehbare Entwicklungsgrundlage. Des Weiteren stellt die finale Bereitstellung einer entwickelten Flask-Applikation mittels einer virtuellen Umgebung (virtualenv) eine Lösung für Produktivumgebungen dar. Die virtuellen Umgebungen werden gekapselt von der Standard Python Umgebung gestartet und können somit von

deren Entwicklungen und Anpassungen nicht beeinflusst werden. Dies ermöglicht die Verwendung diverser Erweiterungen in unterschiedlichen Versionen, ohne Konflikte zu verursachen. Neue Versionen von Erweiterungen können somit parallel zur produktiven Web-Applikation getestet und erst bei Erfolg ersetzt werden. (Aggarwal, 2019, p. 10) (Vyshnavi and Malik, 2019, pp. 16–17)

Diese Dissertation verwendet im Kontext der Backend-Entwicklung Python in der Version 3.10.6 und Flask in der Version 2.2.1.

3.2.2 PyCharm

PyCharm ist eine seit 2010 veröffentlichte integrierte Entwicklungsumgebung (IDE), welche für die Verwendung und Entwicklung von Python ausgerichtet ist. Die IDE ermöglicht das Schreiben von Code in einer hohen Qualität durch kontinuierliche Überprüfung, Refaktorisierung und Restrukturierung. PyCharm erhält durch die Steigerung an Produktivität, beispielsweise durch die Bereitstellung von automatischen Code-Vervollständigungen, sowie der Verwaltung und Einbindung von Erweiterungen, insbesondere auch im wissenschaftlichen Bereich, eine hohe Bedeutung. Diese führt dazu, dass die IDE im Kontext von Python eine der meist verwendeten IDEs weltweit darstellt. (JetBrains s.r.o., 2022a) (Nguyen, 2019, pp. 8–9, 13–15)

PyCharm stellt ebenfalls Funktionen zur Einbindung und Verwaltung von Datenbanken bereit, wodurch eine Backend-Entwicklung im Hinblick auf diese komfortabler gestaltet wird. Der innerhalb der IDE entwickelte Code kann direkt gestartet, getestet und sogar im Bereich der Web-Entwicklung auf einem Webserver gestartet werden. Ebenfalls stellt PyCharm einen integrierten Git-Klienten dar, welcher zur Vereinfachung der Versionsverwaltung genutzt werden kann. (Nguyen, 2019, pp. 147–148) (JetBrains s.r.o., 2022b)

Verwendet wird PyCharm im Rahmen der Dissertation zur Backend-Entwicklung in der Professional-Edition, Version 2022.2.

3.2.3 Python Erweiterungen

Im Zuge der Backend-Entwicklung wurden folgende Python Bibliotheken als Erweiterungen verwendet:

psycopg2 in Version 2.9.1. Die Erweiterung stellt ein Adapter für PostgreSQL-Datenbanken im Kontext von Python dar, wodurch SQL-Befehle mittels Python an das

DBMS gesendet und das resultierende Ergebnis weiterverarbeitet werden kann. (Varrazzo, 2021)

python-dotenv in Version 0.19.0. Die Entwicklung ermöglicht das Auslesen bestimmter Variablen-Paare (sogenannte key-value Paare) aus einer separaten Datei für Umgebungsvariablen. Diese Variablen können im Anschluss innerhalb der Code-Umgebung verwendet und verarbeitet werden. (Kumar, 2022)

jsonschema (Version 3.2.0) bietet einen Funktionsumfang zur Erstellung und Validierung von JSON-Konstrukten, wodurch Plausibilitäten geprüft und Nachrichtenformate strikt eingehalten werden können. (Berman, 2013)

plotly in Version 5.2.2. Im Zuge des Backends wird plotly zur Validierung vorbereiteter Diagramm-Definitionen verwendet, bevor die Daten an das Frontend übergeben werden. (Plotly, 2022a)

pandas (Version 1.3.2) dient der Vorbereitung der Diagramm-Definitionen und Nutzung des plotly Funktionsumfangs. (the pandas development team and Plotly, 2022)

Requests in Version 2.26.0 stellt eine übersichtliche und einfache Bibliothek zur Verwendung des HTTP/1.1 Protokolls dar. Requests bildet die Grundlage des Backends zur Nutzung von API-Endpunkten. (Reitz, 2022)

Flask-CORS in Version 3.0.7 nutzt der Ermöglichung von CORS-Anfragen (Cross Origin Resource Sharing), welches insbesondere in der initialen Entwicklung des Gesamtsystems auf lokalen Endgeräten eine Kommunikation zwischen dem Front- und Backend ermöglicht. (Dolphin, 2013)

Flask-Caching (Version 1.10.1) bietet eine Cache-Funktion für ein bereits existentes Flask-Backend. Diese Erweiterung stellt eine Sicherheitsmaßnahme dar, um sämtliche API-Aufrufe ohne Datenverlust abarbeiten zu können. (Burgess and Justin, 2016)

Flask-RESTful (Version 0.3.9) ist ebenfalls eine Erweiterung für ein bestehendes Flask-Backend und ermöglicht die schnelle Generierung von API-Endpunkten auf Basis von REST (Representational State Transfer). (Burke et al., 2020)

3.2.4 Node.js, Vue.js und Vuetify

Node.js stellt eine umfangreiche, auf JavaScript basierende, Umgebung für schnelle und skalierbare serverseitige Netzwerkanwendungen zur Verfügung. Die dadurch im Vordergrund befindliche Entwicklung für Web-Applikationen arbeitet nach dem

Grundsatz von Node.js eventbasiert und asynchron. Das System bietet eine eigene Laufzeitumgebung, welche die Events verwaltet und auch eine threadbasierte Arbeitsweise ermöglicht. (Sun et al., 2018, p. 196) (Ojamaa and Düüna, 2012, pp. 348–349) (Tilkov and Vinoski, 2010, p. 80)

Auf diese Laufzeitumgebung und Funktionen für die Generierung von netzwerkfähigen Anwendungen setzt Vue.js auf, welches für die Entwicklung von Benutzeroberflächen, den sogenannten Frontends, verwendet wird. Vue.js repräsentiert ein leichtgewichtiges Framework für JavaScript und verfügt über eine eigene CLI, welche eine schnelle Entwicklung und das Management von Web-Applikationen ermöglicht. (Rojas, 2019, pp. 10–11) (Peterke, 2019, p. 1) (Passaglia, 2017, p. 7,8)

Neben Funktionen wie bedingtes Rendern von Objekten, Unterstützung von umfänglichen Formularen und der Bereitstellung von Animationen, können weitere benötigte Funktionen performant über Node.js Erweiterungen nachinstalliert werden. Die Erweiterung axios nimmt im Kontext der Nutzung von API-Endpunkten eine primäre Rolle ein, da durch sie ein HTTP-Client bereitgestellt und somit eine Kommunikation innerhalb der Web-Endpunktumgebung ermöglicht wird. (Zabriskie, 2022) (Steyer, 2019, pp. 109, 161, 197)

Auf Grundlage des Vue.js Frameworks erfolgt im nächsten Schritt die Notwendigkeit nach Design-Optionen zur einheitlichen und übersichtlichen Darstellung der generierten Benutzeroberflächen-Objekte. Die Erweiterung Vuetify ermöglicht die Verwendung von materiellen Design-Elementen und Richtlinien, wodurch einheitliche und simple Benutzeroberflächen erstellt, beziehungsweise angepasst werden können. Die Design-Elemente stellen beispielsweise Tabellen, Buttons, Fortschrittsbalken und vordergründig auch Seitenbereiche und Seitenaufteilungen dar. (Leider et al., 2022) (Rojas, 2019, p. 11) (Orlivskiy et al., 2021, p. 569)

Im Zuge der Frontend-Entwicklung wird Node.js (Version 16.16.0) mit dem Vue.js Framework (Version 2.7.8) und dem Material-Design Framework Vuetify (Version 2.6.8) verwendet. Für die Kommunikation zwischen dem Frontend und den API-Endpunkten des Backend wird die Erweiterung axios in der Version 0.27.2 verwendet.

3.2.5 Visual Studio Code

Das open-source Microsoft und Community Projekt Visual Studio Code stellt neben einem klassischen Code-Editor auch eine Entwicklungsumgebung dar, welche durch eine Vielzahl an Erweiterungen, je nach Anwendungsfall, in ihrem Umfang gesteigert

werden kann. Das plattformunabhängige Projekt bietet sämtliche funktionale Unterstützung zur Entwicklung von Code, wie beispielsweise Syntax-Prüfungen und Hervorhebungen. Weiter können die entwickelten Komponenten direkt innerhalb des Systems, je nach verwendeter Programmiersprache und -umgebung gestartet (build) und schrittweise auf Fehler überprüft (debug) werden. (Microsoft, 2022a)

Das Projekt unterstützt vollumfänglich Node.js und JavaScript, wodurch eine Kode-Vervollständigung, Überprüfung und Dokumentation gewährleistet werden kann. Ebenfalls existieren Erweiterungen für die Verwendung von Vue und die Möglichkeit der Nutzung der Vue-CLI. (Microsoft, 2022b) (Microsoft, 2022c) (Vue.js Contributors, 2020)

Visual Studio Code wird im Zuge der Entwicklung der Benutzeroberfläche (Frontend) in Version 1.70 verwendet.

3.2.6 Plotly

Das im Jahr 2013 veröffentlichte Projekt auf open-source Basis dient der webbasierten Visualisierung von Daten im Kontext von Python, welches dann bereits 2015 die weltweit meist verwendete Bibliothek zur grafischen Darstellung wurde. Plotly.js stellt die JavaScript Bibliothek dar, welche aufgrund ihrer Konfigurierbarkeit und Funktionen eine Vielzahl an Darstellungsmöglichkeiten bietet. Neben Scatter-Plots, Torten- und Balkendiagrammen können auch weitere statistisch relevante Diagramme, wie Box-Plots und Histogramme bereitgestellt werden. (Plotly, 2022b) (Plotly, 2022c) (Unpingco, 2021, pp. 251–253)

Die Bibliothek wird derzeit von vielen anderen Frameworks und Sprachen unterstützt, so existiert auch eine Vue.plotly Erweiterung zur Einbindung der Graphen und Diagramme in eine Vue.js Umgebung. Die Erweiterung fungiert als ein sogenannter Wrapper, welcher sämtliche Funktionen, der zu Grunde liegenden Bibliothek, programmatisch verwendbar gestaltet. (Ulrich and Massinger, 2022)

Plotly.js selbst stellt für Anwender bereits bei der Visualisierung Funktionen zum Speichern, Fenstern und Hervorheben zur Verfügung, wodurch selbst bei großen Graphen Ausschnitte erzeugt und separat exportiert werden können. Die Verwendung von Plotly, in Kombination mit großen wissenschaftlichen Datenkomplexen, ist beispielsweise in (von Kurnatowski et al., 2021), (Ali et al., 2016) und (Liu et al., 2021) ersichtlich.

Die Bibliothek wird im Zuge der Dissertation zur graphischen Darstellung von Suchergebnissen innerhalb des Frontends unter Version 1.0.0 verwendet.

3.2.7 Versionsverwaltung Git

Systeme zur Versionskontrolle, sogenannte VCS (version control systems), sind für Arbeiten im Bereich der Softwareentwicklung und der Forschung von einem hohen Stellenwert und zählen mittlerweile zu den essenziellen Fähigkeiten, der in diesen Bereichen handelnden Akteuren. Die Systeme dienen dazu Änderungen an Dateien, Quellcode oder sonstigen Daten zu erfassen, eine Historie bereitzustellen und jeder Zeit einen konkret benötigten Stand wiederherzustellen und verwenden zu können. Die Historie bezieht sich dabei nicht nur auf Änderungen innerhalb der Daten, sondern auch auf die Verwaltungsdaten wie das Änderungsdatum, die beteiligten Akteure an der Änderung und gegebenenfalls deren Rolle im vorliegenden Kontext. (Zolkifli et al., 2018, p. 414) (Haaranen and Lehtinen, 2015, p. 88)

Ein etabliertes System in diesem Bereich stellt die 2005 veröffentlichte open-source Lösung Git dar. Die Lösung stellt ein konkretes Source-Code-Management (SCM) zur Verfügung und kann sowohl lokal als auch zentral Daten übertragen und versionieren.

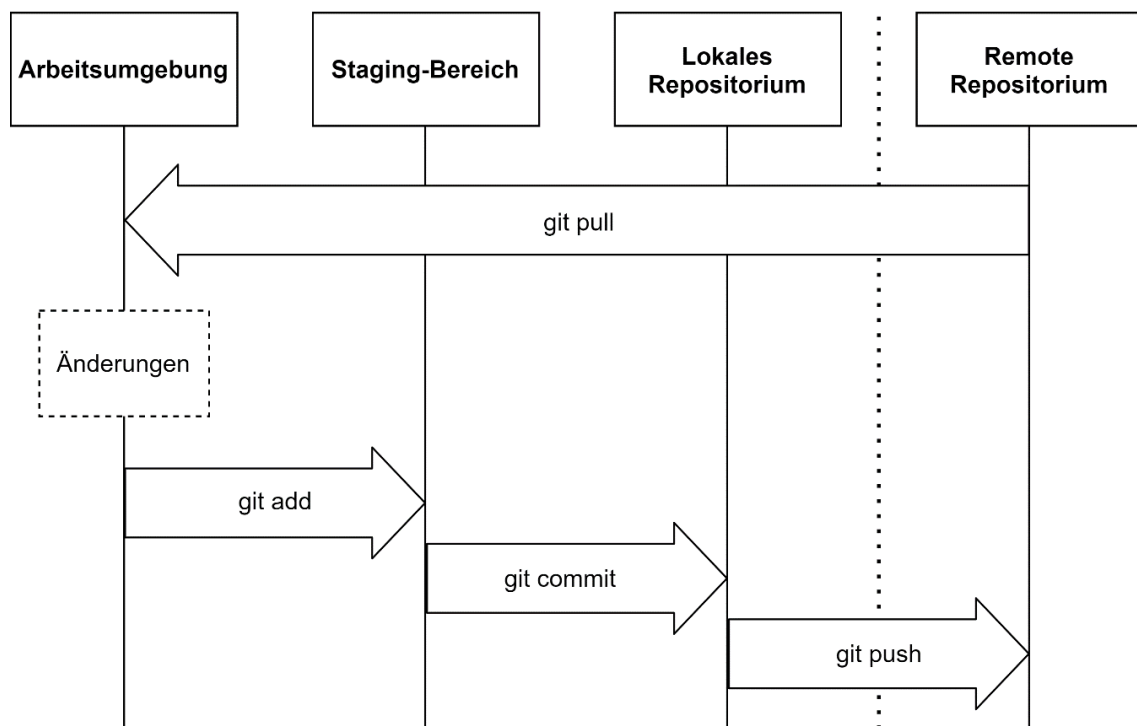


Abbildung 7: Darstellung des Ablaufes der Versionsverwaltung Git (Eigene Darstellung)

Abbildung 7 stellt die grundlegenden Funktionen von Git dar. Die Hauptfunktionalitäten beziehen sich hierbei zunächst auf `git pull`, um den Bestand eines Git-Repositorys mit dem lokalen Stand abzugleichen und bei Bedarf die Daten aus dem Repository anzufordern und gegebenenfalls die lokalen Dateien zu ersetzen. Bearbeitete Dateien können mittels `git add` einer Vorstufe, dem sogenannten Staging-Bereich, hinzugefügt

und im Anschluss mittels *git commit* bestätigt werden. Ein commit enthält hierbei eine konkrete Beschreibung, wodurch Änderungen im Kontext der Historie nachvollziehbarer gestaltet werden. Der Befehl *git push* überträgt dann die mittels *git add* und *git commit* markierten und beschriebenen Dateien mit ihren Änderungen an ein Git-Remote-Repository, wodurch diese im Anschluss zentral zur Verfügung stehen. (Loeliger and McCullough, 2012, pp. 50–53) (Vuorre and Curley, 2018, p. 221)

Im Falle von parallelen Entwicklungen bietet Git sogenannte branches, welche als Abzweigungen des Repositoriums repräsentiert werden. Diese branches können unabhängig voneinander bearbeitet und je nach Fortschritt wieder dem Hauptzweig, dem Master, zugeführt werden. Diese Funktionalität bietet insbesondere die Möglichkeit unterschiedliche Ansätze und Funktionserweiterung zu verfolgen. Bei einem Misserfolg hat der einzelne branch keine Auswirkung auf den Master (Haupt-branch), bei einem Erfolg kann jedoch eine Integration durchgeführt werden. (Blischak et al., 2016, p. 14) (Loeliger and McCullough, 2012, pp. 91–93)

Zur grafischen und auch funktionalen Erweiterung von Git existiert beispielsweise die 2011 entwickelte open-source DevOps-Plattform GitLab. Das System bietet beliebige Projekte und Git-Repositories zu erstellen, inklusive einer Vielzahl an Management-Funktionen. Darunter fällt die Verwaltung des Repositoriums, beispielsweise in Hinblick auf branches, die Einsicht in die Historie und auch die Wiederherstellung konkreter Versionen. Zusätzlich bietet GitLab ein Portfolio an Funktionen zur Durchführung eines agilen Projektmanagements, inklusive Projektplänen, Meilensteinen, einzelnen Aufgaben und eine Verlinkung von Aufgaben mit übersandten Commit-Nachrichten. Weiter erlaubt eine Benutzer- und Rechteverwaltung Personen zu Projekten hinzuzufügen und mittels der definierten Rechte bestimmte Bereiche des Repositoriums und somit der Entwicklung/Forschung einzuschränken. (Choudhury et al., 2020, p. 2) (Baarsen, 2014, pp. 1–2) (GitLab B.V., 2022)

Das Projektmanagement, die Kode-Verwaltung und Versionierung dieser Dissertation erfolgt mittels einer innerhalb des DIZ betriebenen GitLab Instanz (Version 15.03) und der Verwendung von Git (Version 2.37.1).

3.3 Samply.MDR

Im Zuge der 2012 durchgeführten Entwicklung einer Clinical Communication Platform (CCP-IT) des Deutschen Konsortiums für Translationale Krebsforschung und dem 2013 vom BMBF geförderten Projektes OSSE, einer open-source Registerlösung für seltene Erkrankungen, erfolgte die technische Generierung von IT-Werkzeugen, welche diverse Gegebenheiten existenter und neu zu etablierender IT-Strukturen organisieren und verbessern sollte. Die entwickelten Werkzeuge stellten unter anderem ein Metdatenrepositorium dar, welches sich im Nachgang als das Tool Samply.MDR, in Form eines eigenständigen Systems, ausgegliedert hat. Die MDR Lösung basiert primär auf dem Standard ISO/IEC 11179, welcher konkrete Vorgaben für die Struktur und Organisation eines Metdatenrepositoriums vorgibt. (Universitätsklinikum Frankfurt, 2022) (Muscholl et al., 2014) (Kadioglu et al., 2018, p. 51)

Samply.MDR bietet die Möglichkeit auf den beschriebenen Standards lokale Datenelemente konkret zu organisieren, indem eine Strukturierung, Formalisierung und Standardisierung erfolgen kann. Das MDR erzeugt Einträge in Form von Datenelementen, welche mit konkreten Variablen beschrieben werden können. Diese Variablen lassen sich zunächst in einen Namen, eine Definition und eine Validierung unterteilen. Letzteres lässt sich Abbildung 8 und Abbildung 9 entnehmen und kann diverse Informationen wie einen Datentyp oder einen Datenbereich abbilden. (Vengadeswaran et al., 2019, p. 87) (Kadioglu et al., 2018, p. 52)



Abbildung 8: Initiale Definition eines Datenelementes in Samply.MDR (Screenshot aus eigener Systeminstanz)

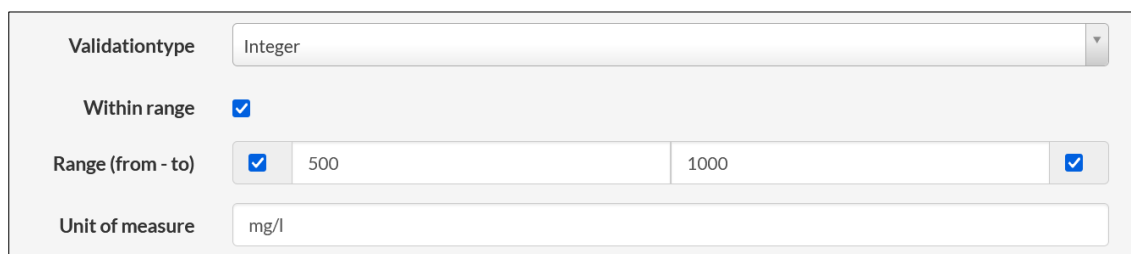


Abbildung 9: Validierungsinformationen eines Samply.MDR Datenelementes (Screenshot aus eigener Systeminstanz)

Eine weitere Funktion des Samply.MDR stellt auch die Abbildung bereits existenter Terminologien und Klassifikationen dar.

Nach einem erfolgreichen Import der Klassifikationsdateien ist die Kode-Struktur innerhalb des Systems hierarchisch ersichtlich und kann detailliert eingesehen werden, wie in Abbildung 10 verdeutlicht. (Vengadeswaran et al., 2019, p. 87)

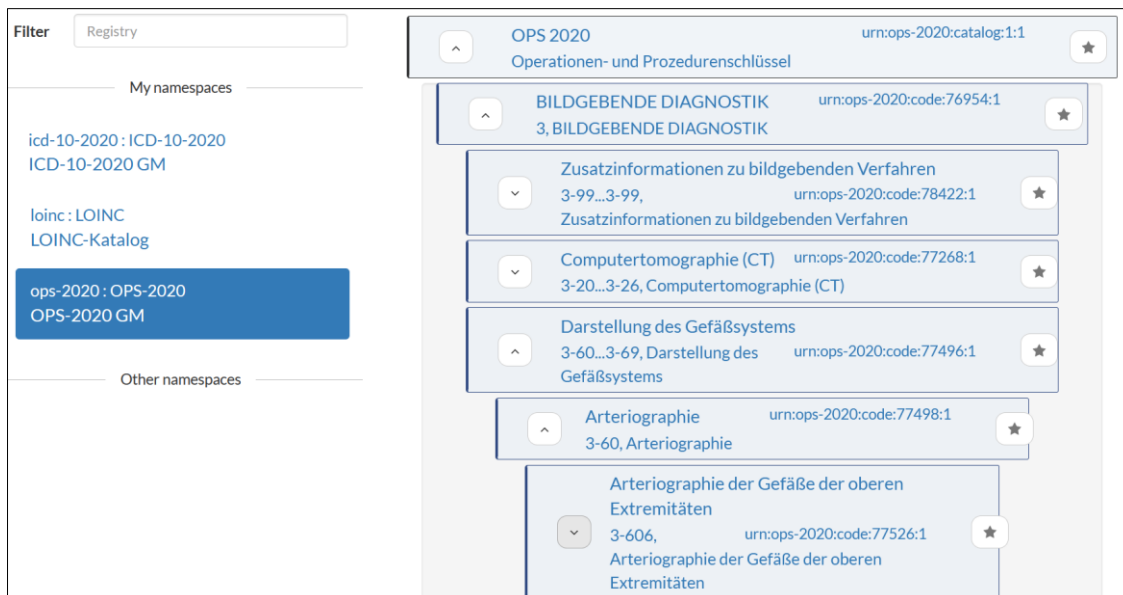


Abbildung 10: Hierarchische Darstellung von Klassifikationen innerhalb von Samply.MDR (Screenshot aus eigener Systeminstanz)

Zur Abfrage dieser konkreten Datenelemente und Strukturen steht eine vollumfängliche REST-API zur Verfügung, welche von anderen Systemen als Informationsgeber genutzt werden kann. (Kadioglu et al., 2016)

Im Zuge dieser Dissertation erfolgt die Verwendung von Samply.MDR (Version 2.3.5-MIRACUM) als Informationsgeber im Kontext der Klassifikations- und Terminologie-Informationen (Semantik und Hierarchie), unter Nutzung der bereitgestellten API.

3.4 Docker

Im Jahr 2010 wurde das jetzt als Docker Inc. bekannte Unternehmen gegründet und stellte ihr Hauptprodukt Docker bereit. Das System stellt eine betriebssystembasierte Virtualisierungsschicht dar, welche nativ auf Funktionen des Linux-Kernels aufbaut und somit bereits vorhandene Strukturen für die Virtualisierung, sogenannte Container, nutzt. Docker isoliert mittels der Kernelfunktionen Ressourcen des zu Grunde liegenden Host-Systems und stellt diese den Containern bereit. Diese gekapselten Bereiche können in einer Vielzahl auf einem einzigen System betrieben werden, ohne Konflikte mit

Ressourcen zu verursachen. Die Verwaltung dieser einzelnen Container obliegt dem sogenannten Docker-Daemon, welcher neben der Docker CLI (command line interface) und der Docker API ein Teil der Docker-Engine darstellt. (Bhat, 2018, pp. 1–2, 20) (Matthias et al., 2016, pp. 23, 103)

Die Schichten und Kapselung im Kontext der Virtualisierung sind in Abbildung 11 dargestellt. (Pittard and Li, 2020, pp. 295, 297) (Mouat, 2016, pp. 4–5)

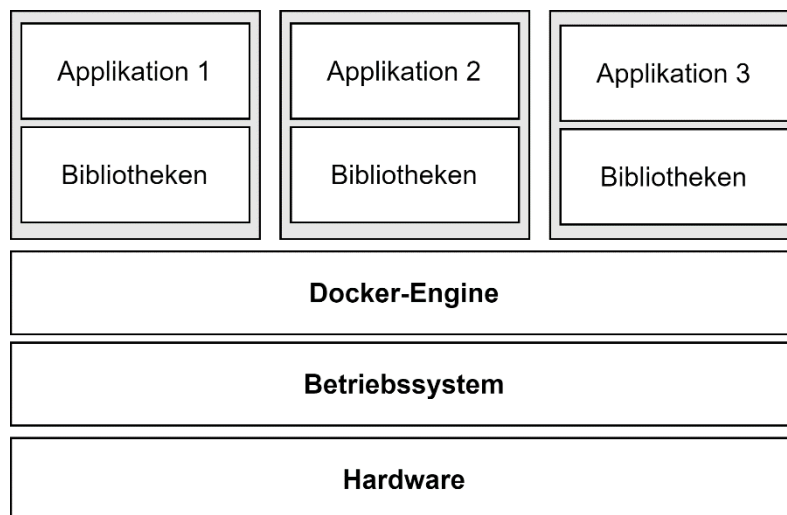


Abbildung 11: Virtualisierungsschichten und Kapselung der Systeme unter Docker (Eigene Darstellung)

Aufgrund der vollständigen Kapselung dieser Container können beispielsweise mehrere Systeme parallel betrieben werden, welche die gleichen Netzwerkressourcen wie Ports benötigen, ohne Konflikte wie Doppelbelegungen zu verursachen. Ebenfalls existiert die Möglichkeit unterschiedliche Speicherkonzepte für die Container bereitzustellen. Hierbei wird unterschieden in Mounts (Einbindung von existenten Speicherbereichen des Host-Systems), Volumes (Erstellte separate Speicherbereiche für den entsprechen Container, ebenfalls auf dem Host-System) und sogenannte tmpfs, welche Speicherbereiche darstellen, die lediglich innerhalb des Hauptspeichers des Host-Systems existieren und somit flüchtig sind. (Docker Inc., 2022a) (Matthias et al., 2016, pp. 111–114)

Sämtliche Container basieren auf sogenannten Images, welche eine Bauanleitung darstellen. Ein Image kann auf einem Dockerfile basieren, welches konkrete, menschenlesbare und stapelbasierte Befehle zur Erstellung enthält. Die Befehle beziehen sich hierbei auf eventuell zu Grunde liegende und schon vorhandene Images wie ein Betriebssystem, Netzwerkkonfigurationen, Anpassungen an Umgebungsvariablen, verwendbare Ressourcenbereiche und auszuführende Befehle zur

Anpassung der zu erstellenden Umgebung. (Merkel, 2014, pp. 4–5) (Boettiger, 2015, p. 74) (Fink et al., 2014)

Eine Steigerung des Vorgehens der Verwendung von Dockerfiles, zur Erstellung einzelner Images und somit lauffähigen Containern, stellt Docker Compose dar. Diese Erweiterung ermöglicht es, unter Bezugnahme von YAML-Dateien, einen Verbund von Containern, gleichzeitig und mit dem Vorteil der Definition von Abhängigkeiten unter diesen, zu erstellen. Ebenfalls können konkrete Anpassungen an den Umgebungen getroffen werden, es benötigt jedoch weiterhin die zu Grunde liegenden Bauanleitungen wie beispielsweise ein Dockerfile. Aufgrund der Möglichkeit Beziehungen zwischen den Containern direkt bei der Erstellung zu definieren, entfällt ein Großteil des Konfigurationsaufwandes im Vergleich zur einzelnen Erzeugung der virtuellen Infrastruktur. (Docker Inc., 2022b) (Bhat, 2018, pp. 154, 161–162)

Sämtliche Systeminstallationen im Kontext dieser Dissertation werden in Form von Containern bereitgestellt. Verwendet wird Docker in Version 20.10.

3.5 Airbyte

Die Software Airbyte stellt eine open-source Datenintegrationslösung dar, welche derzeit als eine der schnellst wachsenden Systeme in diesem Bereich gilt. Das im Jahr 2020 gestartete Projekt verfolgt das Ziel Konnektoren bereitzustellen, um beliebige Datenquellen anzusprechen, die Daten zu exportieren, zu transformieren und abschließend in eine andere Datenquelle integrieren zu können. Das Container-basierte System ermöglicht das schnelle Hinzufügen einzelner Konnektoren und bietet pro Schnittstelle separate und teils feingranulare Konfigurationsmöglichkeiten. Aufgrund der quelloffenen Entwicklung können die Konnektoren beliebig angepasst oder sogar vollständig eigens entwickelt werden. Letzteres wird von Airbyte auch durch die Bereitstellung eines Connector Development Kit (CDK) unterstützt, wodurch laut Aussage der Entwickler-Community ein Konnektor innerhalb weniger Stunden vollständig bereitgestellt und in das Hauptsystem integriert werden kann. (Airbyte, Inc., 2022a) (Business Wire, 2022)

Aufgrund der Extraktion gegebenenfalls sensibler Daten unterzog sich Airbyte im Jahr 2022 einem erfolgreichen SOC 2 Audit, welches sämtliche Komponenten auf Kriterien, wie die Datensicherheit, den Datenschutz, die Verfügbarkeit, die Integrität, die Verarbeitung und die Vertraulichkeit des Systems, konkret prüft. (Business Wire, 2022) (Association of International Certified Professional Accountants, 2022)

Eine im Zuge dieser Dissertation verwendete Kernkomponente bezieht sich auf den Konnektor Postgres. Dieser unterteilt sich in zwei Entwicklungen, welche die Quelle (Source-Connector) und das Ziel (Destination-Connector) repräsentieren. Der Source-Connector extrahiert die definierten Tabellen inklusive deren Daten, wohingegen der Destination-Connector definieren kann, wie die Daten integriert werden. Dies kann mittels einer vollen oder einer inkrementellen Integration erfolgen. Sofern die Konnektoren beide in Kombination genutzt werden, stellt dies ein PostgreSQL Replikationsverfahren dar. Dieses Verfahren ermöglicht die Synchronisation oder das Klonen von PostgreSQL-Datenbanken. Weiter erstellt Airbyte mit seinen Konnektoren vollumfängliche Logs, wodurch stets eine Überwachung des Fortschrittes und der durchgeführten Replikationen oder Integrationen erfolgen kann. (Airbyte, Inc., 2022b) (Airbyte, Inc., 2022c)

Diese Dissertation verwendete eine zentrale, eigens betriebene Airbyte-Instanz (Version v0.39.40-alpha), welche den Postgres-Source-Connector (Version 0.4.43) und Postgres-Destination-Connector (Version 0.3.21) zur Durchführung eines Replikationsverfahren verwendet.

3.6 Keycloak

Keycloak stellt eine open-source Identitäts- und Zugangsmanagement Lösung dar, welche eine Vielzahl an Standards unterstützt. Das System bietet neben einer eigenen Benutzerverwaltung auch die Möglichkeit, bereits vorhandene Identitäts-Anbieter, sogenannte Provider, wie beispielsweise LDAP, einzubinden und als Vermittler zu dienen. Keycloak lässt sich durch eine umfangreiche API an bereits existente Systeme und auch Eigenentwicklungen schnell und zuverlässig anbinden. Der Austausch der Daten erfolgt dann beispielsweise durch die etablierten Protokolle OAuth2, OpenID Connect oder SAML2. (Karasawa et al., 2020, p. 2) (Christie et al., 2017, p. 1) (Red Hat, Inc., 2022)

Das Identitätssystem übernimmt hierbei für die angebundenen Anwendungen den gesamten Login-Prozess und stellt benötigte Authentifizierungs-Tokens bereit, was in Abbildung 12 schematisch dargestellt ist.

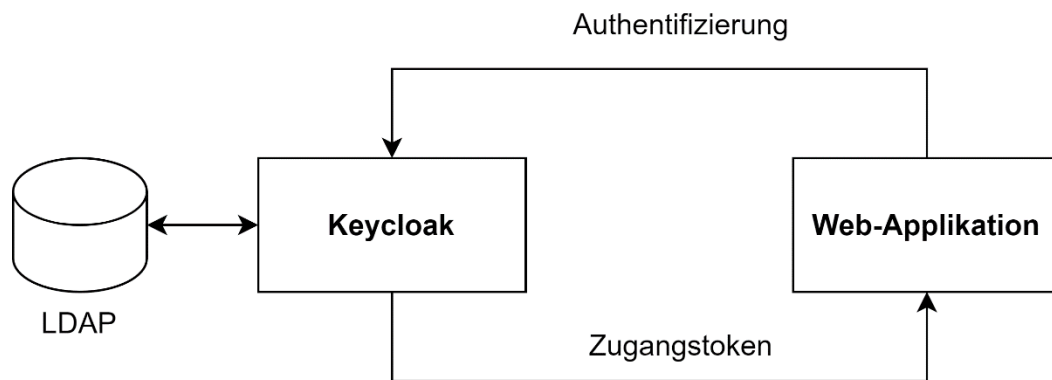


Abbildung 12: Darstellung der Authentifizierung unter Verwendung von Keycloak (Eigene Darstellung)

Keycloak organisiert sich intern in sogenannte Realms, welche wiederum in Clients untergliedert werden. Während ein Realm eine Domäne repräsentiert, stellen Clients die angebotenen IT-Systeme dar. Benutzer können für bestimmte Clients freigeschaltet oder gesperrt werden, wodurch eine Rechteverwaltung ermöglicht wird. Diese Rechteverwaltung und Zuweisung kann sehr feingranular auch auf einzelne Funktionen der angebotenen Applikationen erfolgen. (Norimatsu et al., 2022, pp. 87–88) (N. and Cholli, 2020)

Durch Keycloak kann ein zentrales Authentifizierungssystem bereitgestellt werden, welches auch in Kombination mit weiteren dezentralen Keycloak-Instanzen gekoppelt werden kann. Durch dieses Vorgehen können föderierte Authentifizierungssysteme bereitgestellt werden, welche sogar einen Login über unterschiedliche Standorte hinweg und in fremde IT-Systeme ermöglichen. Dieses Verfahren kommt im Zuge von MIRACUM zum Einsatz. (Prokosch et al., 2018, p. e85)

Diese Dissertation verwendet eine DIZ eigens betriebene Keycloak-Instanz (Version 15.01) als Login-Provider für das zu etablierende Metadatenrepositorium und zur internen Authentifizierung gegenüber weiteren Systemen der DIZ-IT-Landschaft.

3.7 Datenintegrationsplattform

Bereits im Jahr 2014 entstand an der Technischen Hochschule Mittelhessen der Gedanke einer umfangreichen webbasierten Plattform zur Integration medizinischer Forschungsdaten und der Möglichkeit, diese innerhalb der Plattform verwalten und analysieren zu können. (Eberhardt, 2016) (Eberhardt, 2017) Aufgrund des kontinuierlichen Fortschrittes der medizinischen Forschung, der Medizininformatik im klinischen Bereich und dem allgemeinen technischen Fortschritt, wurde die Plattform

erstmalig 2018 unter dem Entwicklungsnamen Datenintegrationsplattform (DIP) neu konzipiert.

Durch den Start des MIRACUM-Projektes erfolgte die erneute Anpassung des technischen Unter- und Aufbaus der DIP. Die modulbasierte Web-Applikation basiert auf einem PHP Backend mit dem Laravel Framework, einem Vue.js Frontend und einer PostgreSQL Datenbank. Die Neukonzipierung und Entwicklung erfolgte zunächst in Bezug auf das Grundgerüst und umfasst neben einem Datenset-basierten Datenmanagement auch ein komplexes Benutzer-, Domänen-, Rollen- und Rechtesystem, welches feingranular angepasst werden kann und plattformintern auch das Teilen von Daten unter Forschenden ermöglicht. Weiter ermöglicht die Plattform die Einbindung von sogenannten Modulen, welche sich in interne und externe Module unterteilt. Interne Module sind eigens entwickelt, direkt in die Plattform integriert und nutzen den vollen Funktionsumfang inklusive Datenspeicher der DIP. Externe Module stellen eigenständige IT-Systeme dar und können mittels Fremdaufruf oder iFrame eingebunden und somit dennoch zentral aus der Plattform heraus genutzt werden. Die Datenintegrationsplattform dient als zentralisiertes Managementsystem und Verbundsystem durch die Bereitstellung aller Tools des Gießener Datenintegrationszentrums. Aus diesem Grund agiert die DIP standortintern als gezielter Einstiegspunkt und somit Portal zum digitalen Werkzeugumfang und der Datenlandschaft. (Sparenberg, 2019) (Seibel, 2021)

Der weitere Funktionsumfang wurde auf Basis von Rückmeldungen durch Forschende und Vorgaben des Datenintegrationszentrums definiert und umgesetzt. Eine für das DIZ wichtige Funktionalität stellt ein Transferstellenmodul dar, welches die primäre Arbeitsumgebung des Transferstellenmitarbeiters darstellt und für die Protokollierung und Bereitstellung der extrahierten Daten aus Forschung und Routine dient. Zu diesem Zweck benötigt die Transferstelle konkrete Informationen zu den relevanten Daten und deren Quellsysteme, welches in den nachfolgenden Kapiteln erarbeitet wird.

Innerhalb der Dissertation wird die Datenintegrationsplattform als Speicherort für durchgeführte Abfragen, sowohl aus Sicht des Forschenden, als auch aus Sicht der Transferstellenmitarbeiter genutzt.

3.8 IT-Grundschutz

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) verdeutlicht mit seinem Standard 200-2 die grundlegende Methodik und Vorgehensweise zur Absicherung von IT-Infrastrukturen aus Sicht sämtlicher beteiligter Komponenten, von der eingesetzten Hardware, über Schichten der Virtualisierung, bis hin zur finalen Anwendungsoberfläche und Klienten. Die hierbei zu betrachtenden Konzepte beginnen mit der Strukturanalyse, welche direkt mit der Definition eines Geltungsbereiches einhergeht. Die Analyse erfasst die beteiligten IT-Komponenten und Ablaufprozesse, wodurch die Details des umzusetzenden oder bereits bestehenden Vorhabens ersichtlich werden. (Bundesamt für Sicherheit in der Informationstechnik, 2017, p. 78) Der nächste Schritt wird durch die sogenannte Schutzbedarfsfeststellung repräsentiert, worin jede in der Strukturanalyse betrachtete Komponente eine konkrete Einschätzung, im Kontext der Vertraulichkeit, Verfügbarkeit und Integrität, erhält. (Bundesamt für Sicherheit in der Informationstechnik, 2017, p. 104) (Bundesamt für Sicherheit in der Informationstechnik, 2022, p. 1)

Die „Modellierung eines Informationsverbunds“ verfolgt im darauffolgenden Schritt die Nutzung der Bausteine des IT-Grundschutz-Kompendiums, welches konkrete Methoden, Vorgaben und Hinweise zur Umsetzung bestimmter technischer und organisatorischer Sachverhalte bereitstellt. (Bundesamt für Sicherheit in der Informationstechnik, 2017, pp. 132–133) Der folgende „IT-Grundschutz-Check“ führt eine Soll-Ist-Analyse zur Prüfung des Sachstandes im Vergleich zu den erarbeiteten Grundschutz Bausteinen durch. (Bundesamt für Sicherheit in der Informationstechnik, 2017, p. 145) Der abschließende Schritt stellt eine umfangreiche Risikoanalyse zur Abschätzung des Restrisikos dar. (Bundesamt für Sicherheit in der Informationstechnik, 2017, p. 152) (Haar and Buchmann, 2019, pp. 482–483)

Im Zuge dieser Dissertation wurde die Entwicklung angelehnt an die Bausteine des IT-Grundschutz-Kompendiums SYS und APP durchgeführt.

4. Ergebnisse

4.1 Projektplanung

Grundlage der Projektplanung stellt ein Lastenheft dar, welches zu Beginn durch die Leitung des MIRACUM-Projektes am Standort Gießen, sowie der DIZ-Leitung und daraus resultierend den Betreuern dieser Dissertation, definiert wurde und die gewünschten Projektzeile enthält. Das Lastenheft ist im Anhang unter Kapitel 13.9 ersichtlich und beschreibt die Notwendigkeit nach rechtlich organisatorischen Ordnungen und einem Portal für Machbarkeitsabfragen. Hierbei wurden unter anderem funktionale Anforderungen, nicht-funktionale Anforderungen und Qualitätsanforderungen definiert. Diese Anforderungen wurden im Detail mit den Autoren besprochen und im Anschluss, in einer ersten Phase des Projektes, die mögliche Umsetzung erarbeitet. Es folgte daraufhin die Erstellung eines Pflichtenheftes, welches im Anhang unter Kapitel 13.10 bereitgestellt ist. Das Pflichtenheft beschreibt die einzelnen benötigten Anforderungen und erfasst neben einer konkreten Beschreibung auch mögliche Wechselwirkungen und konkrete Umsetzungsideen. Wechselwirkungen sind je nach Anforderung im Kontext des eMDR vorhanden, da das System aktiv in die IT-Landschaft des DIZ integriert ist und mittels seiner Schnittstellen als Informationsgeber für weitere Systeme dienen kann. Weiter existieren Wechselwirkungen zwischen der Entwicklung des eMDR und den zu erstellenden Ordnungen, welche klare Vorgaben an das System, im Funktionsumfang und der Organisation, definieren. Im Bereich der Umsetzung der einzelnen Anforderungen wurden bereits konkrete Werkzeuge und Sprachen genannt, welche auch final zum Einsatz kamen. Das Pflichtenheft wurde von den Erstellern des Lastenheftes und somit den Auftraggebern abgenommen.

Auf Basis des fertigen Pflichtenheftes wurde ein Projektplan erzeugt, welcher sämtliche Aspekte des Projektes in einen zeitlichen Ablauf bringt und Abhängigkeiten unter den einzelnen Teilaufgaben herstellt. Dieser Projektplan ist im Anhang unter Kapitel 13.11 dargestellt und definiert konkrete Zeiträume, welche sich grundlegend in folgende Bereiche unterteilen lassen:

01.01.2019 bis 14.06.2021 – Zeitraum vor dem durch die MII beauftragten Audit

15.06.2021 bis 01.12.2022 – Zeitraum nach dem durch die MII beauftragten Audit

Der erste Zeitraum stellt primär die Aufbauphase des DIZ dar. Hierbei stehen die rechtlichen und organisatorischen Ordnungen sowie der erste Prototyp des eMDR im Vordergrund. Auf Basis der Ergebnisse des Audits wird der darauffolgende Zeitraum für

die Nachbearbeitung, Anpassung und Optimierung genutzt. Weiter erfolgt die weitere Funktionsentwicklung für das eMDR und die Rückmeldung der Gremien zu den erstellten Ordnungen sowie deren Inkrafttreten.

4.2 Organisatorische und rechtliche Grundlagen

Im Zuge der Etablierung des Datenintegrationszentrums, speziell im Hinblick auf die Herausgabe und Nutzung von Daten unter derer primär auch das etablierte eMDR aus einer organisatorischen und rechtlichen Sicht fällt, wurden drei Ordnungen am Standort Gießen veröffentlicht, welche in den nachfolgenden Unterkapiteln detaillierter behandelt werden. Der Zusammenhang und die Notwendigkeit der Ordnungen werden in Abbildung 13 verdeutlicht.

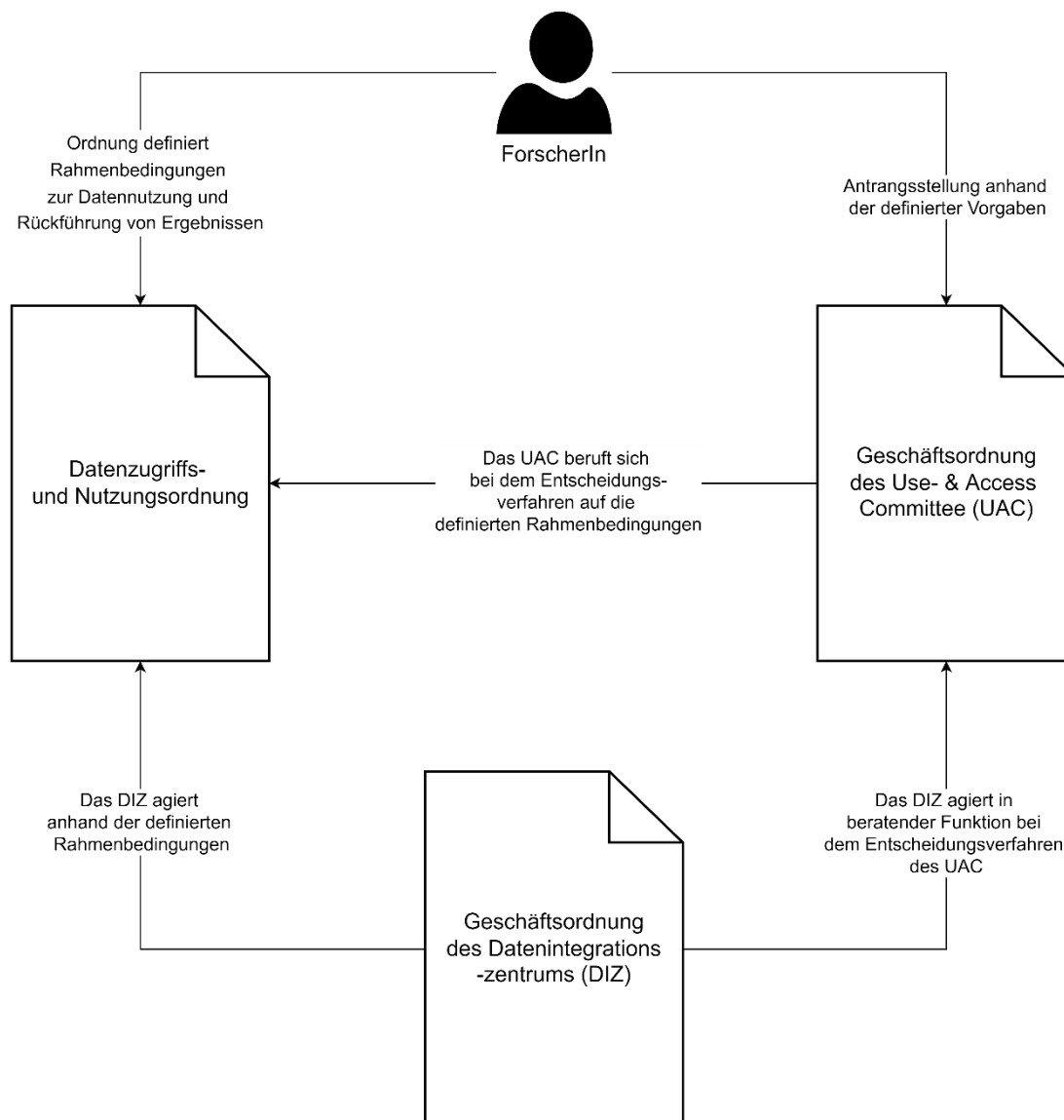


Abbildung 13: Zusammenhang rechtlicher und organisatorischer Rahmenbedingungen (Eigene Darstellung)

Die Ordnungen wurden parallel zum Aufbau des Gießener DIZ und der zu entwickelnden Systeme von der DIZ-Leitung im Zuge dieser Dissertation erarbeitet und von den Vertretern der Betreiber des DIZ¹ für die Weitergabe an die beteiligten Gremien vorbereitet. Anschließend wurden die Ordnungen von den rechtlichen Vertretern der beiden Organisationen (Universitätsklinikum Gießen und Marburg GmbH - Standort Gießen (UKGM-Gi) und Justus-Liebig-Universität Gießen (JLU)) detailliert geprüft und von der Geschäftsführung des UKGM-Gi und dem Präsidium der JLU freigegeben.

4.2.1 Datenzugriffs- und Nutzungsordnung

Die Datenzugriffs- und Nutzungsordnung stellt die primäre Ordnung (Anhang 13.1) dar und regelt konkret die Vorgänge und Vorgaben zur Nutzung von Daten, nicht nur im Kontext des Datenintegrationszentrums, sondern auch daran angegliederte Prozesse und sonstige am Fachbereich Medizin durchgeführten Datennutzungen. Die primären Bereiche stellen die Pflichten der Forschenden sowie die vertraglichen Vorgaben und Rahmenbedingungen im Zuge einer Datennutzung dar. Um eine Nutzung des eMDR als Machbarkeitsportal innerhalb der DIZ-Umgebung ermöglichen zu können, musste in dieser Ordnung bereits eine entsprechende Grundlage geschaffen werden, welche sich jedoch primär auf Vorgaben / Bedingungen bei der Umsetzung des Systems beziehen. Die nachfolgenden Bedingungen ergeben sich konkret aus der am Standort etablierten Datenzugriffs- und Nutzungsordnung:

1. Authentifizierung der Forschenden

Das eMDR benötigt eine konkrete Authentifizierung, um die Machbarkeitsabfragen nur für berechtigte Forschende zu ermöglichen. (*konkret, Anbindung an zentrale Authentifizierungsdienste des DIZ*)

2. Einbringung in das Entscheidungsverfahren des Use- & Access Committee (UAC)

Ein Datenextrakt des eMDR findet Anwendung innerhalb des UAC-Entscheidungsverfahrens (*unkonkret*)

3. Rückübermittlung der Ergebnisse

Bei einer Ergebnisrückführung durch Forschende sollen diese Daten wieder Anwendung innerhalb des eMDR finden (*konkret, Daten werden in die Datenbasis zurückgeführt oder bei Bedarf entfernt*)

¹ Durch die Betreiber des Datenintegrationszentrums (DIZ), der Fachbereich Medizin der Justus-Liebig-Universität Gießen und das Universitätsklinikum Gießen und Marburg GmbH - Standort Gießen, ernannte Interessensvertreter.

Diese Bedingungen nahmen einen aktiven Einfluss auf die Entwicklungsgrundlage des eMDR. Die in diesem Falle noch als *unkonkret* markierte Bedingung wird in der Geschäftsordnung des UAC konkreter erläutert.

4.2.2 Geschäftsordnung des Use- & Access Committee

Jeder Standort des Konsortiums benötigt ein Use- und Access Committee (UAC) zur fachlichen und inhaltlichen Prüfung von Datennutzungsanträgen. Hierbei werden lokale Kompetenzen benötigt, welche insbesondere die wichtigen Aspekte des Datenschutzes, der Ethik und des juristischen Bereiches abdecken. Aufgrund lokaler Gegebenheiten und Besonderheiten unterscheiden sich die UAC-Geschäftsordnungen zwischen den Standorten auch inhaltlich teilweise sehr stark. Am Standort Gießen wird der Fokus auf die Bereiche der Mitglieder (Anzahl und Kompetenzen), die Antragseinreichung (Inhalte) und das Entscheidungsverfahren (Ablauf) gelegt (Anhang 13.2). Analog zur Datenzugriffs- und Nutzungsordnung wurden auch in der UAC-Geschäftsordnung Grundlagen zur Nutzung des eMDR geschaffen, welche in konkreten Anforderungen an das System münden.

1. Antragseinreichung – Angaben zur Feasibility

Im Zuge der Antragseinreichung durch Forschende werden konkrete Angaben zur Feasibility benötigt (*konkret, Exportmöglichkeit der Resultate des eMDR*)

2. Antragsverwaltung – elektronische Plattform

Die Resultate der Machbarkeitsanalyse müssen auch in die elektronische Plattform integriert werden können (*konkret, Exportmöglichkeit der Resultate soll auch zu anderen Systemen erfolgen können*)

3. Entscheidungsverfahren – DIZ-Leitung

Die DIZ-Leitung benötigt zur technischen Überprüfung der Machbarkeit konkrete Informationen zur Herkunft der Daten aus einer Abfrage sowie der Ursprungssysteme (*konkret, interne Erfassung der Quellsysteme pro Datensatz*)

4.2.3 Geschäftsordnung des Datenintegrationszentrums

Die dritte Ordnung stellt die Geschäftsordnung des Datenintegrationszentrums dar und definiert neben den organisatorischen Strukturen auch die bereitgestellten Dienste des DIZ (Anhang 13.3). Hierzu zählt beispielsweise die Datenintegrationsplattform (DIP) als zentrales digitales Werkzeug für Forschende und auch bestimmte damit einhergehende Rahmenbedingungen. Diese Bedingungen erzeugen, wie auch die innerhalb der

anderen Ordnungen genannt, bestimmte Grundlagen und Vorgaben zur Umsetzung des eMDR.

1. Verfügbarkeit

Das System muss eine hohe Verfügbarkeit, bedingt durch die vorhandene Infrastruktur, vorweisen (*konkret, ausfallsichere Installation des Systems*)

2. Infrastruktur – Netzwerk

Das System muss in der Infrastruktur des Gießener Forschungsnetzes (GIFONET) betrieben werden (*konkret, Installation des eMDR auf bereits vorhandener GIFONET-Infrastruktur*)

3. Datenübergabe an Transferstelle

Die Transferstelle agiert innerhalb der DIP und benötigt dort die extrahierten Daten des eMDR (*konkret, Schnittstelle zwischen dem eMDR und der Datenintegrationsplattform*)

Die Erstellung und darauffolgende Freigabe der Ordnungen bilden die technischen und organisatorischen Rahmenbedingungen für das eMDR und das DIZ im Allgemeinen. Die nachfolgenden Kapitel beziehen sich auf die Datengrundlage und die technischen Aspekte zur Umsetzung des Systems.

4.3 Kerndatensatz

Zur Etablierung einer grundlegenden Datenbasis, auf Basis von Metadaten, wird zunächst eine Konkretisierung des benötigten Datenumfangs benötigt. Dieser ergibt sich im Rahmen des Datenintegrationszentrums aus Erfahrungen klinischer Mitarbeiter, Mitarbeiter der klinischen EDV, Erfahrungen durch das MIRACUM-Konsortium und aus diversen Erkenntnissen der Medizininformatik-Initiative für benötigte Daten zur Durchführung von Studien. (Zabka et al., 2019) (Ulrich et al., 2021) Speziell letzterer Aspekt ist nicht nur für einen dezidierten Use-Case des MIRACUM-Konsortiums von besonderer Bedeutung, sondern das etablierte erweiterte Metadatenrepositorium (eMDR) soll als direkter Informationsgeber für klinische Datenabfragen und somit für die grundlegende Machbarkeitsanalyse klinischer Fragestellungen und Studien dienen.

Der durch das Datenintegrationszentrum definierte Kerndatensatz bezieht sich auf vier Kategorien, welche die Patientendaten, Klinikdaten und Diagnose- sowie Prozeduren-Daten darstellen und in Abbildung 14 ersichtlich ist.

Dieser definierte und minimale Kerndatensatz bildet die Grundlage des eMDR und stellt die Basis sämtlicher Abfragen dar. Aufgrund der modularen Struktur ist diese Grundlage

stets erweiterbar. Kapitel 4.6.1 und 4.6.2 verdeutlichen, wie diese Daten erweitert und optimiert wurden.



Abbildung 14: Komponenten des definierten Kerndatensatzes (Eigene Darstellung)

4.4 Konzeptionelle Trennung der Informationen

Zu Beginn der Erstellung der Datenbasis wurde ein besonderer Fokus auf die Wartbarkeit und Erweiterbarkeit des Systems gelegt. Bei diesem Aspekt wurde das System konzeptionell in zwei Bereiche untergliedert, welche die Semantik und die eigentliche Information repräsentieren. Die Semantik wird durch die Terminologien und Klassifikationen LOINC, ICD und OPS abgedeckt, die eigentliche Information stellt dar, wie oft ein Vorkommen innerhalb der klinischen Informationssysteme und innerhalb definierter Zeiträume vorhanden ist. Eine Verdeutlichung hierzu ist in der nachfolgenden Abbildung 15 dargestellt.

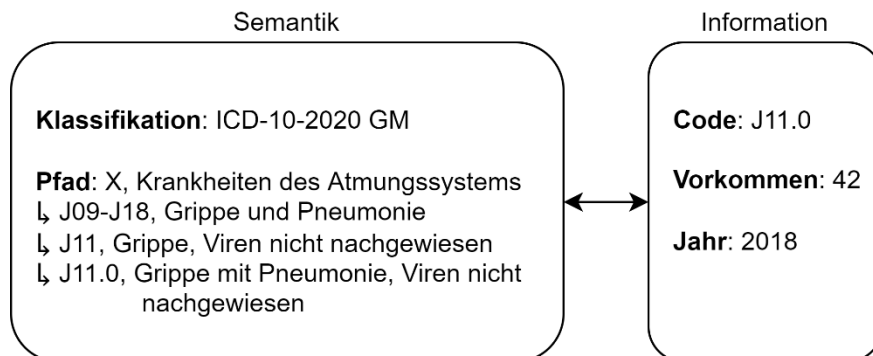


Abbildung 15: Beispielhafte Trennung der relevanten Informationskategorien (Eigene Darstellung)

Speziell aus Sicht der Wartbarkeit weisen diese beiden Arten von Daten grundlegende Unterschiede auf. Die eigentlichen Informationen zu Vorkommen von klinischen Sachverhalten müssen für die Zukunft im Kontext des Kerndatensatzes dynamisch erweitert werden können. Weiter muss es möglich sein, einzelne oder eine Vielzahl an Datensätze, beispielsweise bei datenschutzrechtlichen Änderungen oder bei einem Widerruf der Datennutzung durch Patienten, aus dem System restlos entfernen zu können. Diese Informationen sind jedoch stets an die entsprechende Semantik gebunden, welche in der Regel in neuen Versionen veröffentlicht werden. Falls diese beiden Bereiche in einer gemeinsamen Datenbasis gehalten werden sollen, so stellt sich bereits an dieser Stelle eine logische Trennung heraus. Zur Wartung der Semantik wird aus diesem Grund eine entsprechende Katalogverwaltung benötigt, dessen Inhalte den Informationen zugewiesen werden können. Aus Sicht der Informationen sollen Datenbankabfragen durchgeführt und grafisch aufbereitet werden.

Da es sich bei diesen beiden Bereichen somit um in sich geschlossene und logisch getrennte Systeme handelt, wurde sich auch für eine physische Trennung des erweiterten Metadatenrepositoriums entschieden. Das eMDR selbst behandelt den Aspekt der Vorkommen medizinischer Informationen, wohingegen der Bereich der Semantik vollständig ausgelagert wird. Für diesen Zweck wurde sich für das Tool Samply.MDR entschieden, dessen Etablierung im nachfolgenden Kapitel verdeutlicht wird.

Die wichtige Kommunikation dieser beiden Bereiche, zur Kopplung der Information mit der Semantik, muss selbstverständlich weiterhin gegeben sein. Diese würde in einem gemeinsamen System auf Datenbankebene erfolgen, welches bei einer Trennung nun jedoch auf einer anderen Systemschicht erfolgen muss.

4.5 Semantische Ausgliederung – Samply.MDR

Das durch den MIRACUM Standort Frankfurt am Main bereitgestellte Samply.MDR wurde im Rahmen eines Projektmeilensteines etabliert und dient der Verwaltung von Terminologien und Klassifikationen. Das System wurde an den standorteigenen Authentifizierungsdienst Keycloak angebunden, um eine Administration für definierte Benutzer zu ermöglichen. Diese stellen die Systemadministratoren und Systemverantwortlichen innerhalb des Datenintegrationszentrums dar. Eine Nutzung für weitere Personengruppen ist nicht vorgesehen, da das System lediglich zur Verwaltung der genannten Terminologien/Klassifikationen und somit der Semantik dienen soll.

Die initialen Kataloge wurden von dem Team der Medical Informatics Group (MIG) aus Frankfurt bereitgestellt und stellen OPS-301, ICD10-GM und LOINC dar. Diese wurden in das System importiert und deren Inhalte als Datenelemente definiert, wodurch diese detailliert, mittels bereitgestellter Schnittstellen, angesprochen werden können. Hierbei können aus Sicht eines Terms sämtliche Informationen wie der Kode, die genaue Definition, eine gegebenenfalls vorhandene Beschreibung, die Katalog- / Klassifikationsversion und die Hierarchie bis zur Wurzel angezeigt und abgefragt werden, wie in der nachfolgenden Abbildung 16 ersichtlich ist.

The screenshot displays a hierarchical view of a LOINC catalog. The hierarchy is as follows:

- Labor (urn:loinc:dataelementgroup:10:6)
- Allergiediagnostik (urn:loinc:dataelementgroup:116:1)
- Globalmarker (urn:loinc:dataelementgroup:117:1)
- IgE 19113-0 (urn:loinc:dataelement:1253:1)

Below the hierarchy, there are two main sections:

Validation: A dropdown menu is set to 'Float'.

Unit of measure: The field contains '[IU]/L'.

Slots: A table with the following data:

Name	Value
fhir-coding-system	http://loinc.org
terminology-code	19113-0

Abbildung 16: Aufbau und Struktur der Semantik durch Samplly.MDR (Screenshot aus eigener Systeminstanz)

Das System selbst stellt aus technischer Sicht einen Verbund von drei virtuellen Containern dar und ist aus Sicherheitsgründen sowie best-practise Ansätzen für IT-Infrastrukturen netzwerkseitig isoliert.

Aufgrund der Tatsache, dass das System innerhalb des MIRACUM-Konsortiums als standortübergreifendes Werkzeug etabliert wurde, erfolgt auch von zentraler Stelle eine Bereitstellung von Produkt- und Sicherheitsaktualisierungen, wodurch auch eine

kontinuierliche Wartung des Systems gegeben ist. Die Pflege der Terminologien und Klassifikationen, welche innerhalb des Systems als Kataloge dargestellt werden, kann sowohl durch den Standort Gießen erfolgen, wird jedoch im Rahmen der Projektlaufzeit zunächst durch die Konsortialpartner übernommen.

Durch die Verwendung des Sapply.MDR ist der Aspekt der Katalogverwaltung und die Bereitstellung der semantischen Informationen für das eMDR abgeschlossen.

4.6 eMDR

Das konzeptionierte und entwickelte eMDR wurde von Seiten des Backends und Frontends getrennt. Die nachfolgenden Unterkapitel beschreiben die initialen Planungsschritte, Vorgaben, Konzepte und die letzten Endes entwickelten separierten Komponenten.

4.6.1 Konzeptionierung und interne Vorgaben

Die bereits aus den generierten Ordnungen resultierenden Vorgaben wurden zunächst als Basis für die initiale Konzeptionierung verwendet. Tabelle 1 verdeutlicht die einzelnen Vorgaben und die entsprechenden Auswirkungen auf die Entwicklung und das System.

Vorgabe durch Nutzungs- und Geschäftsordnungen	Auswirkung
Authentifizierung nur für berechtigte Forschende	Backend und Frontend
Ergebnisrückführung	Backend
Exportmöglichkeit der Resultate für Anwender	Frontend
Exportmöglichkeit in andere Systeme	Backend
Erfassung der Quellsysteme pro Datensatz	Backend – Datenbank
Ausfallsichere Installation des Systems	Installation – Methode
Installation auf GIFONET Infrastruktur	Installation
Datenexport zur DIP	Backend und Frontend

Tabelle 1: Vorgaben und Auswirkungen der Nutzungs- und Geschäftsordnungen

Weitere Vorgaben erfolgten durch das Team des Datenintegrationszentrums (Backend)

und durch bereits initial vorhandenes Feedback durch Forschende (Frontend). Diese Vorgaben bezogen sich seitens des DIZ primär auf Schnittstellen, welche jedoch bereits über die Vorgaben aus den Ordnungen erfasst wurden. Seitens der Forschenden sollte ein besonderer Fokus auf eine simple, übersichtliche und schnelle Benutzeroberfläche gelegt werden. Als Referenzwert wurde hierbei die Software i2b2 genannt, welche im Zuge der Entwicklung und Etablierung des Datenintegrationszentrums Forschenden in Projekten zu Testzwecken zur Verfügung gestellt wurde. Das eMDR soll im Vergleich zu dieser Anwendung deutlich simpler gestaltet sein.

Eine der wichtigsten Vorgaben im Zuge der Konzeptionierung ist der initial vorhandene Funktionsumfang seitens der Explorationsmöglichkeiten der Datenbestände. Diese wurden im Vorfeld durch die Betreuer dieser Dissertation sowie durch das Datenintegrationszentrum definiert. Hierbei ist zu beachten, dass die Datenbasis bereits einen großen Umfang vorweisen sollte, die Auswertungsmöglichkeiten sich jedoch zunächst auf simple und kurze Auswertungen beziehen sollen. Als konkrete Vorgabe wurde definiert, einzelne Codes aus den Klassifikationen abzufragen und das Gesamtvorkommen inklusive der Semantik einsehen zu können. Der daraus resultierte und erweiterte Funktionsumfang ist in Kapitel 4.6.4 ersichtlich.

Im Zuge der Generierung eines Kerndatensatzes erfolgte bereits eine direkte Konzeptionierung des Datenbankschemas, welches auch einen direkten Einfluss auf den Funktionsumfang vorweist. Das Schema bildet die Grundlage des Systems und wurde hochgradig modular und simple erstellt. Dies ermöglicht eine nachträgliche Ergänzung von Informationen, welche zusätzlich erfasst werden sollen. Ein Ausschnitt des finalen Datenbankschemas ist in Abbildung 17 dargestellt.

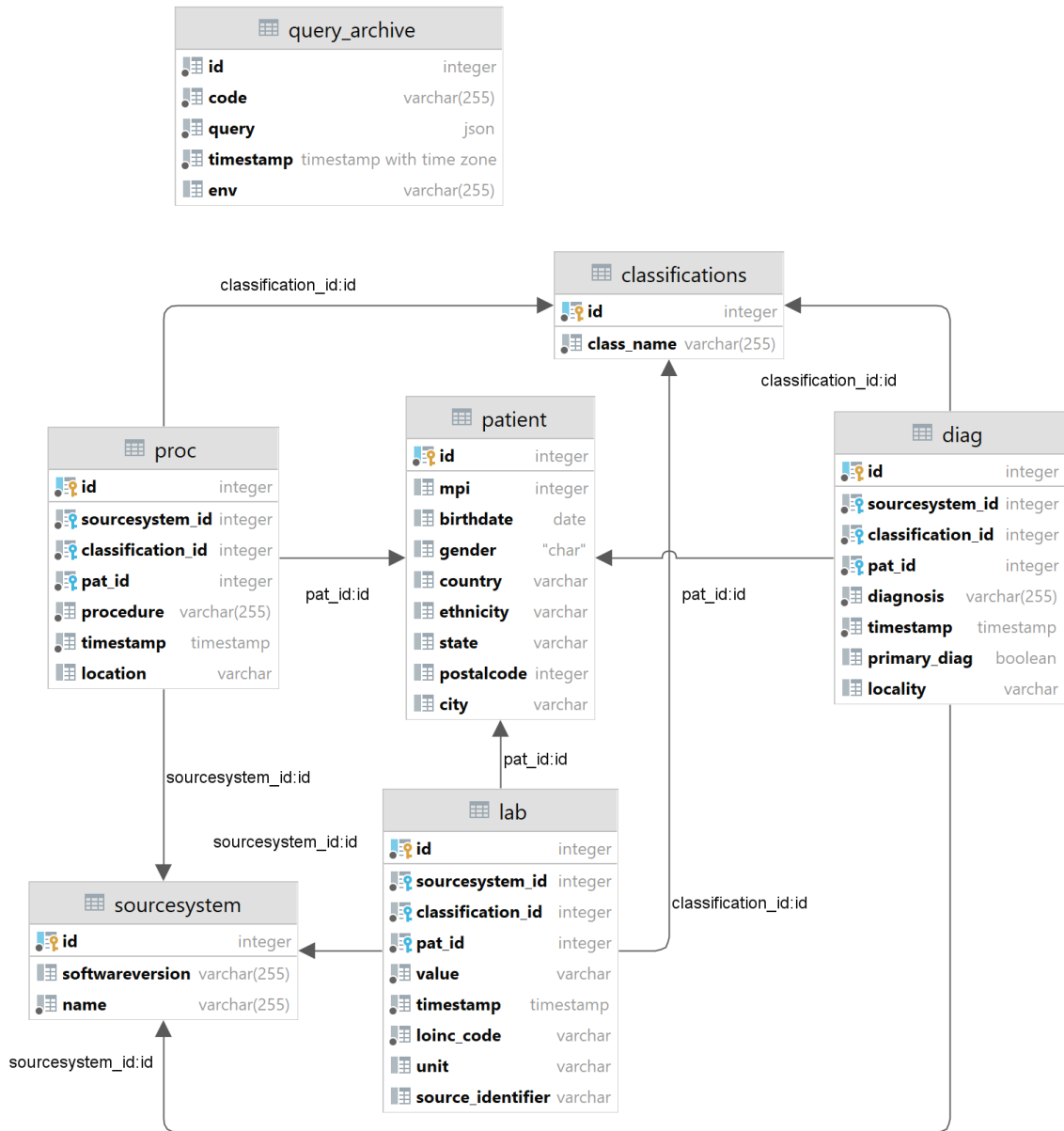


Abbildung 17: Ausschnitt aus dem Datenbankschema des eMDR (Eigene Darstellung)

Die nachfolgenden Kapitel verdeutlichen die erstellten Systemkomponenten auf Basis des gesamten Datenbankschemas.

4.6.2 Backend

Die Initialisierung des Backends erfolgte durch den Einsatz einer PostgreSQL-Datenbank und dem Datenbank-Managementtool *pgAdmin*. Die Datenbank wurde verwendet, da die gesamte DIZ-Infrastruktur vollständig auf PostgreSQL basiert und im Kontext von Schnittstellen, Erweiterbarkeit und Performance diverse Vorteile einher bringt, welche sich im täglichen Gebrauch bewährt haben.

Der in Kapitel 4.3 beschriebene Kerndatensatz musste für die benötigten Funktionen noch erweitert werden, was in Abbildung 17 ersichtlich ist. Das finale Datenbankschema erweitert zunächst die Informationen des Patienten durch das Feld MPI, welches als konkretes Pseudonym für den Patienten verwendet wird. Basierend auf diesem Pseudonym erhält jeder Patient eine interne ID, welche systemintern mit allen relevanten Informationen verknüpft wird (Fremdschlüssel-Attribut).

Zur Umsetzung der Vorgabe, zur Identifikation des Quellsystems im Kontext eines Datensatzes, wurde die Tabelle *sourcesystem* eingeführt, worin die Quellsysteme inklusive ihrer Softwareversion gelistet und mit jedem integrierten Datensatz verknüpft sind. Die Version wird hierbei dazu verwendet, um bei späteren softwarespezifischen Fehlern des Quellsystems gegebenenfalls fehlerhafte Datensätze rückwirkend entfernen und neu einspielen zu können. Ein weiterer wichtiger Aspekt wird durch die Tabelle *classifications* abgedeckt. Innerhalb dieser Tabelle werden sämtliche verwendeten Klassifikationen und Terminologien gelistet und mit den Datensätzen verknüpft. Durch diese Relation ist es möglich, jederzeit die Klassifikation zu einem Datensatz anzuzeigen und darauf basierend die Semantik abzufragen.

Der in Kapitel 4.3 verdeutlichte Kerndatensatz wurde in Form der Tabellen *patient*, *stays*, *visit*, *diag*, *lab* und *proc* übernommen. Die aufenthaltsrelevanten Aspekte wie *stays* und *visit* lassen einen internen Rückschluss zu, um den Weg des Patienten innerhalb des Klinikums nachzuvollziehen. Die Tabellen *diag*, *lab* und *proc* beinhalten die eigentlichen krankheits- und behandlungsrelevanten Datensätze inklusive einiger Besonderheiten. So wird beispielsweise in der Tabelle für Diagnosen (*diag*) eine Unterscheidung zwischen primärer- und sekundärer Diagnose getroffen, weiter wird auch die Lokalität erfasst. Die Tabelle *lab* erfasst nicht nur die Laborcodes inklusive deren Werte, sondern auch die entsprechende Einheit, da diese sich lokal gegebenenfalls von der hinterlegten Einheit der LOINC-Terminologie unterscheidet.

Weiter wurde das Schema vorausschauend bereits um die Tabelle *body*, *vital* und *vital_type* erweitert. Dieser werden im Zuge der initialen Dateneinleitung nicht befüllt, dienen jedoch der Erweiterbarkeit des Systems und bilden Körpereigenschaften der

Patienten (z.B. Größe und Gewicht) sowie Vitalparameter ab. Dies ermöglicht bei einer Erweiterung des Systems eine feingranulare Abfrage- beziehungsweise Auswertungsmöglichkeit.

Einer der wichtigsten Aspekte des Datenbankschemas stellen die Views dar. Diese dienen dem Vorhalten bereits summierter Ergebnisse, um schnelle Abfragen zu gewährleisten. Die Views *diag_counter*, *lab_counter* und *proc_counter* summieren die Vorkommen basierend auf dem Code im Kontext der integrierten Jahre auf, wohingegen die Views *diag_counter_total*, *lab_counter_total* und *proc_counter_total* nicht den Zeitraum berücksichtigen, sondern lediglich das Gesamtvorkommen widerspiegeln. Eine Besonderheit stellt die View *lab_counter_values* dar, welche das Minimum, Maximum und den Durchschnitt der Laborwerte pro Kode darstellt. Dies wird im Kontext einer speziellen Ansicht für das Frontend benötigt und wird im nachfolgenden Kapitel weiter behandelt.

Ein weiterer wichtiger Bestandteil der Datenbank stellt die Tabelle *query_archive* dar, welche im Kapitel 4.6.4 des Funktionsumfangs des eMDR beschrieben ist.

Der nächste Aspekt aus Sicht des Backends wird durch die Integration der Daten in die generierte Datenbasis repräsentiert. Das gesamte Backend wurde mittels Python und dem Framework Flask entwickelt. Flask dient der schnellen und soliden Bereitstellung von API-basierten Backends, welches im Falle des eMDR dringend benötigt wurde. Im Zuge der Nutzung einer PostgreSQL Datenbank wurde das Python-Plugin *psycopy2* eingesetzt, welches das etablierteste Plugin zur Kommunikation mit der genannten Datenbank darstellt. (Varrazzo, 2021)

Aufgrund der zunächst unklaren Entwicklungen des DIZ und den Rahmenbedingungen der Medizininformatik-Initiative wurde eine Integration von §21-Datensätzen (§21 KHEntgG) zunächst einer Schnittstelle für FHIR-Importe vorgezogen. Der Importer ermöglicht die entsprechenden CSV-Dateien direkt in die Datenbasis zu integrieren und erzielt dadurch eine direkte Nutzbarkeit. Dies wurde speziell durch die im Zuge der standortspezifischen Freigabe zur Nutzung von §21-Daten zu Forschungszwecken ermöglicht. Im Anschluss wurde, im Zuge interner DIZ-Projekte zu Ausbildungszwecken, mit einer Studierendengruppe ein initialer FHIR-Importer und Exporter entwickelt. Dies ermöglicht eine direkte Datenintegration durch klinische Systeme und stellt das eMDR im gleichen Zuge, durch die Exportmöglichkeit, als Datengeber dar. Durch die beiden Importstrategien können die integrierten Daten auch im Umfang stets erweitert oder reduziert werden, beispielsweise bei datenschutzrechtlichen Änderungen.

Durch Bereitstellung des Funktionsumfangs zur Befüllung der Datenbasis musste, zur Nutzung der semantischen Informationen, eine entsprechende API-Kommunikation mit dem Samplly.MDR hergestellt werden. Eine Dokumentation der nutzbaren Endpunkte wurde durch die Kollegen der MIG Frankfurt bereitgestellt und wurde entsprechend genutzt. (Universitätsmedizin Mainz - IMBEI, 2015) Die benötigten Schnittstellen beziehen sich konkret auf die Abfrage der Semantik zu einem gegebenen Kode. Die nachfolgende Tabelle 2 verdeutlicht die genutzten Endpunkte und die Auswirkungen auf die Datenlage.

Endpunkt	Nutzung / Auswirkung
<i>/rest/api/mdr/dataelements/ [Katalog]/catalog/search/[Parameter]</i>	Kodesuche für den entsprechenden Katalog, in diesem Falle ICD und OPS. Abfrage übermittelt auch die Hierarchie innerhalb der Kataloge.
<i>/rest/api/mdr/namespaces/ [LOINC-Katalog]/search? query=[Parameter]</i>	Kodesuche für LOINC, da diese Terminologie nicht als Katalog innerhalb des Samplly.MDR verwaltet wird.
<i>/rest/api/mdr/dataelementgroups/ [LOINC-Katalog-URN]/members</i>	Hierarchie-Abfrage für die LOINC-Terminologie.

Tabelle 2: Endpunkte des Samplly.MDR

Die verwendeten Endpunkte ermöglichen somit den benötigten Funktionsumfang im Kontext der Semantik zu nutzen und bieten sogar weitere Informationen, die im späteren Funktionsumfang weiter erläutert werden.

Durch die somit vorhandene Abfragefunktion der semantischen Informationen wurden Endpunkte für das Frontend zur Bereitstellung benötigt. Diese dienen dem primären Funktionsumfang und sollen sowohl die Anzahl an Vorkommen, als auch die dazugehörige Semantik bereitstellen. Die nachfolgende Tabelle 3 verdeutlicht die Schnittstellen inklusive deren Funktion.

Die genaue Funktionsweise und die Sinnhaftigkeit dieser Endpunkte werden im Kapitel 4.6.4, dem Funktionsumfang des eMDR, genauer definiert.

Endpunkt	Nutzung / Auswirkung
<i>/Observation</i>	Extrahiert eine Observation (FHIR-Ressource) auf Basis einer übermittelten pseudonymisierten Patienten-ID.
<i>/Observation/vital-signs</i>	Extrahiert eine Observation im Kontext der Vitalparameter (FHIR-Ressource) auf Basis einer übermittelten pseudonymisierten Patienten-ID.
<i>/Procedure</i>	Extrahiert eine Procedure (FHIR-Ressource) auf Basis einer übermittelten pseudonymisierten Patienten-ID.
<i>/DiagnosticReport</i>	Extrahiert einen Diagnostic Report (FHIR-Ressource) auf Basis einer übermittelten pseudonym. Patienten-ID.
<i>/Patient</i>	Bei einem GET wird der Patient auf Basis einer übermittelten pseudonymisierten Patienten-ID als FHIR-Export übermittelt. Bei einem POST wird der Patient in das System integriert.
<i>/Bundle</i>	Bei einem GET wird ein Bundle der gesamten Patienteninformationen auf Basis einer übermittelten pseudonym. Patienten-ID als FHIR-Export übermittelt. Bei einem POST wird das Bundle verarbeitet und in das System integriert.
<i>/Totalcount</i>	Stellt die insgesamt innerhalb des eMDR vorrätigen Diagnosen, Prozeduren und Laborwerte (summiert) für das Frontend bereit.
<i>/Count</i>	Extrahiert basierend auf einem übermittelten Kode und Katalog die Anzahl an Vorkommen sowie die Semantik.
<i>/Join</i>	Extrahiert basierend auf zwei übermittelten Kodes und Katalogen inklusive Mengenoperation die Anzahl an Vorkommen sowie die die jeweilige Semantik.
<i>/getQuery</i>	Extrahiert basierend auf einer übermittelten UUID eine archivierte Auswertung.
<i>/Plot</i>	Übermittelt die Anzahl an Vorkommen pro Jahr an das Frontend zur Darstellung eines Balkendiagramms auf Basis eines übermittelten Kodes inklusive des Kataloges.
<i>/Boxplot</i>	Übermittelt eine vorformatierte Plot-Anweisung für das Tool plotly an das Frontend zur Darstellung eines Box-Plot auf Basis eines übermittelten LOINC-Kodes.

/RevokeConsent	Entfernt restlos sämtliche Vorkommen eines Patienten auf Basis einer übermittelten pseudonymisierten Patienten-ID aus dem eMDR und übermittelt die entfernte Anzahl an Vorkommen.
----------------	---

Tabelle 3: Endpunkte des eMDR

4.6.3 Frontend

Das Frontend wurde mittels Vue.js und dem Framework Vuetify im Sinne einer Single Page Application (SPA) entwickelt. Die primären Herausforderungen stellten den Login, mittels bereits in dem Datenintegrationszentrum vorhandener Authentifizierungsmethoden und der simplen übersichtlichen Bedienung, dar. Der Login konnte direkt mit der Anbindung des lokalen Keycloak-Dienstes ermöglicht werden.

The screenshot shows the configuration page for a Keycloak client named 'eMDR'. The configuration includes the following fields and settings:

- Client ID:** eMDR
- Name:** (empty)
- Description:** (empty)
- Enabled:** ON
- Always Display in Console:** OFF
- Consent Required:** OFF
- Login Theme:** (empty)
- Client Protocol:** openid-connect
- Access Type:** public
- Standard Flow Enabled:** ON
- Implicit Flow Enabled:** OFF
- Direct Access Grants Enabled:** ON
- Root URL:** http://FrontendURL:Port
- * Valid Redirect URIs:** http://FrontendURL:Port/*
- Base URL:** (empty)
- Admin URL:** http://FrontendURL:Port/*
- Web Origins:** http://FrontendURL:Port
- Backchannel Logout URL:** (empty)
- Backchannel Logout Session Required:** ON
- Backchannel Logout Revoke Offline Sessions:** OFF

Abbildung 18: Beispielkonfiguration des Keycloak Clients für das eMDR (Screenshot aus eigener Systeminstanz)

Hierzu wurde innerhalb der Keycloak Instanz ein neuer Client für das eMDR angelegt und entsprechend konfiguriert. Die Konfiguration ist in Abbildung 18 ersichtlich. Im Anschluss wurde zu Testzwecken Mitarbeitern des DIZ die Rechte zu dem eMDR, mittels Clientzuweisung, eingerichtet und der Login konnte erfolgreich mit den bestehenden Login-Daten der Mitarbeiter durchgeführt werden.

Im Sinne der Authentifizierung wurde frontendseitig auch zusätzlich das Token der Authentifizierung verarbeitet, welcher im Zuge der Nutzung des eMDR für die Schnittstelle zur DIP benötigt wird. Die Verwendung wird in Kapitel 4.6.5 der Datenintegrationsplattform erläutert.

Aufgrund der Vorgabe, dass das Frontend möglichst simpel und übersichtlich gestaltet sein sollte, wurden lediglich drei Bereiche definiert. Der erste Bereich dient einer informellen Übersicht über den Gesamtbestand an Daten innerhalb des Systems, welcher in Abbildung 19 verdeutlicht wird. Diese Daten werden in einem bestimmten Falle in Bezug auf die Systeminstallation leicht anonymisiert. Der Hintergrund wird in Kapitel 4.7 der multiplen Systeminstallationen weiter erläutert.

Der zweite Bereich wird durch die typische Navigationsbar definiert und verfügt im derzeitigen Entwicklungsstand über die Möglichkeit der einfachen Abfrage, der einfachen Mengenoperation und einer Funktion für die interne Transferstelle. Zu Hilfezwecken existiert noch ein Navigationspunkt für die Themen Hilfe und Kontakt, welcher konkret auf das Datenintegrationszentrum als Ansprechpartner verweist. Der dritte Bereich stellt den Mittelpunkt der Benutzeroberfläche dar und bietet einen direkten Einstiegspunkt in die Auswertungsumgebung. Die einfache Suche erfragt direkt einen Code und den korrespondierenden Katalog ab und bietet das konkrete Resultat. Die Navigationsleiste sowie der Auswertungsbereich sind in Abbildung 19 ersichtlich.



Abbildung 19: Funktionsumfang und Eingabeüberprüfung des Frontends (Screenshot aus eigener Systeminstanz)

Das Frontend nutzt zur Bereitstellung der Funktionalitäten die entwickelten API-Schnittstellen des Backends und konsumiert diese mittels dem Tool *axios*. Die für das Frontend benötigten Daten werden dementsprechend auch bereits durch das Backend

aufgearbeitet und in der benötigten Form bereitgestellt, um das Frontend weiterhin leichtgewichtig und so reduziert wie möglich zu halten. Sämtliche Eingaben über die Oberfläche sind als Formularfelder realisiert und werden über das Frontend auf einfache Plausibilität geprüft. Diese vermeidet zunächst jedoch nur leere Eingaben im Kontext der Codes sowie die falsche Eingabe der Kataloge / Terminologien, welches über ein Dropdown gelöst wurde. Dies ist ebenfalls in Abbildung 19 der einfachen Suche ersichtlich.

Der weitere Funktionsumfang, welcher über das Frontend in Zusammenspiel mit dem Backend realisiert wurde, ist im nachfolgenden Kapitel definiert.

4.6.4 Funktionsumfang

Die grundlegenden Funktionen des eMDR werden über das Backend mittels einer Datei für Umgebungsvariablen (env-Datei) angepasst. Innerhalb der Datei wird nicht nur die Datenbankverbindung eingerichtet, sondern auch eine Unterscheidung des Systems in *Routine*, *DIZ* und *Forschung* getroffen. Der Hintergrund sowie die Auswirkung dieser Unterscheidung ist in Kapitel 4.7 erläutert. Weiter werden innerhalb der env-Datei die benötigten Datenelemente des Smply.MDR hinterlegt, welche dadurch für die API-Abfragen verwendet werden können. Das Frontend verfügt ebenfalls über eine env-Datei, welche die URL, beziehungsweise IP-Adresse des Backends und die Verbindungsinformationen des anzubindenden Keycloak-Dienstes vorhält. Durch die derzeitige Anbindung des lokalen Keycloak-Dienstes ist ein Login für DIZ-Mitarbeiter, aufgrund der hinterlegten DIZ-LDAP-Verbindung, nach der entsprechenden Rechtezuweisung, möglich. Eine Anbindung an das Active Directory des GIFONET, um einen Zugang für alle forschenden Mitglieder des Fachbereichs Medizin zu ermöglichen, konnte erfolgreich umgesetzt werden.

Weiter verfügt das Backend über diverse Logging-Funktionalitäten, welche sich nicht nur auf Systeminformationen, sondern auch auf die durchgeführten Abfragen beziehen. Die Systemlogs beziehen sich zum einen auf die durchgeführten Datenbankoperationen und lassen einen Rückschluss auf gegebenenfalls aufgetretene Verbindungsfehler zu, zum anderen wird die FHIR-Schnittstelle komplett protokolliert und verdeutlicht welche Datensätze integriert und exportiert wurden. Die letzte und wichtigste Logging-Funktion stellt die Protokollierung der durchgeführten Abfragen dar. Jede durchgeführte Abfrage wird innerhalb der Datenbank in der Tabelle *query_archive* protokolliert, beziehungsweise archiviert. Die gespeicherten Informationen beziehen sich hierbei auf den Code, den Katalog, einem Zeitstempel der Durchführung, der Installations-Status

des Systems (Routine, DIZ oder Forschung) sowie die aus der Auswertung gewonnenen Resultate, inklusive der systeminternen pseudonymisierten Patienten-Identifikatoren. Die Archivierung der Abfragen wird insbesondere für die DIZ interne Transferstelle benötigt, der Kontext der Nutzung dieser Informationen wird im folgenden Verlauf weiter erläutert.

Die wichtigste Funktionalität des eMDR wurde mittels der einfachen Abfrage von Vorkommen realisiert. Wie bereits im Kontext des Frontends verdeutlicht, erfolgt eine simple Abfrage nach einem Kode inklusive Klassifikation/Terminologie. Das Resultat daraus wird mittels der Anzahl an Vorkommen sowie der dazugehörigen Semantik visualisiert. Im Zuge dessen wurde eine weitere Funktion umgesetzt, welche die semantische Abhängigkeit beziehungsweise Hierarchie des abgefragten Kodes innerhalb der Klassifikation verdeutlicht. Die gesamte Darstellung einer beispielhaften Abfrage, inklusive der dazugehörigen semantischen Hierarchie, ist in Abbildung 20 ersichtlich.

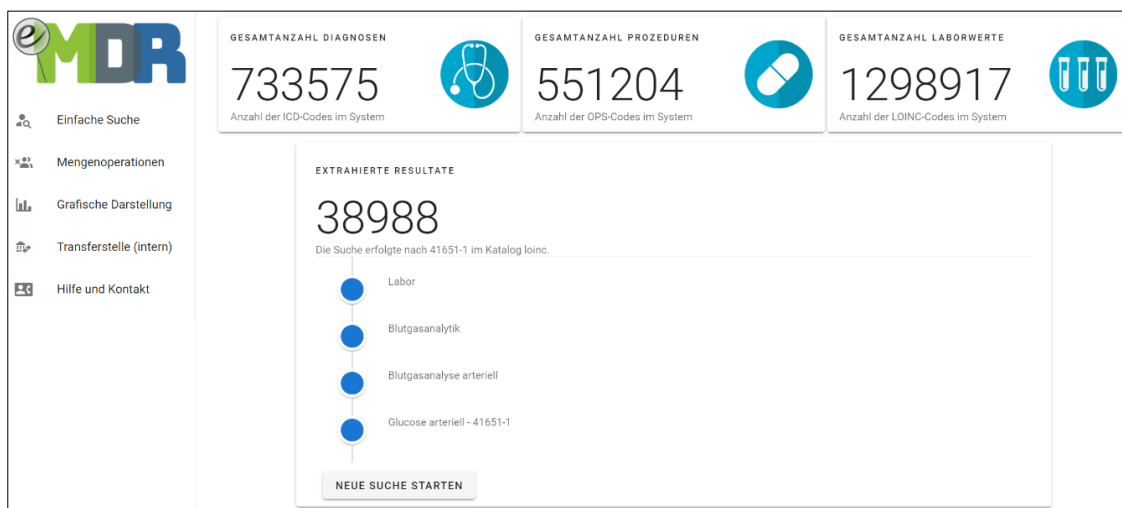


Abbildung 20: Darstellung der Resultate einer Beispielabfrage (Screenshot aus eigener Systeminstanz)

Als zusätzliche Funktion wurde eine komplexere Suche implementiert, welche die Möglichkeit bietet, die Mengenoperationen *Vereinigung*, *Schnittmenge* und *symmetrische Differenz* auf zwei Suchparameter anzuwenden (Abbildung 21). Die Ergebnisse werden analog zur einfachen Suche inklusive beider semantischer Hierarchien dargestellt.

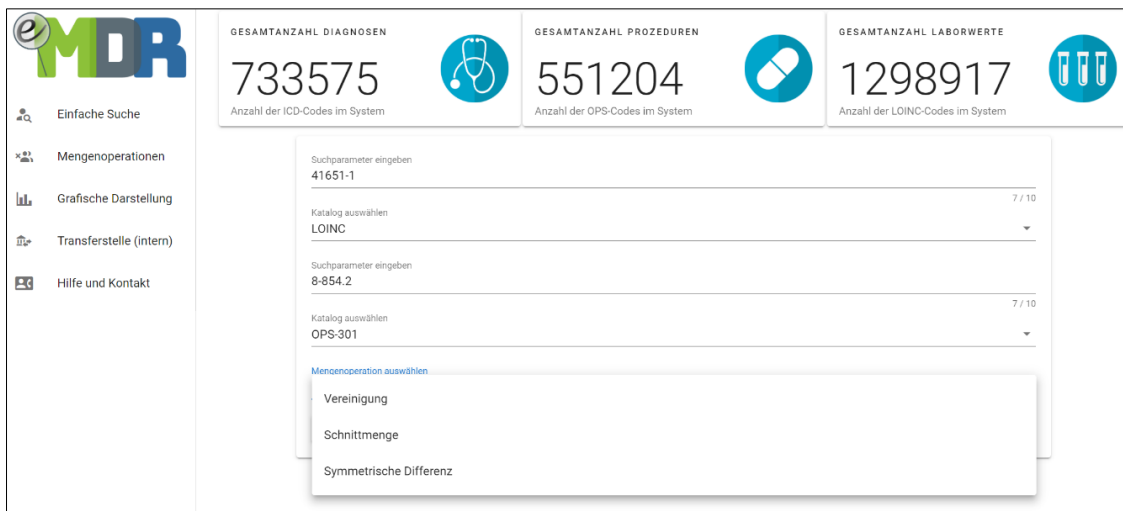


Abbildung 21: Funktionsumfang der Mengenoperationen (Screenshot aus eigener Systeminstanz)

Eine weitere Funktion stellt die Plot-Funktionalität dar, welche mittels dem Python Paket `plotly` umgesetzt wurde. Dies ermöglicht eine grafische Darstellung, welche über integrierte Funktionen auch als Bildausschnitt separat abgespeichert werden kann. Weiter ermöglicht `plotly` Ausschnitte der Grafik zu erzeugen und mittels Mouse-over-Effekt die entsprechenden Werte anzuzeigen. Grundsätzlich lassen sich alle Auswertungen mittels eines Balkendiagramms darstellen, wie in Abbildung 22 verdeutlicht wird.

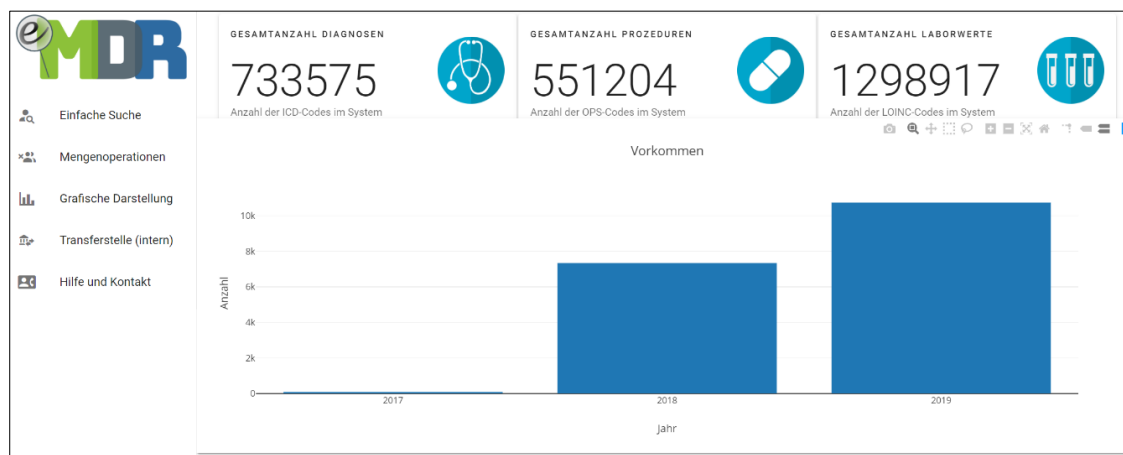


Abbildung 22: Verwendung von Plotly zur grafischen Ergebnisdarstellung (Screenshot aus eigener Systeminstanz)

Weiter existiert die Möglichkeit die Darstellung von LOINC-Anfragen von einem Balkendiagramm zu einem Box-Plot umzustellen. Hierfür dient die in Kapitel 4.6.1 beschriebene Tabelle `lab_counter_values`, wodurch der jahresbasierte Box-Plot generiert werden kann.

Die letzte und für den Ablauf innerhalb des Datenintegrationszentrums wichtigste Funktionalität, stellt die Option der Transferstelle dar. Die Funktion basiert auf der

beschriebenen Archiv- beziehungsweise Logging-Funktionalität des Backends und versieht jede durchgeführte Abfrage bei der Archivierung mit einer sogenannten UUID der Version 4, einer (pseudo)zufällig generierten ID, welche intern an die Datenintegrationsplattform übermittelt wird und von Mitarbeitern der Transferstelle geprüft werden kann. Dies dient zum einen der Überprüfung der Auswertung auf Korrektheit (fälschungssicher) und zum anderen dem Zugriff auf die hinterlegten Patienten-Identifikatoren, welche die Grundlage der Arbeit der Transferstelle bilden. Eine beispielhafte Überprüfung ist in Abbildung 23 ersichtlich und verdeutlicht den Umfang der bereitgestellten Daten.

Der weitere und abschließende Funktionsumfang wird im nachfolgenden Kapitel weiter beschrieben.

The screenshot shows the MDR (Medical Data Repository) interface. At the top, there are three summary cards:

- GESAMTANZAHL DIAGNOSEN:** 733575 (Anzahl der ICD-Codes im System)
- GESAMTANZAHL PROZEDUREN:** 551204 (Anzahl der OPS-Codes im System)
- GESAMTANZAHL LABORWERTE:** 1298917 (Anzahl der LDINC-Codes im System)

Below these cards is a section titled "ARCHIVIERTE QUERY" displaying a JSON query object:

```

{
  "code": "995faf73-8bef-4f1d-ac02-95cfc00e3bee",
  "query": {
    "catalogue": "icd",
    "definition": "A01.0, Typhus abdominalis",
    "counter": 6,
    "tree": [
      "I, Bestimmte infektiöse und parasitäre Krankheiten",
      "A00-A09, Infektiöse Darmkrankheiten",
      "A01, Typhus abdominalis und Paratyphus",
      "A01.0, Typhus abdominalis"
    ]
  },
  "mpis": [
    2019043135,
    2019193988,
    2019192004,
    2019022187,
    2019053391
  ]
},
  "env": "diz",
  "timestamp": "2021-08-31 07:29:41.551355+02:00"
}

```

Abbildung 23: Darstellung der Query-Überprüfung durch die Transferstelle (Screenshot aus eigener Systeminstanz)

4.6.5 Integration in die Datenintegrationsplattform

Eine der wichtigsten und abschließenden Funktionalität stellt die Datenübermittlung an die Datenintegrationsplattform dar. Im Zuge der Entwicklung des eMDR wurde auf Basis der Vorgaben evaluiert, inwiefern eine einfache und benutzerfreundliche Übermittlung der Daten und somit Integration erfolgen kann. Zu diesem Zweck existiert die Möglichkeit nach jeder Abfrage die Option „DIP Export“ auszuwählen.

The screenshot shows a modal window with a blue header containing a close icon and the title "In einen Datensatz der DIP speichern". Below the header, there are three form fields. The first is a dropdown menu labeled "Domäne" with the selected value "Institute of Medical Informatics". Below it is a small instruction: "Um den Datensatz auswählen zu können, in den Sie die Daten hochladen möchten, müssen Sie zuerst die Domäne auswählen, in der sich der Datensatz befindet." The second field is another dropdown menu labeled "Datensatz" with the selected value "Study example 1". Below it is the instruction: "Wählen Sie den Datensatz aus, in den die Daten hochgeladen werden sollen." The third field is a text input labeled "Bezeichnung der Daten" containing the text "Feasability export study 1". Below it is the instruction: "Geben Sie eine eindeutige Bezeichnung der Daten an". At the bottom right of the form is a green button with the text "SPEICHERN".

Abbildung 24: Export der Resultate zur DIP (Screenshot aus eigener Systeminstanz)

Diese Option öffnet ein entsprechendes Pop-up, welches die Auswahl der Benutzerdomäne (bei mehreren verfügbaren Domänen) sowie eines Datasets innerhalb der DIP ermöglicht, wie in Abbildung 24 ersichtlich.

Datasets stellen die Grundlage der DIP dar und repräsentieren die unterschiedlichen Arbeitsumgebungen der Forschenden. Diese direkte Auswahl der Datasets wird ermöglicht, indem das Authentifizierungstoken von Keycloak genutzt wird, um die API der DIP anzusprechen. Aufgrund der Tatsache, dass sich der Login des eMDR und der Login der DIP dem gleichen Keycloak bedienen, ist die Token-Authentifizierung zwischen den beiden Systemen technisch möglich. Nachfolgend eine tabellarische Auflistung der genutzten API-Endpunkte der DIP.

Endpoint	Nutzung / Auswirkung
/profile/rights	Extrahiert die Rechte innerhalb der Plattform für den authentifizierten Benutzer.
/datasets	Extrahiert eine Liste an verfügbaren Datensets, auf die der authentifizierte Benutzer Zugriff hat.
/data	Erzeugt ein neues Datenelement.
/domain/switch/	Wechselt, sofern die Rechte ausreichend sind, die aktuelle Domäne des authentifizierten Benutzers.

Tabelle 4: Verwendeten Endpunkte der Datenintegrationsplattform

Durch diese Funktionalitäten existiert eine direkte Kopplung des eMDR mit dem Kernsystem des Datenintegrationszentrums und ermöglicht eine Weiterverarbeitung der extrahierten Daten im Kontext der Antragsstellung bei Forschungsvorhaben. Forschende können nach erfolgreicher Übertragung die Daten über die DIP einsehen. Dies ist in der nachfolgenden Abbildung 25 und Abbildung 26 ersichtlich.

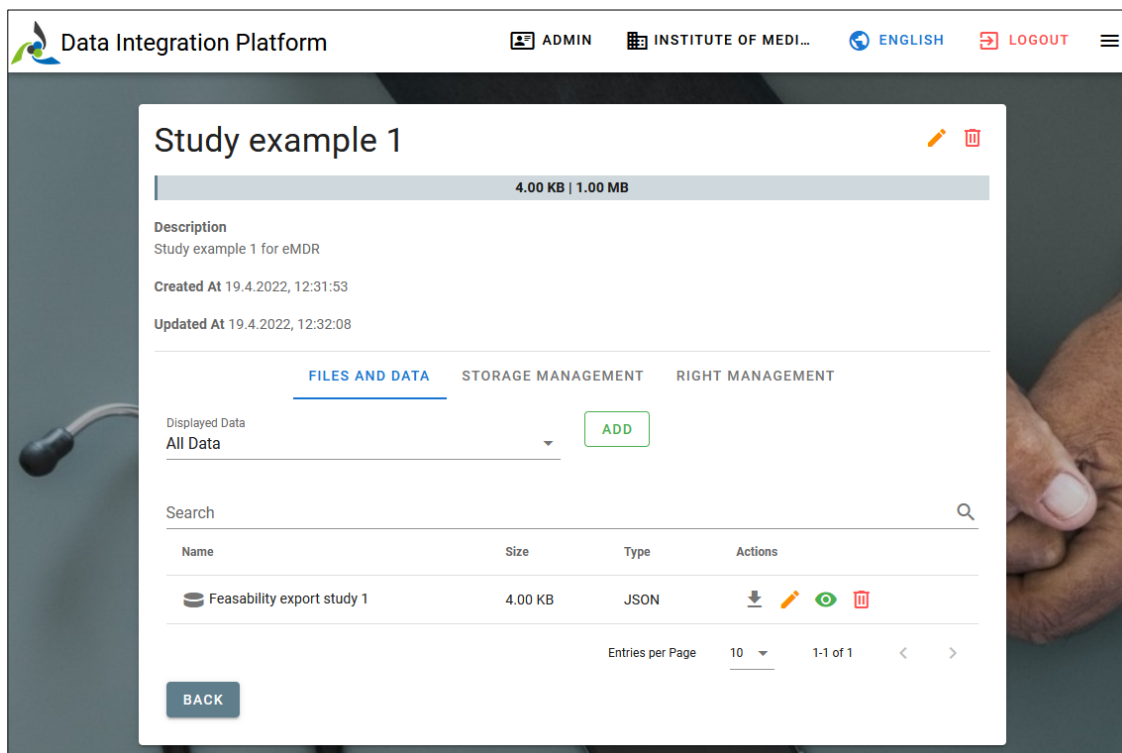


Abbildung 25: Ansicht der Feasibility Abfrage innerhalb der DIP (Screenshot aus eigener Systeminstanz)

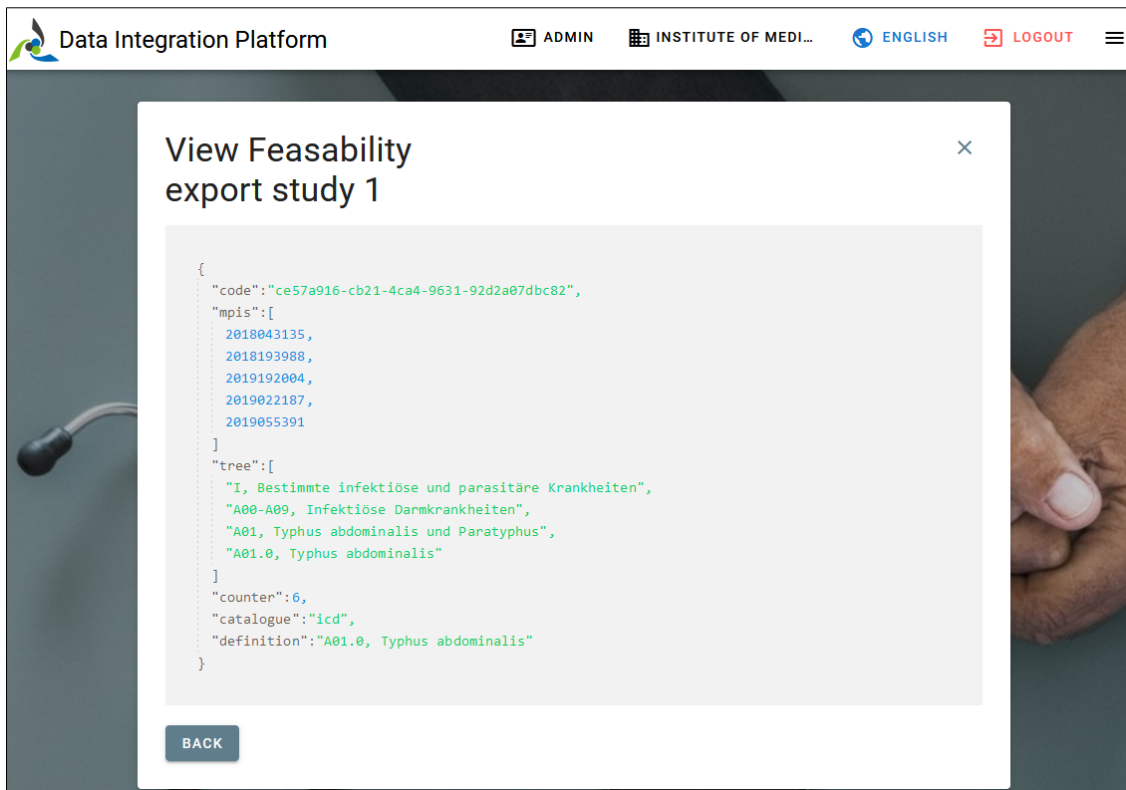


Abbildung 26: Exportierte Feasibility-Abfrage als JSON innerhalb des DIP Datensets (Screenshot aus eigener Systeminstanz)

4.7 Multiple Systeminstallationen

Das eMDR muss speziell aus datenschutzrechtlichen Gründen an unterschiedlichen Stellen, innerhalb der Infrastrukturen des Standortes Gießen, unterschiedliche Informationen enthalten. Hierbei ist es essenziell, dass sensible Patienteninformationen keinesfalls den klinischen Routinebereich verlassen. Des Weiteren ist auch die Anonymisierung von Daten für Abfragen im Kontext der Forschung und somit außerhalb der klinischen Routine von Nöten. (Abbildung 27)

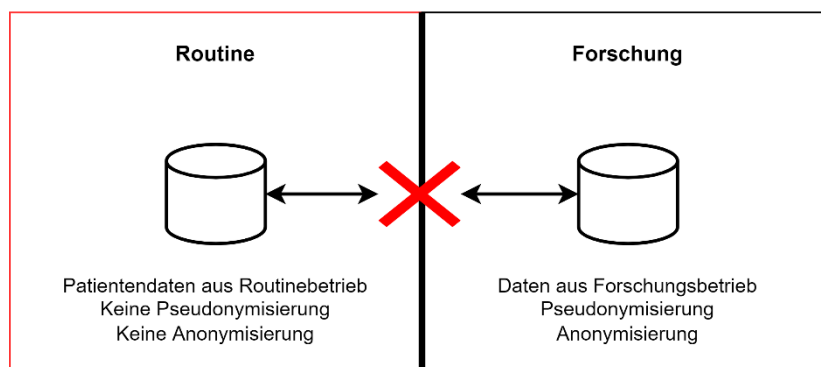


Abbildung 27: Übersicht der Infrastrukturen im Hinblick auf den Datenschutz (Eigene Darstellung)

Aus Gründen der IT-Systemsicherheit wurde sich für multiple Systeminstallationen entschieden. Dies erzeugt in sich geschlossene Systeme, wodurch beispielsweise gesorgt wird, dass bei einem Sicherheitsvorfall innerhalb der Instanz im Forschungsnetzwerk (GIFONET) keine Daten der klinischen Routine-Instanz betroffen sind. Zu diesem Zweck wurde ein Ampelsystem entwickelt, welches in Abbildung 28 ersichtlich ist.

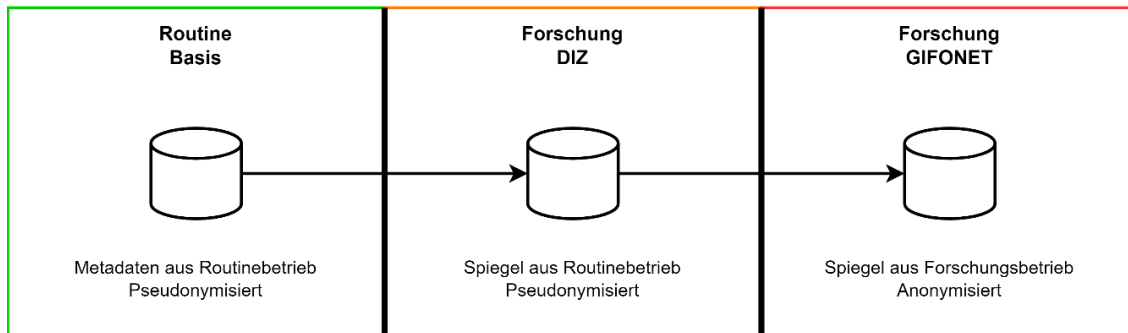


Abbildung 28: Ampelsystem zum Zwecke der Systemtrennung am Standort (Eigene Darstellung)

Die „grüne“ Installation des eMDR stellt die Installation innerhalb der klinischen Routineumgebung dar. Die systeminternen Patienten-Identifikatoren sind wie in Kapitel 4.6.4 beschrieben ohnehin pseudonymisiert, jedoch herrschen innerhalb dieser Instanz keinerlei Beschränkungen bei der Abfrage von Informationen. Dies bedeutet, dass die genaue Anzahl identifizierter Vorkommen über die Benutzeroberfläche bereitgestellt werden und im Hintergrund die internen pseudonymisierten vorgehalten werden. Diese IDs können selbstverständlich nur im Kontext der standortinternen ID-Management-Lösung genutzt werden. Weiter dient die Instanz als initialer Datenursprung und wird von den einzelnen Kommunikationskanälen mit den klinischen Informationen / Metadaten bespielt.

Die „gelbe“ Installation stellt die Instanz des Datenintegrationszentrums dar und ist inhaltlich identisch mit der Installation in der klinischen Routine. Dieses System dient der reinen Durchführung von Machbarkeitsanfragen im Kontext des Ablaufs einer Forschungsfragestellung. Die genaue Nutzung und der dazugehörige rechtlich-organisatorische Ablauf in diesem Kontext ist in Kapitel 4.9 ersichtlich. Das System hält ebenfalls die internen IDs vor und speichert diese im Rahmen der Abfrage-Historie ausschließlich für die standorteigene Transferstelle. Die Transferstelle dient DIZ-intern der Extraktion von Daten aus der klinischen Routine und der Forschung für genehmigte Forschungsanfragen und betrachtet dabei insbesondere die Aspekte der Pseudonymisierung und Anonymisierung.

Die abschließende „rote“ Installation stellt eine optionale und anonymisierte Installation innerhalb des Gießener Forschungsnetzwerkes für eine externe Nutzung dar. Die Inhalte der Datenbasen sind zunächst identisch mit den anderen Instanzen, jedoch werden die pseudonymisierten IDs aus dem System entfernt. Weiter werden Ergebnisse von Abfragen auf der Benutzeroberfläche anonymisiert. Dies sorgt dafür, dass die Anzahl an Vorkommen prinzipiell gerundet dargestellt wird und Resultate unter einem definierten Schwellwert nicht dargestellt werden. Aufgrund der Tatsache, dass laut der Datenzugriffs- und Nutzungsordnung des Standortes Gießen alle Forschungsanfragen über eine entsprechende Antragsplattform abgewickelt werden müssen, kommen derzeit nur die Instanzen „grün“ (klinische Routine) und „gelb“ (Datenintegrationszentrum - GIFONET) zum Einsatz. Die „rote“ Instanz (GIFONET extern) wird zukünftig zunächst für standortübergreifende Abfragen durch externe Datenintegrationszentren benötigt.

Zur Wahrung der Patientenrechte muss das eMDR, bei einem Widerruf der Datennutzung durch den Patienten, die betroffenen Datensätze restlos entfernen können. Dies wäre aus Sicht der multiplen Instanzen problematisch, da speziell bei der anonymisierten Instanz ein Rückschluss auf die betroffenen zu entfernenden Datensätze unmöglich wäre. Zu diesem Zweck wird eine Installation der Software *Airbyte* verwendet. Die Software beginnt bei der initialen Routineinstanz des eMDR und synchronisiert die gesamte Datenbasis täglich (daily full sync) in die Instanz des Datenintegrationszentrums. Dieses Vorgehen stellt eine PostgreSQL-Datenbankreplikation dar. Hierbei wird die Synchronisation stets vollständig protokolliert und kann über die Benutzeroberfläche von *Airbyte* überwacht werden. Im Anschluss wird die synchronisierte Datenbasis der DIZ-Instanz in die Instanz des GIFONET extern synchronisiert. Dort werden mittels Datenbank-Operationen sämtliche internen Patienten-IDs entfernt, wodurch die Anonymisierung gewährleistet wird. Im Falle eines Widerrufs durch einen Patienten werden somit lediglich die definierten Datensätze aus der Instanz innerhalb der klinischen Routine entfernt. Diese Information, konkret die zu entfernende ID mit allen zugehörigen Datensätzen, muss das eMDR von Seiten des ID- und Consent-Managements erhalten. Aufgrund der vollständigen Synchronisation werden somit täglich die Datenbasen der DIZ und GIFONET Instanzen überschrieben, wodurch auch die betroffenen Datensätze entfernt werden. Dieser Ablauf wird in der nachfolgenden Abbildung 29 verdeutlicht.

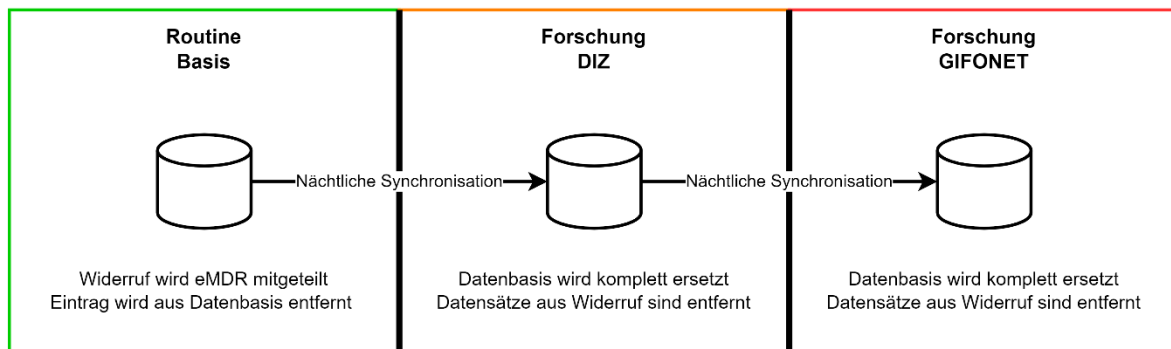


Abbildung 29: Datensynchronisation und Auswirkung im Kontext des Ampelsystems (Eigene Darstellung)

4.8 Containerbasierte Bereitstellung

Die benötigten Systemkomponenten werden isoliert in Form von Containern bereitgestellt. Eine Installation, beziehungsweise das Ausrollen der Container, erfolgt mittels der Virtualisierungssoftware Docker oder Podman. Zunächst werden die einzelnen Komponenten des Samply.MDR benötigt, welche durch gewartete Installationsanleitungen, sogenannte Images, bereitgestellt werden. Hierbei sind vor Erstellung der Container einige Anpassungen vorzunehmen, welche sich primär auf Login-Daten für die Datenbasis des Samply.MDR und Verbindungsinformationen zu lokalen Authentifizierungsdiensten (Keycloak) beziehen. Aufgrund der Tatsache, dass Samply.MDR ein primäres Werkzeug des Datenintegrationszentrums darstellt, entfällt diese Installation bei weiteren gewünschten Instanzen des eMDR. Hierfür wird stets auf eine der beiden zentralen Instanzen von Samply.MDR zurückgegriffen, welche sowohl im GIFONET, als auch in der klinischen Routineumgebung existieren.

Das eMDR selbst besteht ebenfalls aus einer gebündelten Installationsanleitung (docker-compose), welche separat die benötigten Dienste bereitstellt. Hierbei wird zunächst ein Container für die benötigte PostgreSQL-Datenbank bereitgestellt und mittels definierter Umgebungsvariablen bereits mit einer Datenbank, einem Datenbankbenutzer und einem entsprechenden Passwort initialisiert. Diese Anpassungen werden in einer separaten Datei für Umgebungsvariablen, der sogenannten „.env“-Datei definiert. Die weiteren Container stellen abschließend das Backend und das Frontend dar. Der Backend Container basiert auf einer leichtgewichtigen Python Installation, welche mittels Flask den benötigten Funktionsumfang bereitstellen kann und auf dem zuvor erstellten Datenbankcontainer basiert. Der Frontend Container bietet eine leichtgewichtige Vue.js Installation. Beide Container werden mit sogenannten Volumes initialisiert, wodurch definierte Ordner des Host-Betriebssystems mit Ordnern innerhalb der Container gekoppelt werden können. Dieses Vorgehen wird bei dem eMDR dazu verwendet, um den benötigten Quellcode

des Systems in die entsprechenden Container dynamisch integrieren und ändern zu können und um wichtige Systemprotokolle (Logs) aus dem Container auslesen zu können. Dies ermöglicht eine kontinuierliche Wartung der einzelnen Komponenten und unterstützt die Fehleranalyse durch Sichtung der Systemprotokolle.

Die verwendeten Umgebungsvariablen für die Datenbankinitialisierung wurden für die Container des Backends und Frontends um diverse Voreinstellungen erweitert. Hierbei ist der Name der erstellten Container anpassbar, um bei multiplen Installationen eine Abgrenzung erzeugen zu können. Weiter wird für jede erstellte Installation des eMDR ein eigenes virtuelles Subnetz erstellt, worin die Container stets definierte IP-Adresse erhalten. Das Subnetz und somit der Netz-Präfix ist hierbei dynamisch anpassbar, die Container selbst erhalten jedoch stets die gleichen Host-IP-Adressen. Hierbei erhalten alle Datenbank-Container die Endung 1 (XXX.XXX.XXX.1), das Backend erhält die Endung 2 (XXX.XXX.XXX.2) und das Frontend abschließend die Endung 3 (XXX.XXX.XXX.3). Durch dieses Schema sind aus Sicht des Netzwerkverkehrs und gegebenenfalls benötigter Analysen eine schnelle Identifikation der einzelnen Komponenten möglich (Abbildung 30).

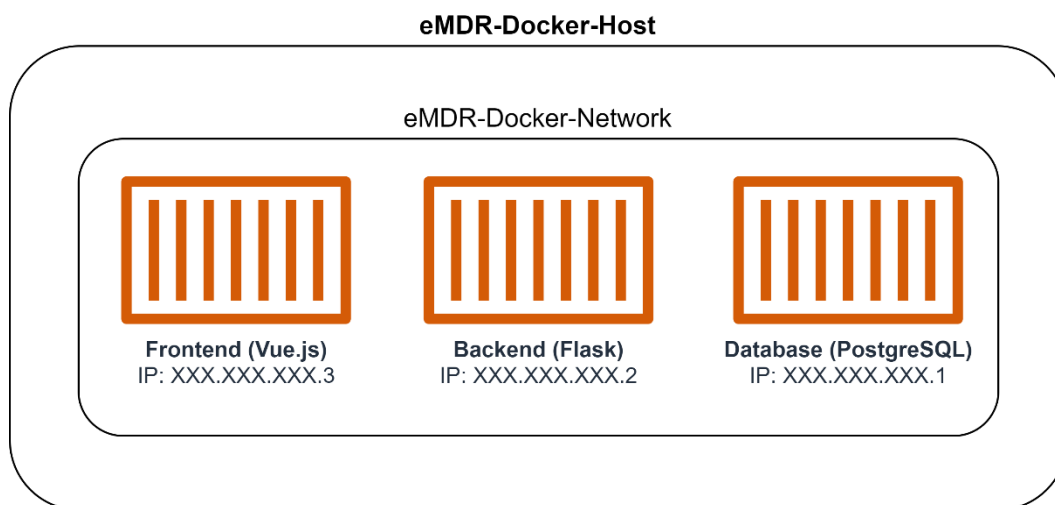


Abbildung 30: Darstellung der Komponenten der Container-Infrastruktur (Eigene Darstellung)

Eine reine Installation der benötigten Infrastruktur für Frontend und Backend ist durch dieses Verfahren mit lediglich einem Befehl möglich. Abschließend sind dann die Integration des Quellcodes und das Erstellen einer Python Umgebung für das Backend erforderlich.

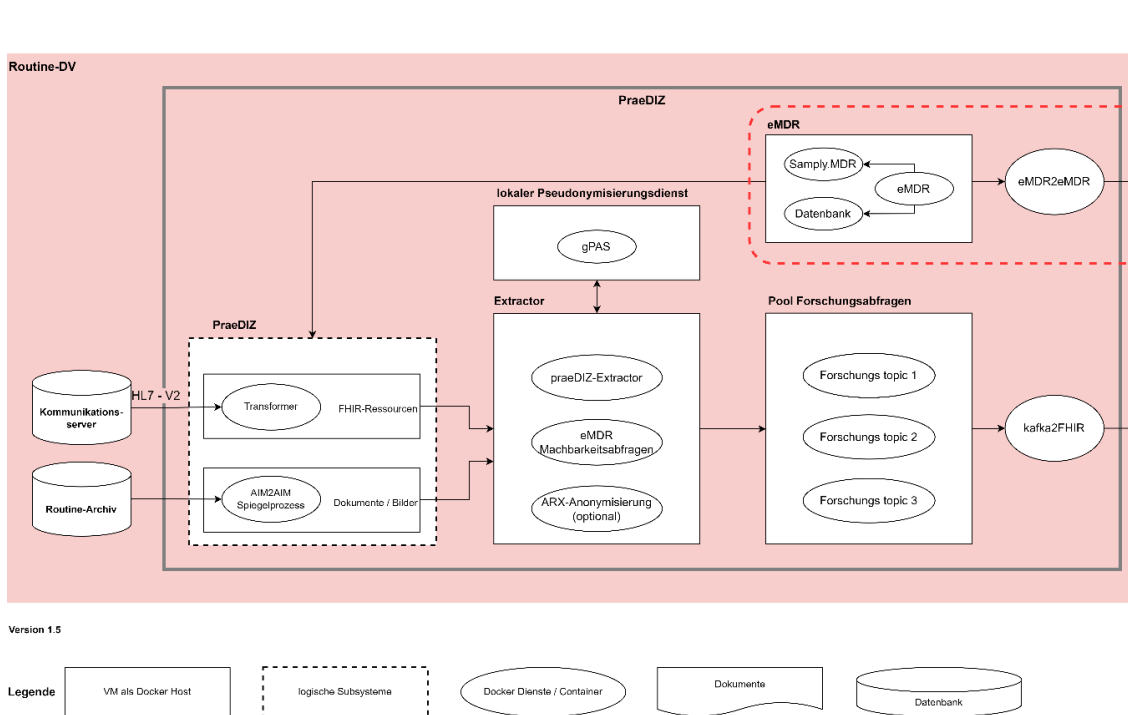


Abbildung 32: Ausschnitt des technischen DIZ-Workflows - Routinebereich (Eigene Darstellung des DIZ)

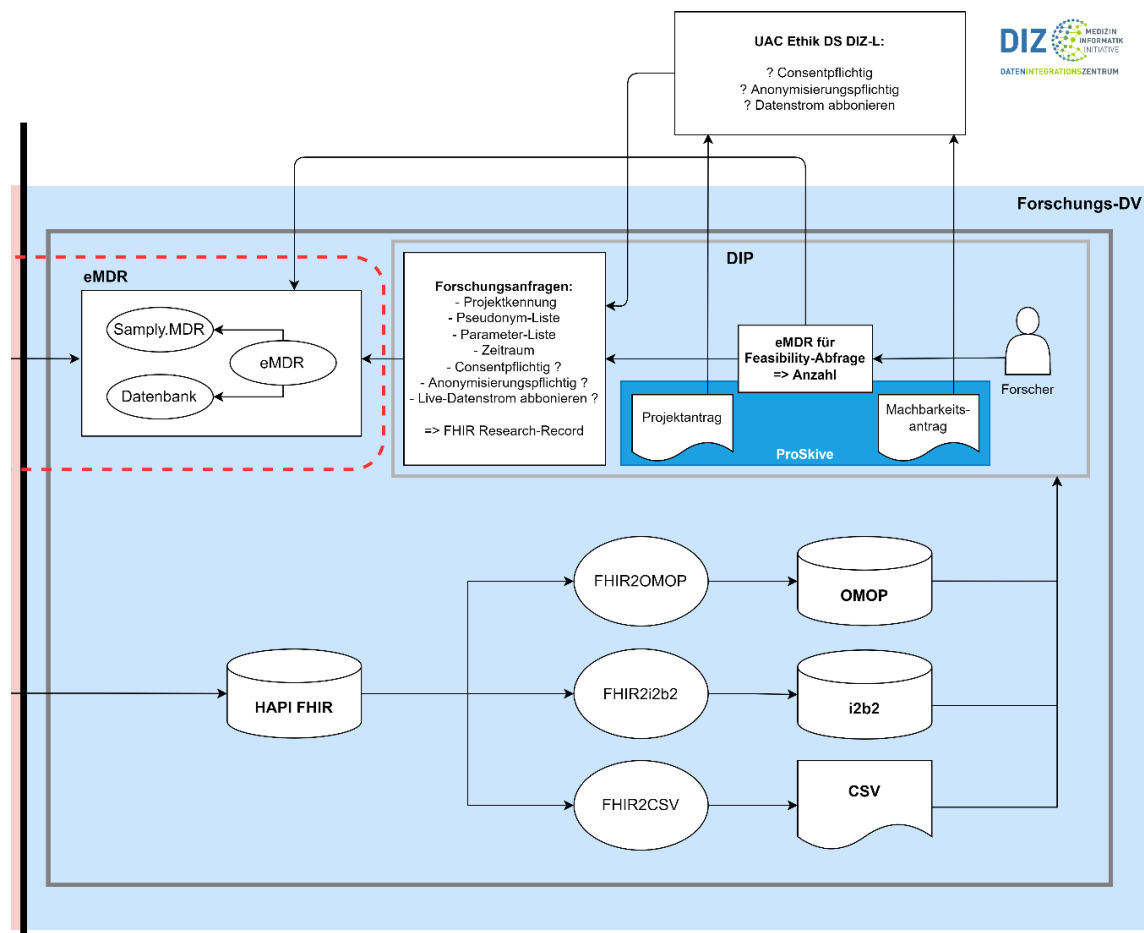


Abbildung 33: Ausschnitt des technischen DIZ-Workflows - Forschungsbereich (Eigene Darstellung des DIZ)

In Abbildung 31 ist die technische Sicht des DIZ dargestellt und verdeutlicht insbesondere die, in dem Kapitel 4.7 beschriebene, multiple Systeminstallation, inklusive des Spiegelprozesses der Datenbank. Abbildung 34 dient der Bereitstellung des relevanten Detailausschnittes.

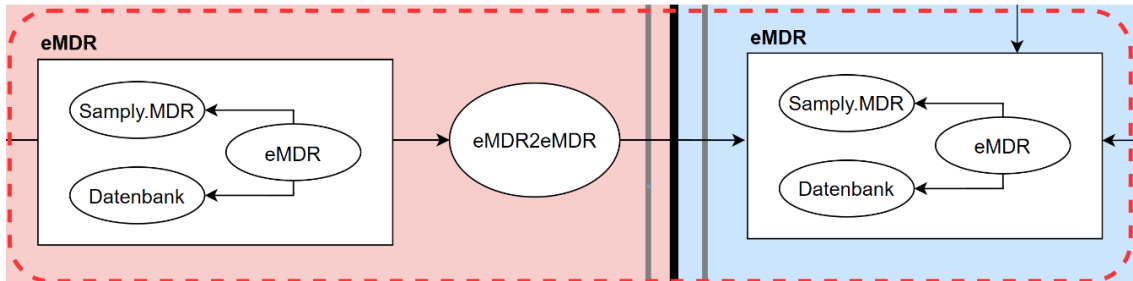


Abbildung 34: eMDR relevanter Ausschnitt des technischen DIZ-Workflows (Eigene Darstellung des DIZ)

Das eMDR dient aus dieser technischen Sichtweise als Informationsgeber für die Transferstelle, indem die entsprechenden Patienten-Identifikatoren für die Extraktionen verwendet werden können. Dies ist im derzeitigen manuellen Arbeitsauflauf mittels der Kopplung zwischen der Datenintegrationsplattform und dem eMDR realisiert. Wie im Funktionsumfang des eMDR in Kapitel 4.6.4 verdeutlicht, werden die innerhalb des Systems extrahierten Resultate in Form einer UUID bereitgestellt und an die DIP für die Forschenden übermittelt. Bei Antragsstellung wird diese UUID angegeben und nach Freigabe durch das UAC von der Transferstelle im eMDR geprüft. Die Prüfung ermöglicht eine Einsicht in die resultierenden Patienten-Identifikatoren, wodurch eine manuelle Extraktion durch die Transferstellenmitarbeiter ermöglicht wird. Der weitreichende Ablauf für Forschungsvorhaben aus Sicht der Forschenden ist in Abbildung 35 ersichtlich. Abbildung 38 zeigt einen konkreten Ausschnitt innerhalb des Ablaufes zur Verdeutlichung des Einsatzes des eMDR.

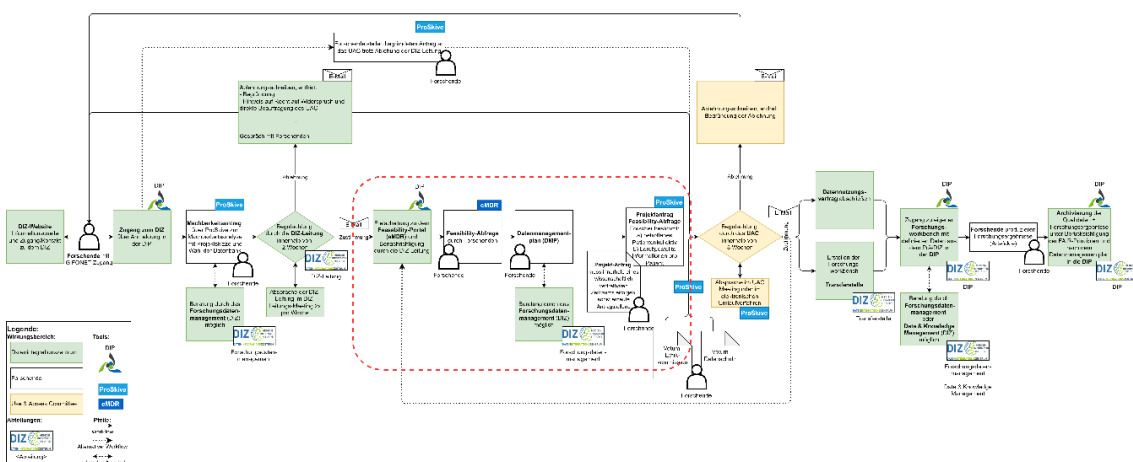


Abbildung 35: Ausschnitt des organisatorischen DIZ-Workflows vor Projektantrag (Eigene Darstellung des DIZ)

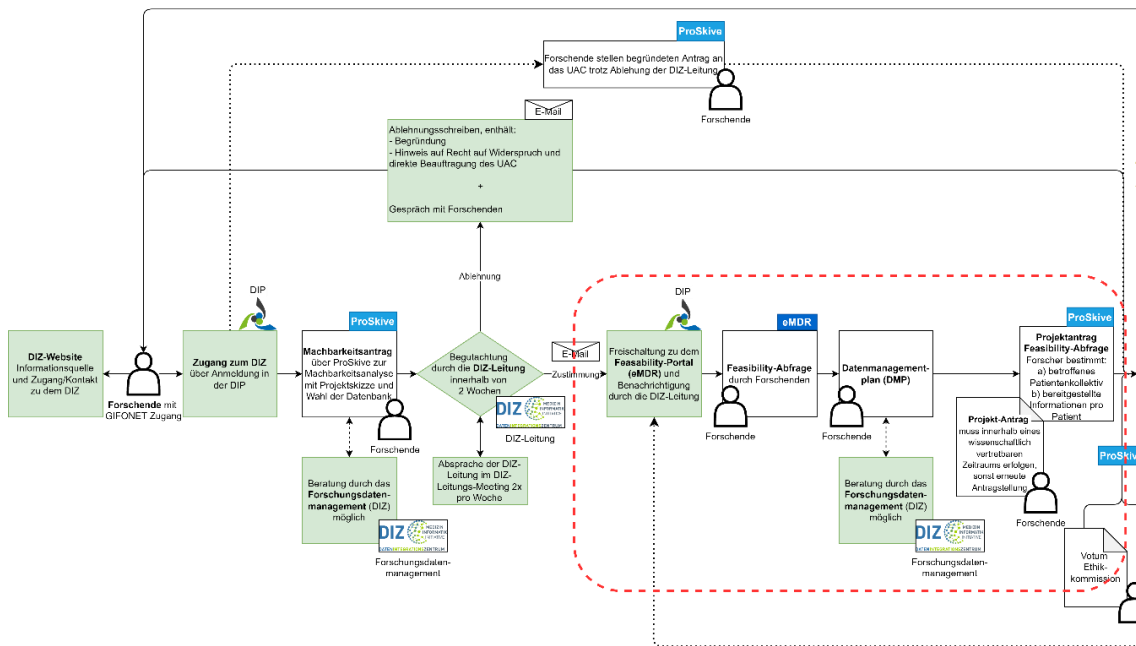


Abbildung 36: Gesamtbild des organisatorischen DIZ-Workflows (Eigene Darstellung des DIZ)

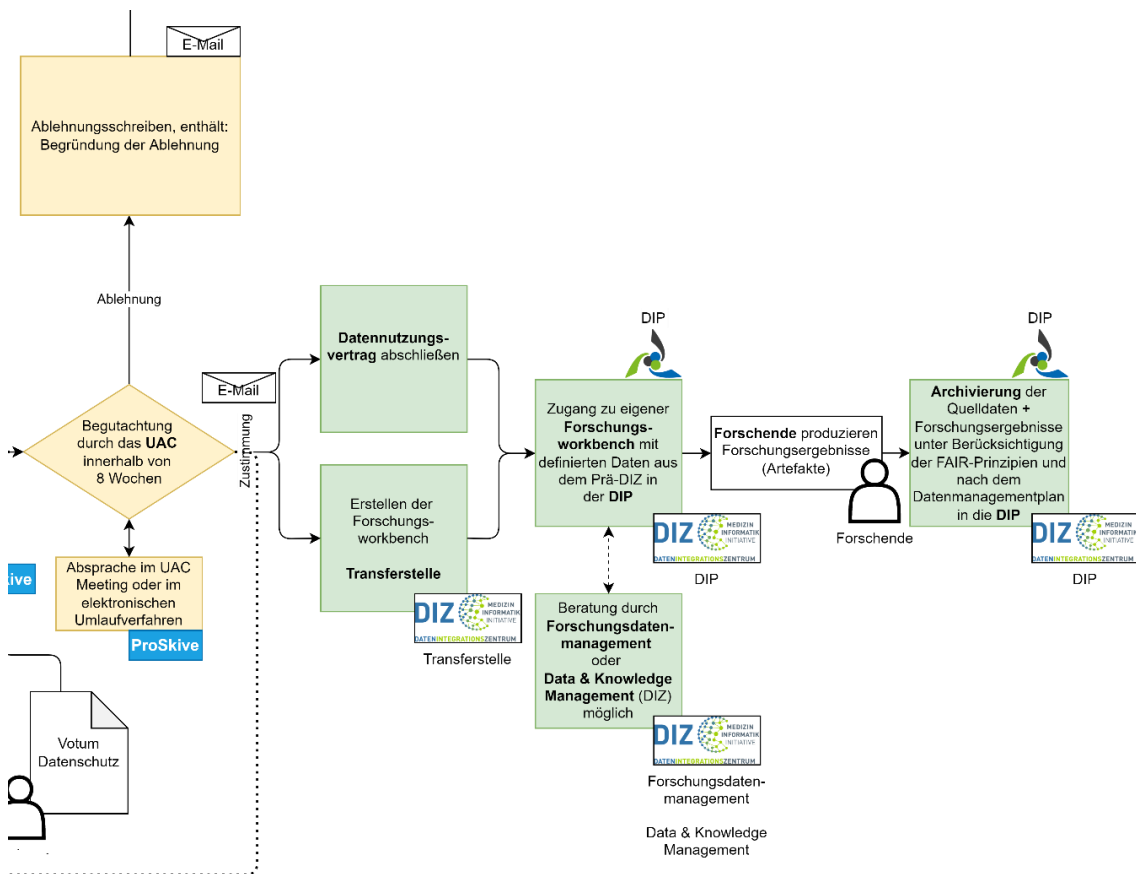


Abbildung 37: Ausschnitt des organisatorischen DIZ-Workflows nach Projektantrag (Eigene Darstellung des DIZ)

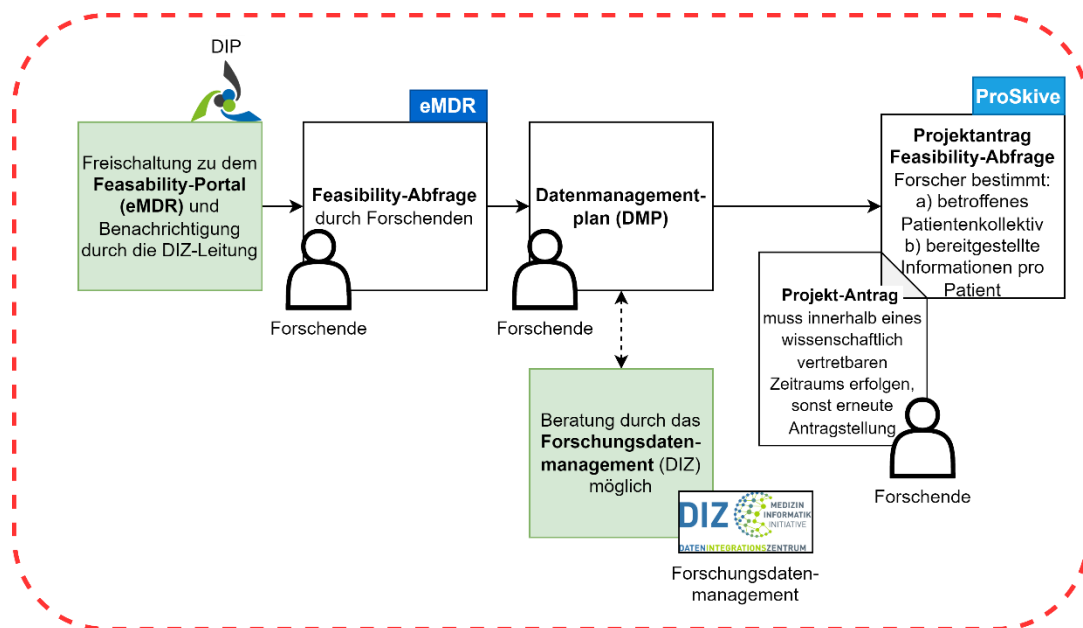


Abbildung 38: eMDR relevanter Ausschnitt des organisatorischen DIZ-Workflows (Eigene Darstellung des DIZ)

Der Ablauf verdeutlicht auch die Integration innerhalb der Datenintegrationsplattform, da der Zugang zu dem eMDR mittels einer Rechtezuweisung innerhalb der DIP vollzogen wird. Ein weiterer wichtiger Bestandteil eines Forschungsvorhabens, welches an das UAC gestellt wird, ist der Datenmanagementplan (DMP). Die Machbarkeitsabfrage und die daraus resultierenden Ergebnisse bilden die Grundlage für diesen DMP. Ein wichtiger Aspekt ist hierbei das Datenvolumen, welches aus der Anzahl an Vorkommen resultiert und eine direkte Auswirkung auf die Langzeitarchivierung im Sinne des benötigten Speichervolumens verdeutlicht.

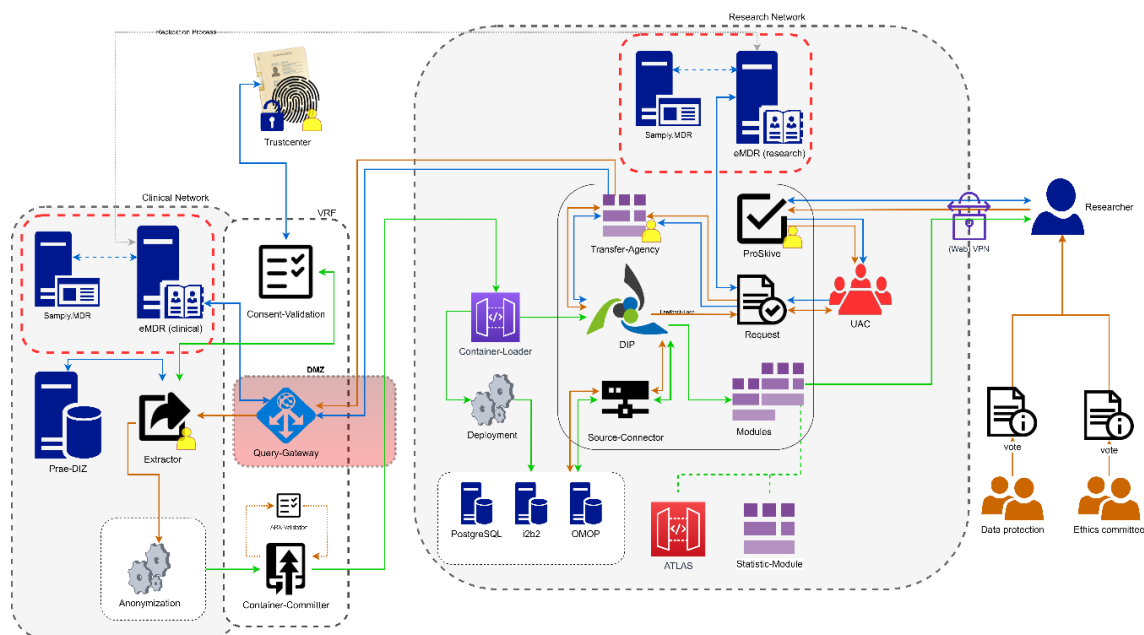


Abbildung 39: Gesamtbild der Systemsicht des DIZ-Workflows (Eigene Darstellung des DIZ)

Die dritte Sicht stellt die Systemsicht innerhalb des DIZ dar und ist in Abbildung 39 ersichtlich.

Die Detailausschnitte sind in Abbildung 40 und Abbildung 41 verdeutlicht und repräsentieren die beiden systematisch getrennten Infrastrukturbereiche.

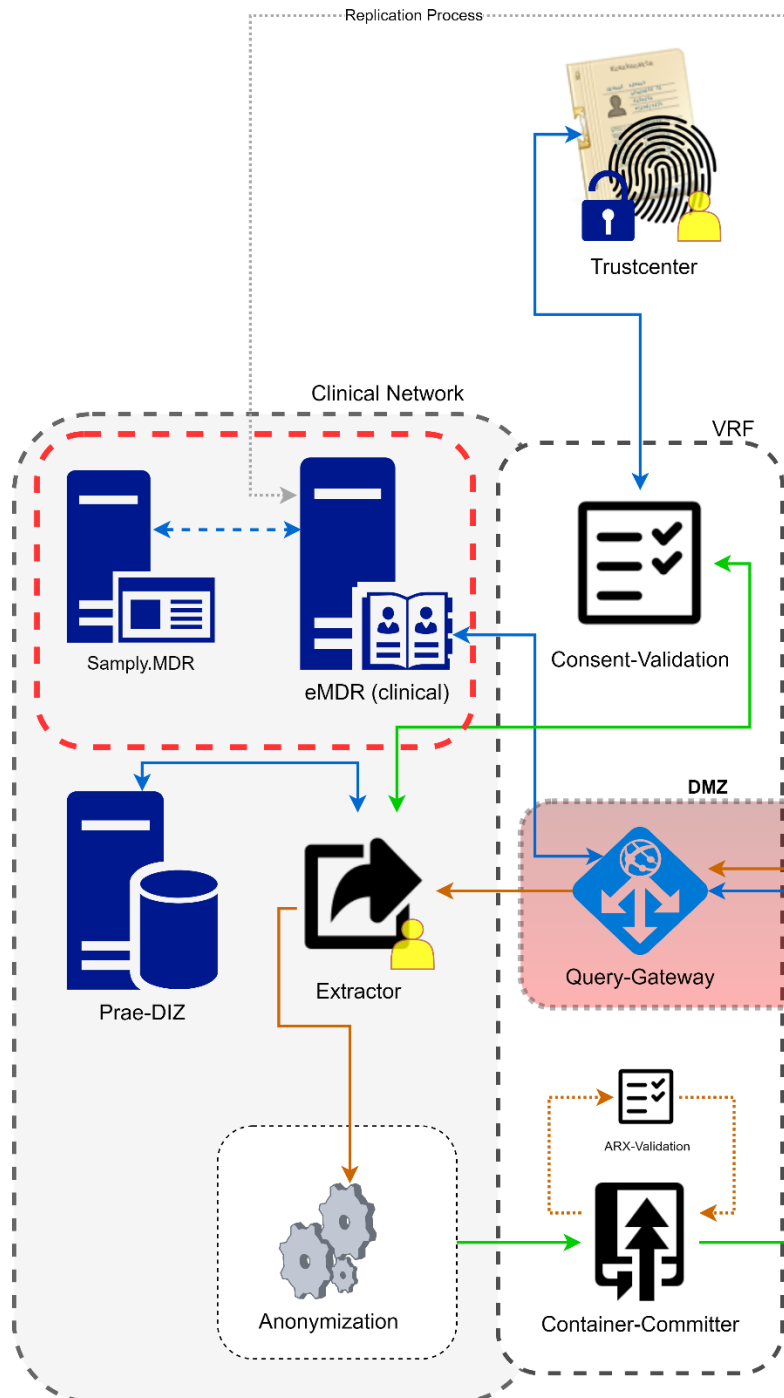


Abbildung 40: Ausschnitt der Systemlandschaft im Routinebereich (Eigene Darstellung des DIZ)

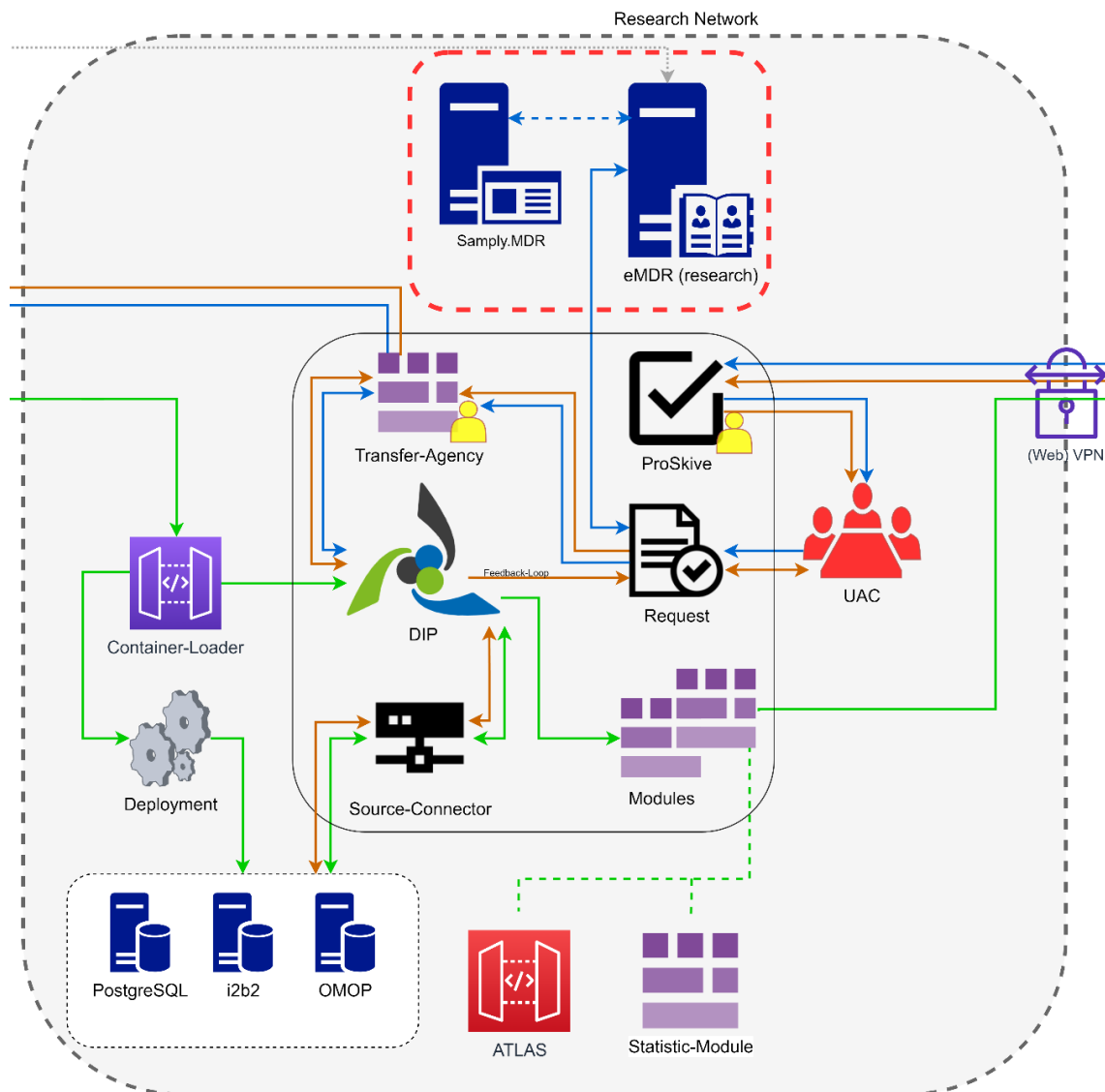


Abbildung 41: Ausschnitt der Systemlandschaft im Forschungsbereich (Eigene Darstellung des DIZ)

Eine große Herausforderung stellt die Homogenität der gesamten Systemlandschaft innerhalb des DIZ dar. Wie in den Abbildungen ersichtlich, gliedert sich das eMDR an zwei zentralen Stellen zwischen diversen, parallel entwickelten und teilweise auch schon vorhandenen Systemen ein. Dies ist nur durch die Etablierung gemeinsamer Schnittstellen und größtenteils zentraler Datenbanken möglich. Weitere Informationen zu der technischen Infrastruktur des DIZ und dessen Verwaltung können Kapitel 13.7 entnommen werden. Die dargestellten Abläufe und Systemsichten, inklusive der Einbringungen des eMDR, wurden im Zuge einer Auditierung, beauftragt durch das Bundesministerium für Bildung und Forschung im Kontext der Medizininformatik-Initiative, im Juni 2021 begutachtet. Das Ergebnis war die Einschätzung des Gießener Datenintegrations-zentrums als „Fortgeschritten“, was im Vergleich zu anderen Datenintegrationszentren im oberen Drittel lag.

4.10 Umsetzung des IT-Grundschutzes

Die für das eMDR relevanten Bausteine des IT-Grundschutz-Kompendiums stellen die Module SYS und APP dar. Hierbei lässt sich das System auch in diese beiden Kategorien, die Infrastruktur (SYS) und die Webapplikation (APP) unterteilen. Nachfolgend werden die essenziellsten Module (Basis-Anforderungen) und deren Status verdeutlicht. Die Kategorienamen entsprechen den genauen Bezeichnungen aus dem IT-Grundschutz-Kompendium (Bundesamt für Sicherheit in der Informationstechnik, 2022, p. SYS.1.6 4-5).

Die Infrastruktur des eMDR basiert auf mehreren Container-basierten unterteilten Infrastrukturen, welche unter dem Baustein SYS.1.6 Containerisierung zusammengefasst werden.

„SYS.1.6.A1 – Planung des Container-Einsatzes“: Sowohl die Skalierbarkeit, als auch die Verfügbarkeit und benötigte Sicherheit wurde separat für die drei Container des eMDR (Datenbank, Backend, Frontend) im Vorfeld innerhalb des DIZ geprüft und für erfüllt befunden. Die Grundlage hierfür bildet ein auditiertes Infrastrukturkonzept, inklusive eines IT-Sicherheitskonzeptes und bestehende Sicherheitsrichtlinien innerhalb des GIFONET.

„SYS.1.6.A2 – Planung der Verwaltung von Containern“: Eine einheitliche und Administrator-unabhängige Verwaltung der Container wird durch unterschiedliche Aspekte als gelöst angesehen. Zum einen erfolgt das Management von IT-Infrastrukturen innerhalb des Datenintegrationszentrums Gießen durch drei Systemadministratoren, wodurch ein personeller Ausfall eingeschränkt werden kann. Weiter wird die Verwaltung der Container durch eine entsprechende Dokumentation unterstützt und kann, mittels zusätzlicher und bereits innerhalb des DIZ eingesetzter Werkzeuge wie *Portainer*, realisiert werden.

„SYS.1.6.A3 – Sicherer Einsatz containerisierter IT-Systeme“: Bezieht sich primär auf die innerhalb der Container betriebenen Anwendungen. Diese sollen aus Sicht der Performance und Kapselung keine Einschränkungen vorweisen, welches durch die initiale containerbasierte Entwicklung der Anwendungen und Nutzung etablierter Images der Hersteller/Communities umgesetzt wurde.

„SYS.1.6.A4 – Planung der Bereitstellung und Verteilung von Images“: Durch die Etablierung einer separaten Image-Registry für das Datenintegrationszentrum können sämtliche Versionen der eMDR-Images abgerufen und verwaltet werden.

SYS.1.6.A5 bis SYS.1.6.A8 beziehen sich noch rein auf die Sicherheit von Images, die Persistenz von Log-Dateien und der Administrationsmöglichkeiten der Systeme. Alle Bausteine wurden durch die Speicherkonfiguration der Container, die Netzwerksegmentierung vom Host-System und der Verwendung von sicheren Passwort-Safes innerhalb des Datenintegrationszentrums umgesetzt und eingehalten.

Die folgenden Standard-Anforderungen SYS.1.6.A9 bis SYS.1.6.A20 wurden entweder im Zuge dieser Dissertation oder durch das allgemeine IT-Sicherheitskonzept des DIZ Gießen mindestens teilweise erfüllt.

Die erweiterten Richtlinien SYS.1.6.A21 – SYS.1.6.A26 konnten aufgrund des Umfangs und des Einflusses auf die gesamte Infrastruktur der Forschungs- und Routineumgebung nicht betrachtet werden.

In Bezug auf die bereitgestellte Web-Anwendung wurden ebenfalls zunächst die Basis-Anforderungen des Grundschutzes betrachtet. (Bundesamt für Sicherheit in der Informationstechnik, 2022, p. APP.3.1 3-5)

APP.3.1.A1, APP.3.1.A7 und APP.3.1.A14 beziehen sich in erster Linie auf den Schutz des Systems und der Daten durch die Umsetzung einer Authentifizierung. Dies wird durch den Einsatz der DIZ-eigenen Keycloak-Instanz erfüllt und somit auch gleichzeitig auf die administrative Schicht des Datenintegrationszentrums ausgelagert. Weiter erfolgt eine Nutzung und die damit einhergegangene Freischaltung eines Forschenden für das System durch das DIZ und nur durch eine im Vorfeld durchgeführte Freigabe des Nutzungsantrages durch das Use- und Access-Committee des Standortes Gießen. Aufgrund der Tatsache, dass dieses Freigabeverfahren bereits digital durchgeführt und erfasst wird, ist eine Dokumentation und Nachvollziehbarkeit gegeben.

Die Standard-Anforderungen beziehen sich im nächsten Schritt zusammengefasst auf die konzeptionelle und feine Trennung der Systemdienste sowie die sichere Kommunikation dieser Komponenten. Dies erfolgt durch die Trennung des Front- und Backends, sowie der Verwendung der Protokolle HTTP und HTTPS.

Abschließend ergibt sich aus der letzten Anforderungskategorie (Erhöhter Schutzbedarf) der Einsatz einer entsprechenden Firewall, welche wie auch bei den erweiterten Richtlinien der SYS.1.6 Bausteine, aufgrund der Auswirkungen und der bereits vorhandenen Firewalls, nicht betrachtet wurde.

Die Entwicklung des eMDR wurde stets unter Berücksichtigung aktueller sicherheitsrelevanter Ereignisse und aktualisierter Vorgaben zur IT-Sicherheit durchgeführt.

5. Diskussion

Durch die weiterhin fortschreitende Digitalisierung und das mittels neuer Verfahren einhergehende Wachstum an Datenmengen, werden diverse Bereiche verdeutlicht, welche besondere Betrachtung benötigen. Zunächst ist jedoch eine grundlegende Abgrenzung der medizinischen Daten in Forschungsdaten und Daten aus der klinischen Routine nötig, da diese im Alltag unterschiedlich betrachtet und verwaltet werden. Hierbei erfolgt ein wichtiger Blick auf die FAIR-Kriterien, welche die Prinzipien *Findable*, *Accessible*, *Interoperable* und *Reusable* behandeln und für das zeitgemäße Datenmanagement im medizinischen Bereich unerlässlich sind. (Jansen, 2016) *Findable* bezieht sich in erster Linie auf die Verwendung von Metadaten, um auf Basis abstrahierter Bezeichner, sämtliche relevanten und verknüpften Daten innerhalb einer Datenlandschaft identifizieren zu können. Das Prinzip *Accessible* verdeutlicht die Notwendigkeit nach Möglichkeiten Daten auch nach bestimmten Zeiträumen, beispielsweise nach der Laufzeit von Forschungsprojekten, weiterhin aufzurufen und verwenden zu können. *Interoperable* stellt die Verwendung von Standards und allgemein verfügbaren und bekannten Datenrepräsentationen dar, wodurch die Daten auch über unterschiedliche Systeme hinweg weiterhin interpretierbar und verwendbar sind. Das letzte Prinzip *Reusable* beschreibt die Notwendigkeit nach open-source Systemen und öffentlich verwendbare System- und Datenlizenzen, um beispielsweise eine Nutzung für andere Forscher zu ermöglichen und somit die weltweite Forschung zu unterstützen. (Wilkinson et al., 2016)

Die lokale Datenhaltung stellt einen großen Bereich dar, welcher aus Sicht der FAIR-Prinzipien einen direkten und derzeit meist negativen Einfluss auf die gewünschte Datensituation hat. Der derzeitige Stand im Forschungsbereich sieht eine Haltung von Forschungsdaten primär lokal auf dienstlichen Endgeräten und externen Datenträgern (CD, USB-Stick, externe Festplatte) vor, wie auch am Standort Gießen eine 2016 durchgeführte Umfrage verdeutlicht. (Waldschmidt-Dietz and Krippes, 2016, p. 9) Diese Analyse deckt sich auch mit dem internationalen Stand der Erfassung, Haltung und Verwaltung von Forschungs- sowie Studiendaten und lässt sich auf fehlende Vorgaben und Standards zurückführen. (MIT Critical Data et al., 2016, pp. 143–144) (Zozus et al., 2019, pp. 214–215) Dies stellt einen konkreten und ersichtlichen Bruch in der Anwendung der FAIR-Prinzipien dar und kann sich beispielsweise negativ auf Kooperationen und Verbundprojekte auswirken.

In der klinischen Routine herrscht oft eine inhomogene IT-Landschaft, die sich negativ auf die Erfüllung der FAIR-Prinzipien auswirken kann. Die Definitionen der Metadaten

sind teilweise herstellereigen und es fehlen oft Schnittstellen zu anderen Systemen. Dadurch können die FAIR-Prinzipien nicht vollständig umgesetzt werden. (Heckmann, 2022, pp. 308–309) (Boeckhout et al., 2018, p. 932) Dies kann zu Problemen führen, wenn es darum geht, Daten zwischen verschiedenen Systemen auszutauschen oder die Daten für die Analyse und Auswertung zu nutzen. Es ist daher wichtig, dass in der klinischen Routineumgebung die Verwendung von standardisierten Metadaten und Schnittstellen zwischen den Systemen gefördert wird. Dadurch können die Daten interoperabel gemacht werden und eine effektive Nutzung der Daten wird ermöglicht.

Durch die Medizininformatik-Initiative und der daraus resultierenden Konsortien wie MIRACUM werden diese Probleme konkret analysiert und bei den relevanten beteiligten Akteuren adressiert. Die Gründung und Etablierung eines Datenintegrationszentrums ist der initiale Schritt für eine zentrale Anlaufstelle für Forschende, jedoch nicht nur direkt in Bezug auf den Abruf von Daten, sondern primär in Bezug auf die Beratung für ein gutes, zeitgemäßes und FAIR-konformes Datenmanagement. Dies führt im Umkehrschluss zur Schaffung einer Grundlage, welche eine zukünftige Datenzusammenführung und Nutzung stark erleichtert und überhaupt erst ermöglicht. (Hampf et al., 2021) (Eberhardt et al., 2021)

Ein davon zu trennendes und technisches Ziel des DIZ ist die Zusammenführung von Forschungsdaten mit Daten aus der klinischen Routine. Dies ermöglicht gesamtheitliche Anpassungen, Standardisierungen und semantischen Angleichungen der großen, meist inhomogenen Datenlandschaften. Eine wichtige Komponente stellt hierbei die initiale Analyse der Datenbestände dar, welche am Standort Gießen einen großen Einfluss auf diese Dissertation verübte. Die Differenzen in Struktur, Speicherort, Datenmenge und Verfügbarkeit stellen die Grundlage für die Idee eines zentralen Informationsgebers in Form eines Metadatenrepositoriums dar.

Die Zusammenführung und speziell die Nutzung von Daten muss allerdings, insbesondere aus rechtlicher und ethischer Sicht, im Zuge jeder einzelnen Auswertung geprüft werden und auf rechtlichen Grundlagen fundieren. Zu diesem Zweck wurden im Rahmen des MIRACUM-Projektes und primär im Zuge dieser Dissertation drei Ordnungen generiert, welche die rechtlichen Rahmenbedingungen definieren und konkret auf die Pflichten der Forschenden bei dem Umgang mit extrahierten Daten hinweist. Aufgrund des gemeinschaftlichen Betriebes des DIZ von Seiten des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des Universitätsklinikums Gießen und Marburg GmbH (Standort Gießen) erfolgte die

Absprache und Freigabe der Ordnungen stets unter Einbeziehung des Dekanats und der Geschäftsführung.

Ein weiterer Aspekt stellt die Ethik und der Datenschutz bei Auswertungen dar, welche auf der Basis des Datenintegrationszentrums durchgeführt werden. Da diese Aspekte sich für jedes Vorhaben unterscheiden und insbesondere in ihrem Umfang stark variieren können, wurden diese notwendigen Punkte ebenfalls fest in den Ordnungen hinterlegt und müssen zwingend erfüllt werden. Jedes Forschungsprojekt und -Vorhaben durchläuft ein Antragsverfahren, welches bereits identifiziert, welche Daten in welchem Umfang und mit welchen Werkzeugen verarbeitet werden sollen. Zu diesem technisch dokumentierten Vorhaben werden ein entsprechendes Ethikvotum und eine Stellungnahme des Datenschutzes benötigt. Im Anschluss folgen eine Begutachtung und Bewertung des gesamten Vorhabens durch das lokale Use- und Access-Committee, wodurch eine Ablehnung oder Freigabe eines wissenschaftlichen und rechtlich sowie ethisch erfahrenen Personenkreises erfolgt. Dieses Antrags- und Abstimmungsverfahren kann zentralisiert und einheitlich für den Fachbereich Medizin genutzt werden, ohne dass einzelne Institutionen und Zentren selbstständig Vorhaben prüfen müssen, wodurch Dokumentations- und gegebenenfalls Entscheidungsfehler vermieden werden können.

Ein Metadatenrepositorium mit einer Abfrage- und Visualisierungsfunktion soll dieses Antragsverfahren in unterschiedlichen Bereichen unterstützen. Zunächst können Forschende ein konkretes Kollektiv an Patienten auf einer standortweiten Datenbasis identifizieren, wodurch im Nachgang auf Grundlage des Abfrageergebnisses eine Prüfung der Ethikkommission und des Datenschutzes erfolgen kann. Somit werden auch diese Entscheidungsabläufe technisch und dokumentativ unterstützt.

Aufgrund der Problematik von Insellösungen in der heterogenen IT-Systemlandschaft wurde eine Analyse bereits bestehender und etablierter MDR-Lösungen durchgeführt woraus, auf Basis des Funktionsumfangs, drei bestehende Lösungen in die finale Entscheidung einbezogen wurden. Als open-source Lösungen wurden die im wissenschaftlich-medizinischen Bereich angesiedelten Systeme i2b2 und Samply.MDR betrachtet, wohingegen aus kommerzieller Sicht das integrierte MDR der klinischen Archivlösung Synedra.AIM analysiert wurde. Letzteres überstieg aus Sicht des Konfigurationsaufwandes, der fehlenden Flexibilität und der benötigten Investitionen in Anpassungen durch den Hersteller, den Rahmen des Vorhabens und musste somit, trotz der bereits existenten Installation innerhalb des klinischen Umfeldes, ausgeschlossen werden.

Die Software i2b2 stellt ein umfangreiches Werkzeug von einem Data-Warehouse bis zur Datenexploration und Analyse dar und wird im Kontext von MIRACUM als ein Kernbestandteil des IT-Portfolios des Projektes, die sogenannten MIRACOLIX-Tools, angesehen. (The i2b2 Community et al., 2023) (MIRACUM, 2022) Auf der Basis einer 2018 durchgeführten Erfassung des technischen Ist-Zustandes am Standort Gießen und der konkreten Planung der Entwicklung des Datenintegrationszentrums, stellte i2b2 einige Probleme mit seiner Verwendung in den Vordergrund. Primär konnte aufgrund fehlender Expertise in Bezug auf i2b2 keine technische Detail- und Erweiterungsplanung erfolgen. Eine Schnittstellenentwicklung zu diesem und anderen Systemen konnte in ihrem Aufwand nicht abgeschätzt werden und sorgte für ein einheitliches Votum, durch die DIZ-Leitung und der Vertreter der Betreiber des DIZ, gegen die Software zur Nutzung als Metadatenrepositorium. Unabhängig davon wird die Software im Zuge der Etablierung des DIZ eigenständig betrieben, mit dem Anspruch eine Expertise aufzubauen und das System produktiver nutzen zu können. Ein weiterer Aspekt bezog sich auf die Benutzerfreundlichkeit und benötigte Einarbeitungszeit in die Anwendung, welche im Zuge initialer und unabhängiger Tests durch das Team des DIZ, zunächst als nicht positiv gewertet wurde. Aufgrund des Umfangs und des Detailgrades der i2b2 Auswertungen steht die Durchführung einer simplen und reinen Machbarkeitsabfrage in keiner guten Relation zur benötigten aufzuwendenden Zeit. Aus diesem Grunde existierten erste Ideen ein separates System vorzuschalten, welches die initiale Machbarkeit bestätigt und der Software i2b2 im Nachgang die Detailanalysen überlässt.

Die dritte Option stellte Samplify.MDR dar, welches ebenfalls ein primäres Werkzeug der MIRACOLIX-Tools darstellt. Diese reine MDR-Lösung erfüllte aufgrund der bereitgestellten Vielzahl an Schnittstellen und einer ausführlichen Dokumentation sämtliche Kriterien, um eine Administration und Betreuung des Systems zu ermöglichen. Die erfassten Nachteile bildeten sich zunächst aus fehlenden Auswertungsmöglichkeiten und dem Fokus des Systems auf die Abbildung und Unterstützung von semantischen und keinen inhaltlichen Informationen, wie die Anzahl an Vorkommen. Dieser Aspekt wurde im Zuge der Planung der DIZ-Infrastruktur aufgenommen und es ergab sich aus dem Technologievergleich und den benötigten Einzelkomponenten ein konkretes Bild der Systemlandschaft. Dieses verdeutlichte, dass i2b2 einen Einsatz im Bereich von detaillierten Analysen und Samplify.MDR in der Verwaltung semantischer Informationen am DIZ dienen soll, jedoch fehlte der konkrete Aspekt einer initialen Machbarkeitsabfrage, welche innerhalb des neu zu etablierenden Arbeitsablaufes voll integriert ist. Das entwickelte eMDR gliedert sich in diese Systemlücke ein und stellt somit zwar eine Eigenentwicklung dar, nutzt jedoch das bereits vorhandene System

Samply.MDR, stellt Schnittstellen für andere Systeme bereit und kombiniert die inhaltlichen und semantischen Informationen zu Gunsten der Forschenden.

Das eMDR wurde zunächst vollständig auf Basis des Python Flask-Frameworks entwickelt, wobei das Framework sowohl für das Backend, als auch für das Frontend mittels bereitgestellter Render-Funktionen verwendet wurde. Aufgrund der technologischen Entwicklungen und best-practice Ansätzen wurde die Entwicklung der beiden Komponenten strikt getrennt und eine interne Kommunikation wurde mittels API-Schnittstellen realisiert. Das Backend ist somit deutlich leichtgewichtiger und übernimmt die Kommunikation mit der verwendeten PostgreSQL-Datenbank. Das verwendete Datenbank-Schema wurde auf Grundlage der lokal benötigten Daten zur Machbarkeitsanalyse und geplanter Erweiterungen generiert und deckt derzeit sämtliche Informationen ab, welche im Zuge der aktuellen Entwicklungsphase aus Sicht des Datenschutzes verwendet werden können. Diese stellen derzeit die Datenlieferungen des Paragraf 21 Datensatzes nach KHEntgG (Krankenhausentgeltgesetz, siehe (Institut für das Entgeltsystem im Krankenhaus GmbH, 2022)) durch das Controlling dar und können in ihrer ursprünglichen Form (CSV) direkt in das eMDR integriert werden. Das Backend selbst übernimmt nicht nur die Kommunikation mit dem semantischen Part, repräsentiert durch das Samply.MDR, sondern bietet auch eigene Schnittstellen zur Abfrage der Vorkommen von Codes innerhalb der Datenbasis. Das Vue.js Frontend übernimmt hierbei die reine Informationsdarstellung und erfasst durchzuführende Abfragen mittels Formularen, welche an das Backend übermittelt und die Ergebnisse aus der Rückmeldung visualisiert. Die Vorteile dieser Lösung beziehen sich nicht nur auf die Tatsache, dass eine Lücke in der komplexen IT-Systemlandschaft geschlossen wird, sondern das System ist hochgradig durch Umgebungsvariablen konfigurierbar und erweiterbar. Weiter stellt das eMDR ein hauseigenes Produkt des DIZ und somit des Fachbereichs dar, wodurch im Falle dringend benötigter Anpassung oder Handlungsbedarf bei Sicherheitsvorfällen eine direkte Aktion erfolgen kann und keine Abhängigkeit von Dritten vorhanden ist. Im Zuge der kontinuierlichen Verbesserung der DIZ-Systeme und Dienste ist das eMDR ebenfalls in der Liste der Kernsysteme geführt sowie integriert und wird, basierend auf Nutzerfeedback, weiterentwickelt.

Die initiale Entwicklung der Komponenten bezog sich streng auf die durch den IT-Grundschutz definierten Bausteine. Jedes neue System in einer technischen Umgebung stellt automatisch einen neuen Angriffspunkt und somit eine Schwachstelle für die Gesamtstruktur dar, weshalb es unerlässlich ist, bereits die Entwicklung rein auf die IT-Sicherheit ausulegen. Die umgesetzten Bausteine bilden aus den genannten Aspekten ein solides Fundament für eine sichere Inbetriebnahme, sollte jedoch kontinuierlich

geprüft und bei Bedarf an neue Vorgaben angepasst werden. Diese Anpassung erfolgt im Zuge eines auditierten Sicherheitskonzeptes des Datenintegrationszentrums durch den Bereich der Systemadministration und ist somit auch dort in die Wartungs- und Prüfabläufe voll integriert.

Neben den Sicherheitsaspekten sind die wissenschaftlichen Rahmenbedingungen, die FAIR-Prinzipien, unerlässlich und konnten durch die Struktur des eMDR vollständig erfüllt werden. Die Daten innerhalb des Systems sind vollständig Metadaten-zentriert und können exploriert werden, wodurch die Komponente *Findable* erfüllt ist. Sämtliche integrierten Daten verbleiben bis zu einem Patientenwiderruf innerhalb des Systems, unabhängig der Zeiträume des Herkunftsprojektes oder des Alters der Daten. Ein entsprechender Widerruf wird vollständig unterstützt, benötigt jedoch ein bereits existentes Consent-Management, welches die, durch das eMDR bereitgestellte, Schnittstelle nutzen kann. Durch die kontinuierliche Bereitstellung der Daten wird der Aspekt *Accessible* umgesetzt. Das nächste FAIR-Prinzip *Interoperable* erfüllt das eMDR durch die Auslagerung der Semantik an das Smply.MDR, welches das Kriterium erfüllt und die Verwendung von strukturierten und etablierten JSON-Strukturen für den internen Datenaustausch ermöglicht. Eine solche Trennung sorgt im Kontext der ganzheitlichen Systemlandschaft für eine getrennte Betrachtung der beiden Informationen (Vorkommen und Bedeutung) und ermöglicht es aus Sicht der Interoperabilität beide Aspekte separat zu betrachten, anzupassen und als Informationsquelle nutzen zu können. (Kush and Nordo, 2019, p. 381) Das letzte Kriterium *Reusable* kann aus technischer Sicht ermöglicht werden, da das quelloffene System aufgrund seiner Anpassbarkeit auch durch Dritte installiert und verwendet werden kann. Aus Sicht der Daten erfolgt eine Erfüllung des Kriteriums durch die Daten- Zugriffs- und Nutzungsordnung, welche die Regeln zur Datennutzung, Bereitstellung und Veröffentlichung am Standort Gießen vorgibt.

Zusätzliche Funktionen stellen die zentrale Authentifizierung mittels Keycloak und die Bereitstellung einer FHIR-Schnittstelle dar. Beide Komponenten sorgen für eine vollständige Integration in eine bestehende Systemlandschaft. Keycloak erleichtert mit seiner Anbindung an das lokale LDAP des GIFONET die Zugangsverwaltung und bietet Forschenden somit einen zentralen Login für sämtliche Dienste. Dies reduziert den administrativen Aufwand, minimiert Sicherheitsrisiken und beschleunigt die Arbeiten innerhalb der DIZ-Dienste durch eine hintergründige Token-Authentifizierung, wodurch die erneute Eingabe von Login-Daten vermieden wird. Durch dieses Verfahren wird auch die volle Integration des eMDR in die Datenintegrationsplattform ermöglicht, wodurch die auditierten Grundsätze, Wartungs- sowie Sicherheitskonzepte und Betriebsabläufe des

DIZ eingehalten werden. Die FHIR-Schnittstelle hingegen ermöglicht den standardisierten und strukturierten Austausch der medizinischen Daten zwischen den Systemen. Diese sorgt für eine grundsätzliche Interpretation von FHIR-Nachrichten, wodurch entsprechende Objekte in das eMDR integriert oder auch exportiert werden können.

Intern verfolgt das DIZ streng die nachvollziehbare und dokumentierbare Authentizität und Plausibilität sämtlicher Arbeitsabläufe. Zu diesem Zwecke erfolgt ausnahmslos die vollständige Dokumentation der innerhalb des eMDR durchgeführter Abfragen und hält diese primär für die interne Daten-Transferstelle bereit. Die Historie dient mit ihren Inhalten nicht nur zur Vereinfachung der Arbeitsabläufe durch Vorbereitung der internen Extraktions-IDs, sondern bietet auch eine Sicherheit gegen Dokumentationslücken und Falschdokumentationen im Zuge der Antragsstellung zur Datennutzung an das UAC. Weiter erfolgt eine Protokollierung jeglicher in das System integrierter und im Kontext der FHIR-Schnittstelle auch exportierter Daten. Die Synchronisation beziehungsweise Replikation der Daten über die unterschiedlichen Infrastrukturen und Installationen hinweg wird zusätzlich separat, durch die Software Airbyte, protokolliert und kann auf dieser Basis analysiert werden.

Im Zuge kontinuierlicher und paralleler Entwicklungen wurden stets alternative und neue Ansätze aus Sicht des Technologie-Stacks für das eMDR geprüft und teilweise in einer initialen Version entwickelt. Der technologische Fortschritt brachte im Laufe der Entwicklung des eMDR das Python FastAPI-Framework hervor, welches ebenfalls auf die Bereitstellung von Web-Services, jedoch mit primärem Fokus auf Performance, ausgelegt ist. (Ramírez, 2022) Eine Übertragung des Backend auf ein neues Framework stellt keine Probleme dar, muss jedoch im Vorfeld auf Sinnhaftigkeit geprüft werden. Ein weiterer betroffener Bereich wird durch die Datenbank repräsentiert, welche derzeit relational strukturiert ist und mit einem objektorientierten NoSQL-Ansatz verglichen werden muss. Insbesondere in den letzten Jahren stellte sich MongoDB als schnelle, skalierbare, zuverlässige und alternative Lösung zu bestehenden großen Datenbanken dar und muss aus diesem Grund zwingend in Betracht gezogen werden, insbesondere auch aufgrund der wachsenden Datenmengen innerhalb des eMDR. (Anusha et al., 2021, p. 1833) Analysen zu den Unterschieden zwischen PostgreSQL und MongoDB laufen in unterschiedlichen wissenschaftlichen Bereichen bereits seit Jahren, weshalb diese Entwicklung weiterhin betrachtet und auf das eMDR angewandt werden muss. Derzeit existiert parallel zur aktuellen Entwicklung eine Daten-identische MongoDB Testinstanz, welche FHIR-Nachrichten direkt als Objekte integriert und als Grundlage für kommende Benchmarking-Analysen des DIZ verwendet wird.

Ursprünglich sah die Konzeptphase des eMDR auch die Anwendung, Nutzung und Exploration von Ontologien auf Basis der Metadaten vor. Dieses Vorhaben wurde im Einvernehmen mit der Leitung des Datenintegrationszentrums, unter Berücksichtigung der strukturellen Trennung der Informationen und der Semantik, auf einen späteren Projektzeitraum verschoben. Gründe hierfür stellen konkret die Nutzung von Samplify.MDR dar, welches für die nächsten Entwicklungsschwerpunkte den Aspekt der Ontologien andenkt und ohnehin für den Bereich der Semantik verwendet wird. Die Verwendung und Einbindung von Ontologien stellen aus Sicht des DIZ, der Weiterentwicklung des eMDR und der medizinischen Datenverarbeitung im Kontext der Wissensgenerierung, eine essenzielle Notwendigkeit dar. (Fung and Bodenreider, 2019, p. 317)

6. Schlussfolgerung

Das entwickelte erweiterte Metadatenrepositorium eMDR ist basierend auf dem aktuellen Entwicklungsstand vollständig in die technischen und organisatorischen Betriebsabläufe des Datenintegrationszentrums integriert und wurde im Zuge des MIRACUM-Projektes in diesem Gesamtbild auch auditiert. Das derzeit verwendete relationale Datenbankschema bietet bereits diverse Erweiterungsmöglichkeiten zur Abbildung der Patientenaufenthalte, im Kontext zeitlicher und örtlicher Informationen innerhalb des Klinikums sowie die metadaten-zentrierte Organisation von Vitaldaten.

Die nächsten Schritte der Entwicklung werden durch die Anpassung der FHIR-Schnittstelle repräsentiert, welche derzeit eine interne Nachrichten-Validierung durchführt, diese jedoch nicht mit dem innerhalb des DIZ verwendeten Schema übereinstimmt. Diese neue Validierung muss auf Basis der HAPI-FHIR Instanz des DIZ erfolgen und bietet durch eine Anpassung die vollständige und systemweite Kompatibilität des eMDR mit klinischen Systemen.

Im Zuge des technologischen Fortschrittes muss eine Überprüfung der verwendeten Komponenten in Form von Benchmarks erfolgen. Eine erste Gegenüberstellung muss im Bereich des Backend zwischen den Frameworks Flask und FastAPI in den Kategorien Verfügbarkeit und Performance erfolgen. Hierbei gilt es eine hohe Nutzungs- und somit Systemauslastung zu simulieren und das Verhalten beider Frameworks im Detail zu betrachten. Im nächsten Schritt muss das verwendete Datenbankmanagementsystem PostgreSQL dem Produkt MongoDB gegenübergestellt werden. Aufgrund des objektorientierten NoSQL-Konzeptes von MongoDB, mit zusätzlichen Indexier-

Funktionen für große Datenmengen, stellt das System einen komplett entgegengesetzten Ansatz dar. Das Benchmarking muss hierbei auf die Performance und die Abfragen-Geschwindigkeit, basierend auf der Anzahl integrierter Daten innerhalb des eMDR, erfolgen.

Die bisher bereitgestellten Systemfunktionen bilden den benötigten Funktionsumfang im Kontext des Datenintegrationszentrums, inklusive der internen Daten-Transferstelle und des allgemeinen Betriebskonzeptes, aus technischer Sicht vollständig ab. Der jedoch wichtigste durchzuführende Aspekt stellt eine Usability-Studie dar, welche im Zuge der aktuellen Ausbaustufe des Datenintegrationszentrums zeitnah erfolgen kann. Die Studie soll die simple Durchführbarkeit von Machbarkeitsstudien und die Zeitersparnis bei Forschenden im Vergleich zu den herkömmlichen Methoden bestätigen. Ein weiteres Ergebnis der Studie soll auch Erweiterungsmöglichkeiten verdeutlichen und gewünschte Funktionen der Forschenden erfassen, um das benutzerorientierte Datenintegrationszentrum in seinen Diensten kontinuierlich verbessern zu können.

Zwei dringend benötigte und noch offene Punkte werden zum einen durch das Consent-Management am Standort Gießen und das allgemeine Datenschutzkonzept des DIZ repräsentiert. Das Consent-Management benötigt den Aufbau und die Anbindung einer entsprechenden Treuhandstelle, welches es im Zuge des MIRACUM-Projektes und des Aufbaus des Datenintegrationszentrums ebenfalls umzusetzen gilt. Die Etablierung der Treuhandstelle wurde an eine externe Institution ausgelagert, wodurch ausschließlich eine Anbindung erforderlich ist, welche bis spätestens Quartal 4 2023 erfolgen soll. Durch diese Anbindung kann das pseudonymisierte Datenmanagement und die damit verbundene Nutzung der Widerruffunktion des eMDR erfolgen.

Aufgrund der großen zusammengeführten Datenmengen wird ein ausführliches Datenschutzkonzept benötigt, welches im Zuge der DIZ-Entwicklung erstellt wurde und das eMDR vollumfänglich einschließt (Anhang 13.8). Das Konzept befindet sich derzeit noch in der Prüfung durch die lokalen Datenschutzparteien und bildet mit seiner Freigabe den Startschuss für die produktive Nutzung des eMDR in allen drei Bereichen der Umgebungen klinische Routine, Forschungsnetz und Extern (ebenfalls Forschungsnetz). Bevor diese Freigabe erfolgt, verfügt das System bereits über den Datenbestand der Paragraf 21 KHEntgG Datensätze, welche im Zuge des MIRACUM-Projektes eine Freigabe zur Nutzung am Standort erhalten haben.

Durch die Unterstützung des IT-Grundschutzes des BSI und der FAIR-Prinzipien, sowie die Verwendung eines aktuellen und etablierten Technologie-Stacks, stellt das eMDR eine fortschrittliche, sichere und wissenschaftlich nutzbare Lösung eines Metadaten-

repositoriums und ein damit einhergehendes Machbarkeitsportal dar, welches im Zuge des durch das BMBF beauftragten Audits bestätigt wurde. Weiter schließt die Lösung existente Lücken in der standortspezifischen IT-Systemlandschaft und fungiert als Bindeglied zwischen Informationsnehmern, sowohl technisch, als auch personell. Die erstellten organisatorisch und rechtlich relevanten Ordnungen bilden die Grundlage für die weitere Entwicklung und Zukunft der Medizininformatik und der klinischen Forschung am Fachbereich Medizin der Justus-Liebig-Universität Gießen.

Zur Gewährleistung der Weiterentwicklung und Wartung des Gesamtsystems wurde innerhalb des DIZ eine interne Fokusgruppe gegründet, welche sich aus den Teilbereichen der Softwareentwicklung, Systemadministration und Datenintegrationsplattform zusammensetzt. Nach zentraler Freigabe des Datenschutzkonzeptes erfolgt eine Integration der Daten durch die interne Transferstelle des Datenintegrationszentrums.

7. Zusammenfassung

Einleitung

Die Medizininformatik-Initiative in Deutschland verdeutlicht den Bedarf an offenen und standardisierten Datenstrukturen sowie die Zusammenführung von Datenbereichen aus der medizinischen Forschung und der klinischen Routine. Im Zuge dessen sollen Datenintegrationszentren (DIZ) gegründet werden, welche diese Aufgabe an den jeweiligen Standorten übernehmen. Eine Grundlage des Projektes sind Metadaten, die für einheitliche und standardisierte Analysen und Machbarkeitsabfragen genutzt werden können. Hierbei gilt es jedoch rechtliche und organisatorische Grundsätze zu beachten und für den Standort zu etablieren. Darüber hinaus wird ein Metadaten-Repository benötigt, welches die Grundlage für Machbarkeitsabfragen und -Analysen bildet.

Material und Methoden

Für die Implementierung des Metadaten-Repositorys wurde ein Python-Backend mit dem Flask-Framework entwickelt, das auf einer PostgreSQL-Datenbank aufsetzt. Das Frontend wird durch Vue.js mit dem Material Design Framework Vuetify dargestellt. Darüber hinaus nutzt das entwickelte System das Metadaten-Repository Smply.MDR und wurde containerbasiert nach Richtlinien des BSI-IT-Grundschutzes aufgebaut.

Ergebnisse

Zum Aufbau des DIZ und als Basis für das Repository wurden für den Standort Gießen drei verschiedene rechtlich-organisatorische Ordnungen entwickelt. Das daraus implementierte erweiterte Metadaten-Repository (eMDR) bietet die Möglichkeit erste Machbarkeitsabfragen mit Visualisierung, auf Basis eines definierten Kerndatensatzes, durchzuführen. Darüber hinaus erfolgt eine Authentifizierung über eine zentrale Keycloak-Instanz sowie eine containerbasierte Bereitstellung der Komponenten. Das System ist technisch und organisatorisch in alle Arbeitsabläufe des DIZ integriert und erfüllt die Anforderungen des BSI-IT-Grundschutzes. Zudem ist das System, durch eine Synchronisationsfunktion, in allen infrastrukturell getrennten Systembereichen installiert.

Diskussion

Die etablierten Strukturen bieten eine ideale Grundlage für die weitere Forschung am Standort Gießen im Rahmen des DIZ und fungieren, insbesondere aus Sicht aus eMDR, als Bindeglied, in der Rolle eines Informationsgebers, in der inhomogenen Systemlandschaft. In Zukunft muss das System Benchmarking-Verfahren und Usability-Studien bestehen, um auf Basis der Ergebnisse erweitert und angepasst zu werden.

Introduction

The medical informatics initiative in Germany highlights the need for open and standardized data structures, as well as the consolidation of data areas from medical research and clinical routine. In the course of this, data integration centers (DIC) will be founded, which will take over this task at the respective locations. One basis of the project is metadata, which can be used for uniform and standardized analyses and feasibility queries. However, legal and organizational principles must be observed and established for the site. Furthermore, a metadata repository is required, which forms the fundament for feasibility queries.

Material and Methods

To implement the metadata repository, a Python backend was developed using the Flask framework, which is layered on top of a PostgreSQL database. The frontend is represented by Vue.js with the material design framework Vuetify. Furthermore, the developed system uses the metadata repository Sampil.MDR and was constructed container-based according to the BSI-IT-Grundschrift.

Results

To establish the DIC and as a basis for the repository, three different organizational orders were developed for the Gießen site. The extended metadata repository (eMDR) implemented from these offers the possibility of initial feasibility queries with visualization options based on a defined core data set. Furthermore, an authentication via a central Keycloak instance as well as a container-based provisioning takes place. The system is technically and organizationally integrated into all workflows and fulfills the requirements of BSI-IT-Grundschrift. Furthermore, the system is installed in all infrastructurally separate organizational areas, which is made possible by synchronizing the systems with each other.

Discussion

The established structures provide a basis for further research at the Gießen site in the context of the DIC and act as a link, in the role of an information provider, in the inhomogeneous system landscape. In the future, the system must pass benchmarking procedures and usability studies in order to be expanded and adapted on the basis of the results.

8. Danksagung

Ganz besonders danke ich Herrn Prof. Dr. Keywan Sohrabi, welcher mir bereits seit meinem Bachelor-Studium mit Rat und Tat beisteht und mich seit vielen Jahren wissenschaftlich begleitet. Vielen Dank für die hervorragende Betreuung, die spannenden Diskussionen, die Kritik und die Bereitschaft mich jederzeit, seit vielen Jahren, in allen Bereichen zu unterstützen.

Großer Dank gilt übergreifend dem gesamten Gießener MIRACUM-Team, welches diese Arbeit im Zuge des Aufbaus des Datenintegrationszentrums unterstützt hat. Vielen Dank an Romina Blasini, Lisa Steinbrecher, Stefanie Lehn, Doris Budasch, Jördis Beck, Sarah Pohl, Johanna Pfaff, Jasmin Spaar, Dr. Birgit Samans, Jan Sparenberg, Christopher Frank, Sascha Seibel, Bernhard Kaus, Nick Heller, Dr. Daniel Amsel, Kai Schmid, Jannik Sehring, Patrick Fischer, Dr. Amir Tabatabaei, Jan Marquardt, Jochen Zohner und Dr. Achim Michel-Backofen. Für die enge Zusammenarbeit und den fachlichen sowie kollegialen Austausch innerhalb des Teams bin ich sehr dankbar.

Besonders danke ich Herrn Prof. Dr. Kurt Marquardt, welcher mich ebenfalls jederzeit unterstützt hat und die organisatorischen Rahmenbedingungen zur Durchführung des Projektes geschaffen hat. Danke für die vielen spannenden Einblicke und Erfahrungen, die ich sammeln durfte.

Weiter gilt ein besonderer Dank Herrn Dr. Achim Michel-Backofen, welcher mich seit Beginn des MIRACUM-Projektes in allen Belangen unterstützt hat. Danke für den wissenschaftlichen, beruflichen aber vor allem sehr persönlichen und privaten Austausch, der mir speziell während der Dissertation sehr viel bedeutet hat.

Ebenfalls danke ich den Kollegen des IMI des Universitätsklinikums Frankfurt (ehemals MIG) für den fachlichen Austausch sowie die Unterstützung im Kontext des Samply.MDR und die Bereitstellung der Lösung. Vielen Dank an Abishaa Vengadeswaran, Dennis Kadioglu und Prof. Dr. Holger Storf.

Der größte Dank gilt meinen Eltern Petra und Andreas, meiner Schwester Julia sowie meinen Großeltern Gerda und Wilhelm, welche mich während meiner gesamten akademischen Ausbildung immer unterstützt haben.

Insbesondere danke ich dabei meinem verstorbenen Großvater und zugleich besten Freund, ohne den ich diesen Weg nicht hätte gehen können.

9. Abbildungsverzeichnis

Abbildung 1: Standorte der MII-Konsortien (TMF e.V., 2020a)	5
Abbildung 2: Förderphasen der MII (TMF e.V., 2020b).....	5
Abbildung 3: Schematischer Aufbau eines Datenintegrationszentrums (Eigene Darstellung).....	6
Abbildung 4: Systemstruktur innerhalb der DIZ-Infrastruktur (Eigene Darstellung)	22
Abbildung 5: Beispielhafte Struktur eines JSON-Objektes (Eigene Darstellung).....	24
Abbildung 6: Darstellung der Erzeugung einer PostgreSQL-View (Eigene Darstellung)	24
Abbildung 7: Darstellung des Ablaufes der Versionsverwaltung Git (Eigene Darstellung)	30
Abbildung 8: Initiale Definition eines Datenelementes in Samply.MDR (Screenshot aus eigener Systeminstanz).....	32
Abbildung 9: Validierungsinformationen eines Samply.MDR Datenelementes (Screenshot aus eigener Systeminstanz)	32
Abbildung 10: Hierarchische Darstellung von Klassifikationen innerhalb von Samply.MDR (Screenshot aus eigener Systeminstanz).....	33
Abbildung 11: Virtualisierungsschichten und Kapselung der Systeme unter Docker (Eigene Darstellung).....	34
Abbildung 12: Darstellung der Authentifizierung unter Verwendung von Keycloak (Eigene Darstellung).....	37
Abbildung 13: Zusammenhang rechtlicher und organisatorischer Rahmenbedingungen (Eigene Darstellung).....	41
Abbildung 14: Komponenten des definierten Kerndatensatzes (Eigene Darstellung)..	45
Abbildung 15: Beispielhafte Trennung der relevanten Informationskategorien (Eigene Darstellung).....	45
Abbildung 16: Aufbau und Struktur der Semantik durch Samply.MDR (Screenshot aus eigener Systeminstanz)	47
Abbildung 17: Ausschnitt aus dem Datenbankschema des eMDR (Eigene Darstellung)	50
Abbildung 18: Beispielkonfiguration des Keycloak Clients für das eMDR (Screenshot aus eigener Systeminstanz).....	55
Abbildung 19: Funktionsumfang und Eingabeüberprüfung des Frontends (Screenshot aus eigener Systeminstanz)	56

Abbildung 20: Darstellung der Resultate einer Beispielabfrage (Screenshot aus eigener Systeminstanz).....	58
Abbildung 21: Funktionsumfang der Mengenoperationen (Screenshot aus eigener Systeminstanz).....	59
Abbildung 22: Verwendung von Plotly zur grafischen Ergebnisdarstellung (Screenshot aus eigener Systeminstanz)	59
Abbildung 23: Darstellung der Query-Überprüfung durch die Transferstelle (Screenshot aus eigener Systeminstanz)	60
Abbildung 24: Export der Resultate zur DIP (Screenshot aus eigener Systeminstanz)	61
Abbildung 25: Ansicht der Feasability Abfrage innerhalb der DIP (Screenshot aus eigener Systeminstanz).....	62
Abbildung 26: Exportierte Feasability-Abfrage als JSON innerhalb des DIP Datensets (Screenshot aus eigener Systeminstanz)	63
Abbildung 27: Übersicht der Infrastrukturen im Hinblick auf den Datenschutz (Eigene Darstellung).....	63
Abbildung 28: Ampelsystem zum Zwecke der Systemtrennung am Standort (Eigene Darstellung).....	64
Abbildung 29: Datensynchronisation und Auswirkung im Kontext des Ampelsystems (Eigene Darstellung).....	66
Abbildung 30: Darstellung der Komponenten der Container-Infrastruktur (Eigene Darstellung).....	67
Abbildung 31: Gesamtbild des technischen DIZ-Workflows (Eigene Darstellung des DIZ)	68
Abbildung 32: Ausschnitt des technischen DIZ-Workflows - Routinebereich (Eigene Darstellung des DIZ)	69
Abbildung 33: Ausschnitt des technischen DIZ-Workflows - Forschungsbereich (Eigene Darstellung des DIZ)	69
Abbildung 34: eMDR relevanter Ausschnitt des technischen DIZ-Workflows (Eigene Darstellung des DIZ)	70
Abbildung 35: Ausschnitt des organisatorischen DIZ-Workflows vor Projektantrag (Eigene Darstellung des DIZ)	70
Abbildung 36: Gesamtbild des organisatorischen DIZ-Workflows (Eigene Darstellung des DIZ)	71
Abbildung 37: Ausschnitt des organisatorischen DIZ-Workflows nach Projektantrag (Eigene Darstellung des DIZ)	71
Abbildung 38: eMDR relevanter Ausschnitt des organisatorischen DIZ-Workflows (Eigene Darstellung des DIZ)	72

Abbildung 39: Gesamtbild der Systemsicht des DIZ-Workflows (Eigene Darstellung des DIZ)	72
Abbildung 40: Ausschnitt der Systemlandschaft im Routinebereich (Eigene Darstellung des DIZ)	73
Abbildung 41: Ausschnitt der Systemlandschaft im Forschungsbereich (Eigene Darstellung des DIZ)	74

10. Tabellenverzeichnis

Tabelle 1: Vorgaben und Auswirkungen der Nutzungs- und Geschäftsordnungen	48
Tabelle 2: Endpunkte des Samply.MDR	53
Tabelle 3: Endpunkte des eMDR	55
Tabelle 4: Verwendeten Endpunkte der Datenintegrationsplattform	62

11. Abkürzungsverzeichnis

API	Application Programming Interface
BDSG	Bundesdatenschutzgesetz
BMBF	Bundesministerium für Bildung und Forschung
BSI	Bundesamt für Sicherheit in der Informationstechnik
CCP	Clinical Communication Platform, Clinical Communication Platform
CDK	Connector Development Kit
CLI	Command Line Interface
CORS	Cross Origin Resource Sharing
CSV	Comma-separated values
DBMS	Datenbank Management Systems
DIC	Data Integration Centre
DIFUTURE	Data Integration for Future Medicine
DIP	Datenintegrationsplattform
DIZ	Datenintegrationszentrum
DMP	Datenmanagementplan
DSGVO	Datenschutzgrundverordnung
eCRF	Electronical Case Report Form
EDV	Elektronische Datenverarbeitung
eMDR	Erweitertes Metadatenrepositorium
ePA	Elektronische Patientenakte
E-Rezept	Elektronisches Rezept
FAIR	Findability, Accessibility, Interoperability, and Reuse
FHIR	Fast Healthcare Interoperability Resources
GIFONET	Gießener Forschungsnetz
HTTP	Hypertext Transfer Protocol, Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
i2b2	Informatics for Integrating Biology & the Bedside
ICD	International Classification of Diseases
ID	Identifikator
IDE	Integrated Development Environment
IEC	International Electrotechnical Commission
IP	Internet Protokoll
ISO	International Organization for Standardization
IT	Informationstechnik
JLU	Justus-Liebig-Universität Gießen
js	JavaScript
JSON	JavaScript Object Notation
KHEntgG	Krankenhausentgeltgesetz
KHZG	Krankenhauszukunftsgesetz
LDAP	Lightweight Directory Access Protocol
LOINC	Logical Observation Identifiers Names and Codes
MDR	Metadatenrepositorium
MIG	Medical Informatics Group
MIRACUM	Medical Informatics in Research and Medicine
MPI	Master Patient Index
OPS	Operationen- und Prozedurenschlüssel
OSSE	Open Source Registersystem für Seltene Erkrankungen
REST	Representational State Transfer

SGB V Sozialgesetzbuch V
SMITH Smart Medical Information Technology for Healthcare
SNOMED CT Systematized Nomenclature of Medicine - Clinical Terms
SOC Service Organization Control
SPA Single Page Application
SQL Structured Query Language
THM Technische Hochschule Mittelhessen
UAC Use- & Access Committee
UKGM-Gi Universitätsklinikum Gießen und Marburg GmbH - Standort Gießen
VCS Version Control System
YAML Yet Another Markup Language

12. Literaturverzeichnis

- Aggarwal, S., 2019. Flask Framework Cookbook: Over 80 proven recipes and techniques for Python web development with Flask, 2nd Edition. Packt Publishing Ltd.
- Airbyte, Inc., 2022a. Product Features | Airbyte - Open-source ELT platform [WWW Document]. URL <https://airbyte.com/why-airbyte> (accessed 8.5.22).
- Airbyte, Inc., 2022b. Postgres Source | Airbyte Documentation [WWW Document]. URL <https://airbytehq.github.io/integrations/sources/postgres/> (accessed 8.5.22).
- Airbyte, Inc., 2022c. Postgres Destination | Airbyte Documentation [WWW Document]. URL <https://airbytehq.github.io/integrations/destinations/postgres/> (accessed 8.5.22).
- Ali, S.M., Gupta, N., Nayak, G.K., Lenka, R.K., 2016. Big data visualization: Tools and challenges, in: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I). Presented at the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), pp. 656–660. <https://doi.org/10.1109/IC3I.2016.7918044>
- Anusha, K., Rajesh, N., Kavitha, M., Ravinder, N., 2021. Comparative Study of MongoDB vs Cassandra in big data analytics, in: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). Presented at the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1831–1835. <https://doi.org/10.1109/ICCMC51019.2021.9418441>
- Appfigures, 2020. Anzahl der im Apple App-Store verfügbaren mHealth-Apps vom 1. Quartal 2015 bis zum 3. Quartal 2020 [WWW Document]. URL <https://de.statista.com/statistik/daten/studie/1191205/umfrage/anzahl-der-bei-apple-verfuegbaren-mhealth-apps/> (accessed 3.2.23).
- Association of International Certified Professional Accountants, 2022. SOC 2® - SOC for Service Organizations: Trust Services Criteria [WWW Document]. AICPA. URL <https://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/aicpasoc2report.html> (accessed 8.5.22).
- Athey, B.D., Braxenthaler, M., Haas, M., Guo, Y., 2013. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. AMIA Summits Transl. Sci. Proc. 2013, 6–8.
- Austin, T., Sun, S., Lim, Y.S., Nguyen, D., Lea, N., Tapuria, A., Kalra, D., 2015. An Electronic Healthcare Record Server Implemented in PostgreSQL. J. Healthc. Eng. 6, 325–344. <https://doi.org/10.1260/2040-2295.6.3.325>
- Baarsen, J. van, 2014. GitLab Cookbook. Packt Publishing Ltd.
- Baker, T., 2000. A Grammar of Dublin Core. -Lib Mag. 6. <https://doi.org/10.1045/october2000-baker>
- Bargmeyer, B.E., Gillman, D.W., 2020. Metadata Standards and Metadata Registries 10.
- Behm, M.E., Klenk, T., 2021. Digitalisierung im Gesundheitssektor, in: Klenk, T., Nullmeier, F., Wewer, G. (Eds.), Handbuch Digitalisierung in Staat und Verwaltung. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 1–12. https://doi.org/10.1007/978-3-658-23669-4_44-2
- Berman, J., 2013. jsonschema 4.9.1 documentation [WWW Document]. URL <https://python-jsonschema.readthedocs.io/en/stable/> (accessed 8.6.22).
- Berufsverband der Deutschen Urologen e.V., 2021. Der Urologe. Springer Medizin Verlag GmbH, Berlin.
- Bhat, S., 2018. Practical Docker with Python; Build, Release and Distribute your Python App with Docker. Apress, Bangalore, Karnataka, India.
- Bholat, D., 2016. Modelling Metadata in Central Banks. <https://doi.org/10.2139/ssrn.2596225>
- Blischak, J.D., Davenport, E.R., Wilson, G., 2016. A Quick Introduction to Version Control with Git and GitHub. PLOS Comput. Biol. 12, e1004668. <https://doi.org/10.1371/journal.pcbi.1004668>

- Bodenreider, O., 2008. Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. Yearb. Med. Inform. 67–79.
- Bodenreider, O., Cornet, R., Vreeman, D.J., 2018. Recent Developments in Clinical Terminologies - SNOMED CT, LOINC, and RxNorm. Yearb. Med. Inform. 27, 129–139. <https://doi.org/10.1055/s-0038-1667077>
- Boeckhout, M., Zielhuis, G.A., Bredenoord, A.L., 2018. The FAIR guiding principles for data stewardship: fair enough? Eur. J. Hum. Genet. 26, 931–936. <https://doi.org/10.1038/s41431-018-0160-0>
- Boettiger, C., 2015. An introduction to Docker for reproducible research. ACM SIGOPS Oper. Syst. Rev. 49, 71–79. <https://doi.org/10.1145/2723872.2723882>
- Brönneke, J.B., Debatin, J.F., 2022. Digitalisierung im Gesundheitswesen und ihre Effekte auf die Qualität der Gesundheitsversorgung. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 65. <https://doi.org/10.1007/s00103-022-03493-3>
- Bruland, P., Doods, J., Brix, T., Dugas, M., Storck, M., 2018. Connecting healthcare and clinical research_ Workflow optimizations through seamless integration of EHR, pseudonymization services and EDC systems | Elsevier Enhanced Reader [WWW Document]. <https://doi.org/10.1016/j.ijmedinf.2018.09.007>
- Bundesamt für Sicherheit in der Informationstechnik, 2022. IT-Grundschutz-Kompendium. Reguvis Fachmedien GmbH, Köln.
- Bundesamt für Sicherheit in der Informationstechnik, 2017. BSI-Standard 200-2 - IT-Grundschutz-Methodik.
- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023a. BfArM - ICD-10-GM [WWW Document]. URL https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/_node.html (accessed 3.10.23).
- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023b. BfArM - OPS [WWW Document]. URL https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/_node.html (accessed 3.10.23).
- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023c. BfArM - LOINC beim BfArM [WWW Document]. URL https://www.bfarm.de/DE/Kodiersysteme/Terminologien/LOINC-UCUM/LOINC-und-RELMA/LOINC-beim-BfArM/_node.html (accessed 3.10.23).
- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), 2023d. BfArM - Informationen zu SNOMED CT [WWW Document]. URL https://www.bfarm.de/DE/Kodiersysteme/Terminologien/SNOMED-CT/Informationen/_node.html (accessed 3.10.23).
- Bundesministerium für Gesundheit, 2022. E-Health [WWW Document]. Bundesminist. Für Gesundh. URL <https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/e/e-health.html> (accessed 3.2.23).
- Bundestag, 2020. Gesetz für ein Zukunftsprogramm Krankenhäuser (Krankenhauszukunftsgesetz – KHZG).
- Bundestag, 1989. Sozialgesetzbuch (SGB) Fünftes Buch (V).
- Burgess, T., Justin, P., 2016. Flask-Caching — Flask-Caching 1.0.0 documentation [WWW Document]. URL <https://flask-caching.readthedocs.io/en/latest/> (accessed 8.6.22).
- Burke, K., Conroy, K., Horn, R., Stratton, F., Binet, G., 2020. Flask-RESTful — Flask-RESTful 0.3.8 documentation [WWW Document]. URL <https://flask-restful.readthedocs.io/en/latest/> (accessed 8.6.22).
- Business Wire, 2022. Following In-Depth Independent Audit, Airbyte Receives Clean SOC 2 Type 2 Report [WWW Document]. URL <https://www.businesswire.com/news/home/20220707005117/en/Following-In-Depth-Independent-Audit-Airbyte-Receives-Clean-SOC-2-Type-2-Report> (accessed 8.5.22).

- Canakoglu, A., Bernasconi, A., Colombo, A., Masseroli, M., Ceri, S., 2019. GenoSurf: metadata driven semantic search system for integrated genomic datasets. *Database J. Biol. Databases Curation* 2019, baz132. <https://doi.org/10.1093/database/baz132>
- Cavelaars, M., Rousseau, J., Parlayan, C., de Ridder, S., Verburg, A., Ross, R., Visser, G.R., Rotte, A., Azevedo, R., Boiten, J.-W., Meijer, G.A., Belien, J.A.M., Verheul, H., 2015. OpenClinica. *J. Clin. Bioinforma.* 5, S2. <https://doi.org/10.1186/2043-9113-5-S1-S2>
- Choudhury, P., Crowston, K., Dahlander, L., Minervini, M.S., Raghuram, S., 2020. GitLab: work where you want, when you want. *J. Organ. Des.* 9, 23. <https://doi.org/10.1186/s41469-020-00087-8>
- Christie, M. A., Bhandar, A., Nakandala, S., Marru, S., Abeysinghe, E., Pamidighantam, S., Pierce, M. E., 2017. Using Keycloak for Gateway Authentication and Authorization.
- Coyle, K., Baker, T., 2009. Guidelines for Dublin Core Application Profiles [WWW Document]. URL <https://www.dublincore.org/specifications/dublin-core/profile-guidelines/> (accessed 10.11.22).
- Crucial Data Solutions, 2023. TrialKit Technical Specifications [WWW Document]. Crucial Data Solut. URL <https://www.crucialdatasolutions.com/trialkit-technical-specifications/> (accessed 3.10.23).
- Cubukcu, U., Erdogan, O., Pathak, S., Sannakkayala, S., Slot, M., 2021. Citus: Distributed PostgreSQL for Data-Intensive Applications, in: *Proceedings of the 2021 International Conference on Management of Data. Presented at the SIGMOD/PODS '21: International Conference on Management of Data, ACM, Virtual Event China*, pp. 2490–2502. <https://doi.org/10.1145/3448016.3457551>
- Cuggia, M., Combes, S., 2019. The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. *Yearb. Med. Inform.* 28, 195–202. <https://doi.org/10.1055/s-0039-1677917>
- Darms, M., Haßfeld, S., Fedtke, S., 2019. Wichtige Gesetze und Standards der IT-Sicherheit im Gesundheitswesen, in: Darms, M., Haßfeld, S., Fedtke, S. (Eds.), *IT-Sicherheit und Datenschutz im Gesundheitswesen: Leitfaden für Ärzte, Apotheker, Informatiker und Geschäftsführer in Klinik und Praxis*. Springer Fachmedien, Wiesbaden, pp. 153–173. https://doi.org/10.1007/978-3-658-21589-7_7
- Das Europäische Parlament und der Rat der europäischen Union, 2016. VERORDNUNG (EU) 2016/ 679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES - vom 27. April 2016 - zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/ 46/ EG (Datenschutz-Grundverordnung) 88.
- Deutsche Gesellschaft für Urologie, 2019. Mitteilungen der DGU - Registerforschung und Medizininformatik-Initiative: BMBF-Förderung für digitales Harnsteinregister. *Urol.* 58, 711–717. <https://doi.org/10.1007/s00120-019-0977-3>
- DIMDI, 2003. Hinweise für die Benutzung OPS-301 [WWW Document]. Nutz. OPS-301. URL <https://www.dimdi.de/static/de/klassifikationen/ops/kode-suche/erwopshtml21/nutzung.htm> (accessed 3.10.23).
- Docker Inc., 2022a. Manage data in Docker [WWW Document]. Docker Doc. URL <https://docs.docker.com/storage/> (accessed 8.5.22).
- Docker Inc., 2022b. Overview of Docker Compose [WWW Document]. Docker Doc. URL <https://docs.docker.com/compose/> (accessed 8.5.22).
- Dolphin, C., 2013. Flask-CORS — Flask-Cors 3.0.10 documentation [WWW Document]. URL <https://flask-cors.readthedocs.io/en/latest/> (accessed 8.6.22).
- Douglas, K., Douglas, S., 2003. *PostgreSQL: A Comprehensive Guide to Building, Programming, and Administering PostgreSQL Databases*. Sams Publishing.
- Drepper, J., Ganslandt, T., Ingenerf, J., Prasser, F., Sax, U., Wendt, T., Zenker, 2017. *MI-I-Metadaten zur Verfügbarkeit von Daten, Auswertungsmöglichkeiten und Kooperationsoptionen*.

- dtv Verlagsgesellschaft, 2021. Sozialgesetzbuch mit Sozialgerichtsgesetz, 50th ed, dtv Beck Texte. dtv Verlagsgesellschaft mbH & Co. KG, München.
- Duval, E., 2001. Metadata Standards: What, Who & Why. *J UCS* 7, 591–601.
- Eberhardt, F., 2017. Konzeptionierung und Entwicklung einer telemedizinischen Plattform zum Austausch und zur Integration medizinischer Daten (Master's Thesis). University of Applied Sciences, Giessen.
- Eberhardt, F., 2016. Konzeptionierung einer kollaborativen Plattform zum Austausch telemedizinischer Daten (Bachelor's Thesis). University of Applied Sciences, Giessen.
- Eberhardt, F., Michel-Backofen, A., Steinbrecher, L., 2021. Leistungskatalog des Datenintegrationszentrums des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen [WWW Document]. URL https://diz.med.uni-giessen.de/docs/DIZ-Leistungskatalog_V1.0_2021-04-29.pdf (accessed 8.15.22).
- Ecma International, 2017. ECMA-404, 2nd edition, December 2017.
- Eggert, S., 2011. Ansatz zur Gestaltung einer zukunftsfähigen IT-Landschaft am Beispiel der Charité – Universitätsmedizin Berlin 13.
- Fangerau, H., Schulz, S., Noack, T., Müller, I., 2017. Medizinische Terminologie : Ein Kompaktkurs. Lehmanns, Berlin.
- Fasel, D., Meier, A., 2016. Big Data: Grundlagen, Systeme und Nutzungspotenziale. Springer-Verlag.
- Fink, J., Librarian, D.S., University, M., 2014. Docker: a Software as a Service, Operating System-Level Virtualization Framework. Code4Lib J.
- Friele, M., Bröckerhoff, P., Fröhlich, W., Spiecker genannt Döhmann, I., Woopen, C., 2020. Digitale Daten für eine effizientere Prävention: Ethische und rechtliche Überlegungen zu Potenzialen und Risiken. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 63. <https://doi.org/10.1007/s00103-020-03147-2>
- Fröhlich, L., 2018. PostgreSQL 10: Praxisbuch für Administratoren und Entwickler. Carl Hanser Verlag GmbH & Co. KG, München. <https://doi.org/10.3139/9783446456419>
- Fung, K.W., Bodenreider, O., 2019. Knowledge Representation and Ontologies, in: Richesson, R.L., Andrews, J.E. (Eds.), *Clinical Research Informatics, Health Informatics*. Springer International Publishing, Cham, pp. 313–340. https://doi.org/10.1007/978-3-319-98779-8_11
- Fyler, D., 2023. Software - tranSMART. I2b2 TranSMART Found. URL <https://i2b2transmart.org/software/> (accessed 3.10.23).
- Gartner, R., 2016. Metadata. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-40893-4>
- Gauck, J., Merkel, A., Gröhe, H., 2015. Gesetz für sichere digitale Kommunikation und Anwendungen im Gesundheitswesen sowie zur Änderung weiterer Gesetze.
- Gehring, S., Eulenfeld, R., 2018. German Medical Informatics Initiative: Unlocking Data for Research and Health Care. *Methods Inf. Med.* 57, e46–e49. <https://doi.org/10.3414/ME18-13-0001>
- GitLab B.V., 2022. GitLab [WWW Document]. GitLab. URL <https://about.gitlab.com/de-de/> (accessed 8.4.22).
- Golder, W., 2007. Hippokrates und das Corpus Hippocraticum: eine Einführung für Philologen und Mediziner. Königshausen & Neumann.
- Gonçalves, R.S., Musen, M.A., 2019. The variable quality of metadata about biological samples used in biomedical experiments. *Sci. Data* 6, 190021. <https://doi.org/10.1038/sdata.2019.21>
- Haar, C., Buchmann, E., 2019. IT-Grundschutz für die Container-Virtualisierung mit dem neuen BSI-Baustein SYS. 1.6. https://doi.org/10.18420/INF2019_65
- Haaranen, L., Lehtinen, T., 2015. Teaching Git on the Side: Version Control System as a Course Platform, in: *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE '15*. Association for Computing Machinery, New York, NY, USA, pp. 87–92. <https://doi.org/10.1145/2729094.2742608>

- Hampf, C., Bialke, M., Geidel, L., Vass, A., Bahls, T., Blasini, R., Blumentritt, A., Boeker, M., Bruns, C., Jandrig, B., Fünfgeld, M., Heinrich, P., Leddig, T., Michel-Backofen, A., Pirkl, A., Rautenberg, M., Simons, F., Stahl, D., Prokosch, H.-U., Hoffmann, W., 2021. A survey on the current status and future perspective of informed consent management in the MIRACUM consortium of the German Medical Informatics Initiative. *Transl. Med. Commun.* 6, 7. <https://doi.org/10.1186/s41231-021-00086-1>
- Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Havinga, Y., Dijkstra, W., de Keijzer, A., 2010. Adding HL7 version 3 data types to PostgreSQL.
- Heckmann, S., 2022. HL7® FHIR® als Grundlage für moderne Digitalstrategien. https://doi.org/10.1007/978-3-658-36226-3_21
- Henke, V., Hülsken, G., Beß, A., Henkel, A., 2022. Digitalstrategie und Strategieentwicklung im Krankenhaus, in: Henke, V., Hülsken, G., Meier, P.-M., Beß, A. (Eds.), *Digitalstrategie im Krankenhaus: Einführung und Umsetzung von Datenkompetenz und Compliance*. Springer Fachmedien, Wiesbaden, pp. 37–57. https://doi.org/10.1007/978-3-658-36226-3_5
- Hippen, A.A., Greene, C.S., 2021. Expanding and Remixing the Metadata Landscape. *Trends Cancer* 7, 276–278. <https://doi.org/10.1016/j.trecan.2020.10.011>
- Hübner, U., Esdar, M., Hüsters, J., Liebe, J.-D., Rauch, J., Thye, J., Weiß, J.-P., 2018. IT-Report Gesundheitswesen: Schwerpunkt – Wie reif ist die IT in deutschen Krankenhäusern? Osnabrück.
- Institut für das Entgeltsystem im Krankenhaus GmbH, 2022. Datenlieferung gem. § 21 Abs.1 KHEntgG, InEK GmbH [WWW Document]. URL https://www.g-drg.de/Datenlieferung_gem._21_KHEntgG/Datenlieferung_gem._21_Abs.1_KHEntgG (accessed 8.9.22).
- ISO/IEC JTC 1/SC 32 Data management and interchange, 2013. ISO/IEC 11179-3:2013, Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes.
- Janda, C., 2019. *Medizinrecht*, 4th ed. UVK Verlag, München.
- Jansen, M., 2016. FAIR Principles [WWW Document]. GO FAIR. URL <https://www.go-fair.org/fair-principles/> (accessed 10.11.22).
- JetBrains s.r.o., 2022a. PyCharm: die Python-IDE von JetBrains für professionelle Entwickler [WWW Document]. JetBrains. URL <https://www.jetbrains.com/de-de/pycharm/> (accessed 8.4.22).
- JetBrains s.r.o., 2022b. Features - PyCharm [WWW Document]. JetBrains. URL <https://www.jetbrains.com/pycharm/features/> (accessed 8.6.22).
- Kadioglu, D., Breil, B., Knell, C., Lablans, M., Mate, S., Schlue, D., Serve, H., Storf, H., Wagner, T., Weingardt, P., 2018. Sampil.MDR – A Metadata Repository and Its Application in Various Research Networks 5. <https://doi.org/10.3233/978-1-61499-896-9-50>
- Kadioglu, D., Weingardt, P., Ückert, F., Wagner, T., 2016. Sampil.MDR – Ein Open-Source-Metadaten-Repository. German Medical Science GMS Publishing House, p. DocAbstr. 425. <https://doi.org/10.3205/16gmds149>
- Kaminski, S., 2016. Python 3, Python 3. De Gruyter Oldenbourg.
- Kapsner, L.A., Mang, J.M., Mate, S., Seuchter, S.A., Vengadeswaran, A., Bathelt, F., Deppenwiese, N., Kadioglu, D., Kraska, D., Prokosch, H.-U., 2021. Linking a Consortium-Wide Data Quality Assessment Tool with the MIRACUM Metadata Repository. *Appl. Clin. Inform.* 12, 826–835. <https://doi.org/10.1055/s-0041-1733847>
- Karasawa, M., Hover, J., Misawa, S., 2020. Federated User Account Management. *EPJ Web Conf.* 245, 07058. <https://doi.org/10.1051/epjconf/202024507058>
- Kennedy, P., 2008. Manifestations of metadata: from Alexandria to the Web – old is new again. *Aust. Libr. J.* 57, 128–146. <https://doi.org/10.1080/00049670.2008.10722461>

- Klauber, J., Geraedts, M., Friedrich, J., Wasem, J., 2019. Krankenhaus-Report 2019 - Das digitale Krankenhaus.
- Komatsoulis, G.A., Warzel, D.B., Hartel, F.W., Shanbhag, K., Chilukuri, R., Fragoso, G., de Coronado, S., Reeves, D.M., Hadfield, J.B., Ludet, C., Covitz, P.A., 2008. caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability. *J. Biomed. Inform.* 41, 106–123. <https://doi.org/10.1016/j.jbi.2007.03.009>
- Kumar, S., 2022. python-dotenv [WWW Document]. URL <https://github.com/theskumar/python-dotenv> (accessed 8.6.22).
- Kush, R., Nordo, A., 2019. Data Sharing and Reuse of Health Data for Research, in: Richesson, R.L., Andrews, J.E. (Eds.), *Clinical Research Informatics, Health Informatics*. Springer International Publishing, Cham, pp. 379–402. https://doi.org/10.1007/978-3-319-98779-8_11
- Kushida, C.A., Nichols, D.A., Jadrnicek, R., Miller, R., Walsh, J.K., Griffin, K., 2012. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. *Med. Care* 50, S82–101. <https://doi.org/10.1097/MLR.0b013e3182585355>
- Landrock, H., Gadatsch, A., 2018. Rechtliche und ethische Aspekte, in: Landrock, H., Gadatsch, A. (Eds.), *Big Data im Gesundheitswesen kompakt: Konzepte, Lösungen, Visionen, IT kompakt*. Springer Fachmedien, Wiesbaden, pp. 63–70. https://doi.org/10.1007/978-3-658-21096-0_6
- Leider, J., Henry, A., Marlow, L., Kael, 2022. Why you should be using Vuetify [WWW Document]. Vuetify. URL <https://vuetifyjs.com/en/introduction/why-vuetify/> (accessed 8.6.22).
- Leipzig, J., Nüst, D., Hoyt, C.T., Ram, K., Greenberg, J., 2021. The role of metadata in reproducible computational research. *Patterns* 2, 100322. <https://doi.org/10.1016/j.patter.2021.100322>
- Liu, C., Dunsmuir, D., Krepiakevich, A., Kissoon, T., Ansermino, M., Trawin, J., 2021. Development of a survey reporting process to facilitate health systems strengthening and quality improvement. *Pediatr. Crit. Care Med.* 22, 270. <https://doi.org/10.1097/01.pcc.0000740504.82641.3a>
- Loeliger, J., McCullough, M., 2012. *Version Control with Git: Powerful Tools and Techniques for Collaborative Software Development*. O'Reilly Media, Inc.
- Lux, T., Breil, B., Dörries, M., Gensorowsky, D., Greiner, W., Pfeiffer, D., Rebitschek, F.G., Gigerenzer, G., Wagner, G.G., 2017. Digitalisierung im Gesundheitswesen — zwischen Datenschutz und moderner Medizinversorgung. *Wirtschaftsdienst* 97. <https://doi.org/10.1007/s10273-017-2200-8>
- Maia, I., 2015. *Building Web Applications with Flask*. Packt Publishing Ltd.
- Marx, G., Rossaint, R., Marx, N. (Eds.), 2021. *Telemedizin: Grundlagen und praktische Anwendung in stationären und ambulanten Einrichtungen*. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-60611-7>
- Matthias, K., Kane, S.P., Lorenzen, K., 2016. *Docker Praxiseinstieg: Deployment, Testen und Debugging von Containern in Produktivumgebungen, mitp Professional*. MITP, [Frechen].
- Meißner, A., 2018. Risiken und Nebenwirkungen der digitalisierten Medizin. *NeuroTransmitter* 29, 30–37. <https://doi.org/10.1007/s15016-018-6385-x>
- Merkel, D., 2014. *Docker: Lightweight Linux Containers for Consistent Development and Deployment*.
- Michel-Backofen, A., 2021. Datenintegrationszentrum Gießen [WWW Document]. Datenintegrationszentrum Gießen. URL <https://diz.med.uni-giessen.de/> (accessed 6.25.22).
- Microsoft, 2022a. Why Visual Studio Code? [WWW Document]. URL <https://code.visualstudio.com/docs/editor/whyvscode> (accessed 8.6.22).
- Microsoft, 2022b. Working with JavaScript in Visual Studio Code [WWW Document]. URL <https://code.visualstudio.com/docs/nodejs/working-with-javascript> (accessed 8.6.22).
- Microsoft, 2022c. Vue JavaScript Tutorial in Visual Studio Code [WWW Document]. URL <https://code.visualstudio.com/docs/nodejs/vuejs-tutorial> (accessed 8.6.22).

- MIRACUM, 2022. MIRACOLIX-Tools. URL <https://www.miracum.org/das-konsortium/datenintegrationszentren/miracolix-tools/> (accessed 8.9.22).
- MIT Critical Data, Johnson, A., Marshall, D., Raffa, J., Saliccioli, J., Paik, K.E., Celi, L.A., Komorowski, M., Ghassemi, M., Charlton, P., Pollard, T.J., Naumann, T., 2016. *Secondary Analysis of Electronic Health Records*, 1st ed. Springer Charm, Cambridge, MA.
- Mouat, A., 2016. *Docker*. dpunkt.verlag GmbH, Heidelberg.
- Murphy, S.N., Weber, G., Mendis, M., Gainer, V., Chueh, H.C., Churchill, S., Kohane, I., 2010. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Inform. Assoc.* 17, 124–130. <https://doi.org/10.1136/jamia.2009.000893>
- Muscholl, M., Lablans, M., Wagner, T.O., Ückert, F., 2014. OSSE – open source registry software solution. *Orphanet J. Rare Dis.* 9, O9. <https://doi.org/10.1186/1750-1172-9-S1-O9>
- N., D.D., Cholli, N.G., 2020. A Review on Identity and Access Management Server (KeyCloak). *Int. J. Secur. Priv. Pervasive Comput. IJSPPC* 12, 46–53. <https://doi.org/10.4018/IJSPPC.2020070104>
- National Cancer Institute, Warzel, D.B., 2023. caDSR Wiki - caDSR - NCI Wiki [WWW Document]. URL <https://wiki.nci.nih.gov/display/caDSR/caDSR+Wiki> (accessed 3.10.23).
- National Institutes of Health, 2023. CEDAR, Better metadata means better science [WWW Document]. CEDAR Better Metadata Means Better Sci. URL <https://metadatascenter.org/> (accessed 3.10.23).
- Nguyen, Q., 2019. *Hands-On Application Development with PyCharm: Accelerate your Python applications using practical coding techniques in PyCharm*. Packt Publishing Ltd.
- Norimatsu, T., Nakamura, Y., Yamauchi, T., 2022. Flexible Method for Supporting OAuth 2.0 Based Security Profiles in Keycloak.
- O'Connor, M.J., Warzel, D.B., Martínez-Romero, M., Hardi, J., Willrett, D., Egyedi, A.L., Eftekhari, A., Graybeal, J., Musen, M.A., 2020. Unleashing the value of Common Data Elements through the CEDAR Workbench. *AMIA. Annu. Symp. Proc.* 2019, 681–690.
- Ojamaa, A., Dūūna, K., 2012. Assessing the security of Node.js platform, in: *2012 International Conference for Internet Technology and Secured Transactions. Presented at the 2012 International Conference for Internet Technology and Secured Transactions*, pp. 348–355.
- Olla, P., Shimskey, C., 2015. mHealth taxonomy: a literature survey of mobile health applications. *Health Technol.* 4, 299–308. <https://doi.org/10.1007/s12553-014-0093-8>
- OpenClinica, LLC, 2023. OpenClinica's Electronic Data Capture (EDC) Platform [WWW Document]. OpenClinica. URL <https://www.openclinica.com/solutions/electronic-data-capture-edc/> (accessed 3.10.23).
- Orlivskiy, S., Deomin, B., Averianova, O., 2021. Pagination And Its Efficient Methods For RESTful Web Services, in: *2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON). Presented at the 2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 567–571. <https://doi.org/10.1109/UKRCON53503.2021.9575139>
- Passaglia, A., 2017. *Vue.js 2 Cookbook*. Packt Publishing Ltd.
- Peterke, L., 2019. *Vue.js kurz & gut*. dpunkt.verlag GmbH, Heidelberg.
- Pfannstiel, M.A., Holl, F., Swoboda, W.J. (Eds.), 2020. *mHealth-Anwendungen für chronisch Kranke: Trends, Entwicklungen, Technologien*. Springer Fachmedien Wiesbaden, Wiesbaden. <https://doi.org/10.1007/978-3-658-29133-4>
- Pittard, W.S., Li, S., 2020. The Essential Toolbox of Data Science: Python, R, Git, and Docker, in: Li, S. (Ed.), *Computational Methods and Data Analysis for Metabolomics, Methods in Molecular Biology*. Springer US, New York, NY, pp. 265–311. https://doi.org/10.1007/978-1-0716-0239-3_15
- Plotly, 2022a. Plotly Open Source Graphing Library for Python [WWW Document]. URL <https://plotly.com/python/> (accessed 8.6.22).

- Plotly, 2022b. Plotly JavaScript Open Source Graphing Library [WWW Document]. URL <https://plotly.com/javascript/> (accessed 8.5.22).
- Plotly, 2022c. Plotly, Inc. [WWW Document]. URL <https://plotly.com/about-us> (accessed 8.5.22).
- Prokosch, H.-U., 2022. Über uns - MIRACUM [WWW Document]. Über Uns - MIRACUM. URL <https://www.miracum.org/ueber-uns/> (accessed 6.25.22).
- Prokosch, H.-U., Acker, T., Bernarding, J., Binder, H., Boeker, M., Boerries, M., Daumke, P., Ganslandt, T., Hesser, J., Höning, G., Neumaier, M., Marquardt, K., Renz, H., Rothkötter, H.-J., Schade-Brittlinger, C., Schmücker, P., Schüttler, J., Sedlmayr, M., Serve, H., Sohrabi, K., Storf, H., 2018. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods Inf. Med.* 57, e82–e91. <https://doi.org/10.3414/ME17-02-0025>
- Prokosch, Prof.Dr.H.-U., 2022. Datenintegrationszentren > MIRACUM [WWW Document]. URL <https://www.miracum.org/das-konsortium/datenintegrationszentren/> (accessed 6.25.22).
- Ramírez, S., 2022. FastAPI [WWW Document]. URL <https://fastapi.tiangolo.com/> (accessed 8.9.22).
- Rector, A.L., Solomon, W.D., Nowlan, W.A., Rush, T.W., Zanstra, P.E., Claassen, W.M.A., 1995. A Terminology Server for Medical Language and Medical Information Systems.
- Red Hat, Inc., 2022. Keycloak Server Administration Guide [WWW Document]. URL https://www.keycloak.org/docs/latest/server_admin/index.html (accessed 8.4.22).
- Regenstrief Institute, Inc, 2023. About LOINC [WWW Document]. LOINC. URL <https://loinc.org/about/> (accessed 3.10.23).
- Reitz, K., 2022. Requests: HTTP for Humans™ — Requests 2.28.1 documentation [WWW Document]. URL <https://requests.readthedocs.io/en/latest/> (accessed 8.6.22).
- ResearchManager, 2021. Electronic Data Capture – ResearchManager. URL <https://my-researchmanager.com/de/electronic-data-capture-de/> (accessed 3.10.23).
- Richesson, R.L., Lynch, C.O., Hammond, W.E., 2019. Developing and Promoting Data Standards for Clinical Research, in: Richesson, R.L., Andrews, J.E. (Eds.), *Clinical Research Informatics, Health Informatics*. Springer International Publishing, Cham, pp. 313–340. https://doi.org/10.1007/978-3-319-98779-8_11
- Rienhoff, O., Semler, S.C., 2015. Terminologien und Ordnungssysteme in der Medizin : Standortbestimmung und Handlungsbedarf in den deutschsprachigen Ländern. MWV Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Riley, J., National Information Standards Organization (U.S.), 2017. Understanding metadata: what is metadata, and what is it for?
- Rojas, C., 2019. *Building Progressive Web Applications with Vue.js: Reliable, Fast, and Engaging Apps with Vue.js*. Apress.
- Rui, Y., Zhijie, X., Tianjie, X., Yu, Z., 2020. Multi-modal big data knowledge aggregation for advanced automobile intelligent manufacturing operation and maintenance. *J. Phys. Conf. Ser.* 1653, 012026. <https://doi.org/10.1088/1742-6596/1653/1/012026>
- Sagiroglu, S., Sinanc, D., 2013. Big data: A review, in: 2013 International Conference on Collaboration Technologies and Systems (CTS). Presented at the 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42–47. <https://doi.org/10.1109/CTS.2013.6567202>
- Santaló, J., Berdasco, M., 2022. Ethical implications of epigenetics in the era of personalized medicine. *Clin. Epigenetics* 14, 44. <https://doi.org/10.1186/s13148-022-01263-1>
- Seibel, S.H., 2021. Konzeptionierung und Entwicklung eines Support- und Feedbackmanagements für Benutzer, sowie einer Fehler- und Anwendungsüberwachung mit Anbindung an die Datenintegrationsplattform (Master's Thesis). University of Applied Sciences, Giessen.
- Semler, S.C., Wissing, F., Heyder, R., 2018. German Medical Informatics Initiative. *Methods Inf. Med.* 57, e50–e56. <https://doi.org/10.3414/ME18-03-0003>

- Shah, J., Rajgor, D., Pradhan, S., McCreedy, M., Zaveri, A., Pietrobon, R., 2010. Electronic Data Capture for Registries and Clinical Trials in Orthopaedic Surgery: Open Source versus Commercial Systems. *Clin. Orthop. Relat. Res.* 468, 2664–2671. <https://doi.org/10.1007/s11999-010-1469-3>
- Singh, M., Verma, A., Parasher, A., Chauhan, N., Budhiraja, G., 2019. Implementation of database using python flask framework. <https://doi.org/10.18535/ijecs/v8i12.4399>
- SNOMED International, 2023. SNOMED CT - Value Proposition [WWW Document]. SNOMED Int. URL <https://www.snomed.org/value-proposition?lang=de> (accessed 3.10.23).
- solidIT, 2022. DB-Engines Ranking [WWW Document]. DB-Engines. URL <https://db-engines.com/de/ranking> (accessed 8.4.22).
- Sparenberg, J., 2019. Implementierung einer Plattform zur Zentralisierung von medizinischen Forschungsdaten und digitalen Werkzeugen (Master's Thesis). University of Applied Sciences, Giessen.
- Steger, F., 2019. Medizinische Terminologie, 5th ed. Vandenhoeck & Ruprecht GmbH & Co. KG, Göttingen.
- Steyer, R., 2019. Webanwendungen erstellen mit Vue.js: MVVM-Muster für konventionelle und Single-Page-Webseiten. Springer Fachmedien Wiesbaden, Wiesbaden. <https://doi.org/10.1007/978-3-658-27170-1>
- Stinson, B., 2001. PostgreSQL Essential Reference. Sams Publishing.
- Stonebraker, M., Rowe, L.A., 1986. THE DESIGN OF POSTGRES 16.
- Studer, T., 2019. SQL zur Datendefinition und Datenmanipulation, in: Studer, T. (Ed.), Relationale Datenbanken: Von den theoretischen Grundlagen zu Anwendungen mit PostgreSQL. Springer, Berlin, Heidelberg, pp. 105–124. https://doi.org/10.1007/978-3-662-58976-2_6
- Sun, H., Bonetta, D., Humer, C., Binder, W., 2018. Efficient dynamic analysis for Node.js, in: Proceedings of the 27th International Conference on Compiler Construction. Presented at the CGO '18: 16th Annual IEEE/ACM International Symposium on Code Generation and Optimization, ACM, Vienna Austria, pp. 196–206. <https://doi.org/10.1145/3178372.3179527>
- Sunyaev, A., Dünnebeil, S., Mauro, C., Leimeister, J.-M., Krcmar, H., 2010. IT-Standards im Gesundheitswesen: Überblick und Entwicklungsperspektiven mit der Einführung Service- Orientierter Architekturen 6.
- Tamersoy, A., Loukides, G., Nergiz, M.E., Saygin, Y., Malin, B., 2012. Anonymization of Longitudinal Electronic Medical Records. *IEEE Trans. Inf. Technol. Biomed.* 16, 413–423. <https://doi.org/10.1109/TITB.2012.2185850>
- The i2b2 Community, Klann, J., Metta, R., 2023. i2b2 Community Wiki [WWW Document]. I2b2 Community Wiki. URL <https://community.i2b2.org/wiki/> (accessed 8.9.22).
- the pandas development team, Plotly, 2022. Pandas Plotting Backend in Python [WWW Document]. URL <https://plotly.com/python/pandas-backend/> (accessed 8.6.22).
- The PostgreSQL Global Development Group, 2022. Appendix D. SQL Conformance [WWW Document]. PostgreSQL Doc. URL <https://www.postgresql.org/docs/14/features.html> (accessed 8.4.22).
- The PostgreSQL Global Development Group, 2018. JSON Functions and Operators [WWW Document]. PostgreSQL Doc. URL <https://www.postgresql.org/docs/9.3/functions-json.html> (accessed 8.4.22).
- Tilkov, S., Vinoski, S., 2010. Node.js: Using JavaScript to Build High-Performance Network Programs. *IEEE Internet Comput.* 14, 80–83. <https://doi.org/10.1109/MIC.2010.145>
- TMF e.V., 2020a. Über die Initiative | Medizininformatik-Initiative [WWW Document]. URL <https://www.medizininformatik-initiative.de/de/ueber-die-initiative> (accessed 3.2.23).
- TMF e.V., 2020b. Phasen | Medizininformatik-Initiative [WWW Document]. URL <https://www.medizininformatik-initiative.de/index.php/de/ueber-die-initiative/phasen> (accessed 6.25.22).

- Townend, D., 2010. Privacy, health insurance, and medical research: tensions raised by European data protection law. *New Genet. Soc.* 29, 477–493. <https://doi.org/10.1080/14636778.2010.528194>
- Ulrich, A., Massinger, M., 2022. Vue.plotly - Wrapper for the charting library plotly.js - Made with Vue.js [WWW Document]. URL <https://madewithvuejs.com/vueplotly> (accessed 8.5.22).
- Ulrich, H., Behrend, P., Wiedekopf, J., Drenkhahn, C., Kock-Schoppenhauer, A.-K., Ingenerf, J., 2021. Hands on the Medical Informatics Initiative Core Data Set - Lessons Learned from Converting the MIMIC-IV. *Stud. Health Technol. Inform.* 283, 119–126. <https://doi.org/10.3233/SHTI210549>
- Universitätsklinikum Frankfurt, 2022. Hintergrund – OSSE [WWW Document]. OSSE. URL <https://www.osse-register.de/hintergrund/> (accessed 8.3.22).
- Universitätsmedizin Mainz - IMBEI, 2015. MDR REST Interface [WWW Document]. URL <https://doc.samply.de/mdr.rest/rest-mdr.html> (accessed 8.15.22).
- Unpingco, J., 2021. Python Programming for Data Analysis. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-030-68952-0>
- Vanderbilt University, 2023. REDCap. URL <https://www.project-redcap.org/> (accessed 3.10.23).
- Varrazzo, D., 2021. PostgreSQL driver for Python — Psycopg [WWW Document]. URL <https://www.psycopg.org/> (accessed 8.6.22).
- Vengadeswaran, A., Neuhaus, P., Hegselmann, S., Storf, H., Kadioglu, D., 2019. Semantically Annotated Metadata: 7.
- von Kurnatowski, L., Stoffers, M., Haupt, C., 2021. Provenance based software dashboards.
- Vue.js Contributors, 2020. Debugging in VS Code — Vue.js [WWW Document]. URL <https://v2.vuejs.org/v2/cookbook/debugging-in-vscode.html> (accessed 8.6.22).
- Vuorre, M., Curley, J.P., 2018. Curating Research Assets: A Tutorial on the Git Version Control System. *Adv. Methods Pract. Psychol. Sci.* 1, 219–236. <https://doi.org/10.1177/2515245918754826>
- Vyshnavi, V.R., Malik, A., 2019. Efficient Way of Web Development Using Python and Flask 6, 4.
- Waldschmidt-Dietz, F., Krippes, C., 2016. Forschungsdaten an der JLU Gießen: Auswertung einer Umfrage aus dem Juli 2016 32.
- Warin, T., 2021. Global Research on Coronaviruses: Metadata-Based Analysis for Public Health Policies. *JMIR Med. Inform.* 9, e31510. <https://doi.org/10.2196/31510>
- Weigend, M., 2020. Python Ge-Packt, mitp Ge-packt. MITP, [Germany?].
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- World Health Organization, 2011. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization.
- Zabka, S., Ammon, D., Ganslandt, T., Gewehr, J., Haverkamp, C., Kiefer, S., Lautenbacher, H., Thun, S., Boeker, M., 2019. Towards a Medication Core Data Set for the Medical Informatics Initiative (MII): Initial Mapping Experience between the German Procedure Classification (OPS) and the Identification of Medicinal Products (IDMP) 12.
- Zabriskie, M., 2022. axios [WWW Document]. URL <https://github.com/axios/axios> (accessed 8.6.22).
- Zeng, M.L., Subrahmanyam, B., Shreve, G.M., 2005. Metadata Quality Study for the National Science Digital Library (NSDL) Metadata Repository, in: Chen, Z., Chen,

- H., Miao, Q., Fu, Y., Fox, E., Lim, E. (Eds.), Digital Libraries: International Collaboration and Cross-Fertilization, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 339–340. https://doi.org/10.1007/978-3-540-30544-6_36
- Zingg, T., Sojer, R., Röthlisberger, F., 2019. Digitalisierung in der ambulanten Gesundheitsversorgung. Schweiz. Ärzetzg. <https://doi.org/10.4414/saez.2019.17521>
- Zolkifli, N.N., Ngah, A., Deraman, A., 2018. Version Control System: A Review. Procedia Comput. Sci., The 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018): Empowering Smart Technology in Digital Era for a Better Life 135, 408–415. <https://doi.org/10.1016/j.procs.2018.08.191>
- Zozus, M.N., Kahn, M.G., Weiskopf, N.G., 2019. Data Quality in Clinical Research, in: Richesson, R.L., Andrews, J.E. (Eds.), Clinical Research Informatics, Health Informatics. Springer International Publishing, Cham, pp. 213–248. https://doi.org/10.1007/978-3-319-98779-8_11

13. Anhang

13.1 Daten-Zugriffs- und Nutzungsordnung

**Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum
des Fachbereichs Medizin der Justus-Liebig-Universität Gießen
und des UKGM / Standort Gießen**

Präambel

Ein zentrales Ziel der Medizininformatik-Initiative¹ des Bundesministeriums für Bildung und Forschung (BMBF) ist es, einheitliche Rahmenbedingung für einen bundesweit einheitlichen Datenzugang und Datenaustausch zu schaffen. Zur Umsetzung dieses Ziels müssen neben der technischen Harmonisierung vor allem einheitliche organisatorisch und rechtlich abgesicherte Rahmenbedingungen für den Datenzugang und die Datennutzung festgelegt werden. Auf nationaler Ebene hat sich die Arbeitsgruppe Data Sharing des Nationalen Steuerungsgremiums (NSG) mit den Grundprinzipien des Datenaustausches befasst und ein Eckpunktepapier einer einheitlichen Nutzungsordnung erarbeitet.

Nachfolgend wurden für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des Universitätsklinikums Gießen und Marburg GmbH (Standort Gießen) (Datenintegrationszentrum Gießen, DIZ) die Eckpunkte dieser nationalen Nutzungsordnung für die lokale Zugriff- und Nutzungsordnung berücksichtigt und gemäß den lokalen Rahmenbedingungen angepasst und weiter ausgeführt. Neben den Regelungen dieser Zugriffs- und Nutzungsordnung sind ergänzend datenschutzrechtliche Bestimmungen (auf Landes- Bundes- und EU-Ebene), Regelungen des Krankenhausrechts, Vorgaben des ärztlichen Berufsrechts, Gesetze zu Patenten und Urheberrechten sowie ggf. weitere rechtliche Rahmenbedingungen zum Schutz von Patienten im Versorgungs- (BGB) oder Studienfall (AMG, MPG) zu beachten.

Das MIRACUM Konsortiums umfasst zehn deutsche Universitätsmedizin-Standorte (Magdeburg, Marburg, Gießen, Frankfurt, Mainz, Mannheim, Erlangen, Greifswald, Dresden und Freiburg). Das Leitthema der Ausschreibung und damit auch des MIRACUM Konsortiums sowie des DIZ ist es, mittels an den Standorten etablierter DIZ, einen verbesserten Datenzugang, eine gemeinsame Datennutzung und (wo nötig auch) Datenaustausch zu ermöglichen. In diesem Kontext ermöglicht das DIZ eine lokale Zusammenführung von Daten und bietet Optionen für spätere standortübergreifende Datennutzung im nationalen Umfeld. Die nachfolgende Daten-Zugriffs- und -Nutzungsordnung beschreibt die Governancestruktur und die Richtlinien, die am DIZ als Organisationsrahmen und Regelwerk zur Datennutzung, -bereitstellung und -austausch verbindlich zur Anwendung kommen sollen.

Die vorliegende Version einer Zugriffs- und Nutzungsordnung fokussiert zunächst auf die Bereitstellung dokumentierter Routine-Versorgungs-Daten für sekundäre Zwecke. Die im Rahmen der Medizininformatik-Initiative aufzubauenden DIZs sollen später zusätzlich auch die Aufgabe haben, primär im Forschungskontext erhobene Daten (aus Studien, Registern, Kohorten etc.) zu integrieren, aufzubereiten und diese dann in Gesamtsichten (Versorgungs- und Forschungsdaten) für Forschungs- und Versorgungszwecke zur Verfügung zu stellen. Da die für MIRACUM definierten Use Cases diese Aspekte der Datenintegration im DIZ noch nicht enthalten, werden auch die Regularien zu diesen Zukunftsoptionen in dieser Nutzungsordnung nicht berücksichtigt. Eine Adaption der Nutzungsordnung kann dann themenaktuell erfolgen. Dabei werden neben den o.g. rechtlichen Regelungen und den jeweils

¹ nachfolgend als BMBF MII bezeichnet

geltenden forschungsethischen und forschungsrechtlichen Normen folgender Grundsatz zugrunde gelegt: Eine Übernahme, Speicherung, Nutzung, und Weitergabe von Forschungsdaten durch das DIZ ist nur mit Zustimmung der Personen möglich, die an der Erzeugung der Daten beteiligt waren (Datengeber der originären F&E-Daten). Die Zustimmung bezieht sich auf den Einzelfall des definierten und von UAC freigegebenen Projektes und hat keine grundsätzliche Dimension.

Ergänzend ist anzumerken, dass in dieser Daten-Zugriffs- und Nutzungsordnung Verweise auf nachgeordnete Dokumente enthalten sind, die Strukturen, Begrifflichkeiten und Details erläutern. Die Texte sind nicht als Erweiterungen des Hauptdokumentes zu verstehen, sondern sollen lediglich Einzelaspekte des Hauptdokumentes im Verständnis transparenter darstellen.

Definitionen

a) Use & Access Committee

Gemäß des Eckpunktepapiers einer einheitlichen Nutzungsordnung ist jedes Konsortium aufgefordert ein oder mehrere Use & Access Committees (UAC) und deren Zuständigkeit für die Datenintegrationszentren (DIZ) festzulegen. Diese UACs führen die fachliche und inhaltliche Prüfung von Anträgen auf Datennutzung durch und involvieren dabei die lokalen Kompetenzen aus den Bereichen Datenschutz, Ethikkommission und Justitiariaten. Sie führen eine Entscheidung zur Nutzungsfreigabe herbei, wobei die Zustimmung identifizierbarer relevanter Datengeber eingeholt wird.

b) Datenempfänger

Als Datenempfänger werden Personen oder Institutionen bezeichnet, die im Rahmen der BMBF MII bereitgestellte Services nutzen, um Daten zu finden und über diese Daten weitere Informationen im Detail zu erhalten. Diese Personen oder Institutionen können

- Teil bzw. wissenschaftliche Mitarbeiter des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des Universitätsklinikum Gießen und Marburg GmbH (Standort Gießen) sein,
- Teil bzw. wissenschaftliche Mitarbeiter eines der geförderten Konsortien sein, oder
- außerhalb eines der teilnehmenden Konsortien angesiedelt sein

In diesem Kontext wird auch auf die Definition für „Consumer“ in OAIS² hingewiesen.

c) Datengeber

Als Datengeber werden Personen oder Institutionen bezeichnet, die im Rahmen der BMBF MII Daten zur Nutzung im Forschungs- oder Versorgungskontext zur Verfügung stellen.

In diesem Zusammenhang wird auch auf die Definition für „Producer“ in OAIS hingewiesen.

d) Antrag

Anträge stellen die inhaltlichen und formalen Rahmenbedingungen eines Forschungsprojekts dar.

e) Transferstelle

Die Transferstelle ist Teil des DIZ und für die technische und ggf. administrative Umsetzung des Data Sharing verantwortlich.

f) Zentrale Antrags- und Registerstelle (ZARS)

Die Antrags- und Registerstelle ist eine zentrale und neutrale Einrichtung der Begleitstruktur, die koordinative und administrative Aufgaben übernimmt. Sie nimmt Anträge von externen Partnern auf Datennutzung an und reicht diese an die zuständigen UACs bzw. Transferstellen der Datenintegrationszentren weiter. Sie führt ein öffentliches, zentrales Register aller Anträge und daraus resultierender Data-Sharing-Projekte und ein Verzeichnis der lokalen Nutzungs- und Verfahrensordnungen.

² CCSDS Reference Model for an Open Archival Information System (OAIS). Recommended Practice. Magenta Book. 2012. Consultative Committee for Space Data Systems: Recommendation for Space Data System Practices, CCSDS 650.0-M-2, <https://public.ccsds.org/pubs/650x0m2.pdf> (Abruf: 2016-11-23).

g) Datennutzungsvereinbarung

Die Datennutzungsvereinbarung regelt alle rechtlichen und administrativen Aspekte für die an einem Data-Sharing-Projekt beteiligten Partner und bezieht sich inhaltlich auf den gemeinsamen Antrag der beteiligten Partner zu einem Datennutzungsprojekt.

1. Struktur des DIZ

1.1 Vertreter der Betreiber des DIZ

Die Vertreter werden im Einvernehmen zwischen dem Dekanat des Fachbereichs Medizin der JLU (Justus-Liebig-Universität) und der Geschäftsführung des UKGM (Standort Gießen) benannt. Auch eventuelle Änderungen dieser Besetzung sind einvernehmlich zwischen den genannten Partnern zu klären. Zu den Aufgaben gehört die Vertretung der Betreiber in Bezug auf das DIZ nach innen und außen sowie die Bereitstellung jener Ressourcen (Personal-, Sach- und Investivmittel), welche aus den MIRACUM-Fördermitteln oder aus ergänzenden Bereitstellungen des Fachbereichs Medizin bzw. des UKGM resultieren. Dazu gehören insbesondere Serverkapazitäten, Softwarelizenzen, Anwendungswerkzeuge und Storage. Ferner entscheiden die Vertreter gemeinsam mit der DIZ-Leitung über die Einstellung der im Rahmen des Projekts benötigten Mitarbeiter. Zu den weiteren Aufgaben gehört die Erstellung und Überarbeitung der mit dem DIZ rechtlich in Zusammenhang stehenden Nutzerordnungen und Regelungen sowie die Abstimmung dieser Regelungen mit dem Dekanat des Fachbereichs Medizin der JLU und der Geschäftsführung von UKGM / Standort Gießen. Des Weiteren bearbeiten die Vertreter im Rahmen der Mitgliedschaft im Gießener UAC Fragestellungen im Zusammenhang mit dieser Daten-Zugriffs- und Nutzerordnung.

Die Vertreter überwachen den zweckgebundenen Einsatz der Ressourcen.

Die aktuelle Ernennung der Vertreter der Betreiber des DIZ ist im Dokument „Ernannte Vertreter der Betreiber des Datenintegrationszentrums des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“ festzulegen.

1.2 Use & Access Committee

Die Mitglieder des Gießener Use & Access Committees (UAC) werden im Benehmen zwischen dem Dekanat des Fachbereichs Medizin der JLU und der Geschäftsführung UKGM / Standort Gießen benannt.

Das UAC sollte Mitglieder haben, die sich aus unterschiedlichen Bereichen zusammensetzen. Eine detaillierte Ausführung der Aufstellung, Aufgaben und Pflichten des UAC ist in der entsprechenden Geschäftsordnung („Geschäftsordnung für das Use & Access Committee (UAC) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“) niedergeschrieben.

1.3. Zusammenarbeit im MIRACUM Konsortium und mit der Geschäftsstelle der BMBF MII

Für das MIRACUM Gesamt-Konsortium soll künftig ebenfalls ein UAC eingerichtet werden. Das Steering Board des Konsortiums (Vertreter im Steering Board für das Dekanat des Fachbereichs Medizin der JLU und das UKGM / Standort Gießen sind die Herren Acker und Marquardt) wird den Aufbau des „MIRACUM UAC“ planen.

In Bezug auf eine national einheitliche Nutzungsordnung wird eine zentrale Geschäftsstelle aufgebaut, welche gemäß des NSG Eckpunktepapiers als neutrale Einrichtung eine Begleitstruktur eingerichtet, die administrative Aufgaben bei der Bearbeitung von Konsortium-übergreifenden Nutzungsanträgen übernimmt und hierfür die zentrale Anlaufstelle darstellt. Das Gießener UAC verpflichtet sich zur konstruktiven und aktiven

Zusammenarbeit sowohl mit dem künftigen MIRACUM UAC als auch mit der zentralen Geschäftsstelle der Begleitstruktur.

2. Grundlagen und Zweck der Datennutzung

Die Nutzung/Austausch von Daten erfolgt auf Grundlage einer datenschutzgerechten Einwilligung des Patienten oder auf Basis anderer gesetzlicher Grundlagen, welche sich auf die medizinische Forschung beziehen. Der Zugriff und die Nutzung der Daten bedürfen eines vorgegebenen Antragsverfahrens. Neben den Standorten des Konsortiums können auch öffentliche und private Einrichtungen Anträge zur Datennutzung für Forschungszwecke stellen. Mit einem Antrag verpflichtet sich der Antragsteller gleichzeitig zur Akzeptanz der vorgegebenen Rahmenbedingungen und Regularien des Konsortiums.

3. Publikationen, Patente, Schutzrechte

Publikationen über die Ergebnisse der Datennutzung aus Use Cases oder anderer Forschungsvorhaben, welche auf DIZ-Datennutzung basieren (und das Genehmigungsverfahren zur Datennutzung durchlaufen haben), sind anzustreben.

Das DIZ ist über alle beabsichtigten Publikationen, die auf der Bereitstellung von Daten beruhen zu informieren, ggf. unter Nutzung der dafür etablierten elektronischen Plattform. In Veröffentlichungen, denen von DIZ zur Verfügung gestellte Daten ganz oder teilweise zugrunde liegen, muss ein Hinweis eingefügt werden, dass diese durch DIZ zur Verfügung gestellt wurden. Für Veröffentlichungen und insbesondere für die Nennung von Institutionen und Personen, die an der Generierung oder Aufbereitung der Daten beteiligt waren, und deren Aufführung als Mitautoren, gelten die Regeln der Guten Wissenschaftlichen Praxis der Deutschen Forschungsgemeinschaft (DFG) und die zu treffenden Vereinbarungen im Datennutzungsvertrag.

Der Datenempfänger darf ohne ausdrückliches schriftliches Einverständnis des Datengebers und der JLU und UKGM keine Patente oder andere gewerbliche Schutzrechte, die sich auf die bereitgestellten Daten beziehen oder durch diese begründet werden, anmelden. Die datenerhebenden Institutionen sind an allen aus der Datenbereitstellung erwachsenden Patenten oder sonstigen Schutzrechten zu beteiligen. Art und Umfang der Beteiligung sollen jeweils im Rahmen einer individualvertraglichen Vereinbarung festgelegt werden.

4. Übernahme, Weitergabe, Freigabe und Aufbewahrung von Daten durch DIZ

4.1. Übernahme von Daten

Die Übernahme von Daten durch das DIZ erfolgt getrennt nach Patientenversorgungs- und Forschungsdaten nach Maßgabe lokal definierter Datentransportverfahren.

Patientenversorgungsdaten:

In einem ersten Schritt wird das DIZ als Wissenschaftsarchiv aus Patientenversorgungsdaten des Klinikums gespeist. Diese Daten werden mit Algorithmen und Organisationswerkzeugen zur grundsätzlich wissenschaftlichen Verwendung aufbereitet. Der Umfang und Inhalt der Daten werden einvernehmlich zwischen UKGM und Fachbereich Medizin der Universität Gießen abgestimmt.

Forschungsdaten:

In einer späteren Phase wird das DIZ mit Forschungsdaten angereichert. Die Übernahme von Forschungsdaten der JLU und die Haltung der Daten im DIZ erfolgt auf Antrag und erfordert die Zustimmung des Datengebers. Der Datengeber kann Nutzungseinschränkungen nach Art, Umfang und Dauer der Nutzung und Verwendung der jeweils übergebenen Daten machen, einschl. Einschränkungen der Weitergabe. Diese Rahmenbedingungen zur Überlassung und Nutzung der DIZ-Daten sind elektronisch zwecks Überprüfung zu hinterlegen.

Die datenschutzrechtlichen Voraussetzungen für die Übernahme, Speicherung, Nutzung und Weitergabe der Daten muss dem DIZ vorgelegt werden. Nur solche werden übernommen, bei denen diese Vorgaben erfüllt sind. Datengeber und Antragsteller haben gemeinsam die Erfüllung der datenschutzrechtlichen Auflagen nachzuweisen. Auch diese entsprechenden Unterlagen sind elektronisch zwecks Überprüfung zu hinterlegen.

Grundsätzlich ist anzumerken, dass innerhalb des DIZ jene aus der Abfrage rückgemeldeten Daten für den Anwender nicht mehr personenbeziehbar sind, da sie lediglich ein internes maschinell erzeugtes Schlüsselement enthalten, welches nur innerhalb des für Anfragen aufgebauten und geschützten konsolidierten Datenpools (Versorgungsdaten und Forschungsdaten) einen Patienten identifizieren kann. Somit gibt es aus der Sicht des DIZ für den Anwender keinerlei Möglichkeit eines Rückschlusses auf die Identität einer betreffenden Person.

4.2. Weitergabe von Daten

DIZ darf Daten an Dritte nur herausgeben, wenn folgende Voraussetzungen vorliegen:

- Vorliegen eines Antrags gemäß Abschnitt 5
- Zustimmung des UAC; bei Forschungsdaten gesonderte Freigabe der Daten durch die jeweiligen Datengeber
- Abschluss eines Datennutzungsvertrags mit dem Datenempfänger gemäß Abschnitt 6
- Vertragliche Verpflichtung des Datenempfängers zur Einhaltung der Vorgaben des Datengebers (s.o.)
- DIZ trifft geeignete Sicherheitsvorkehrungen, um die Vertraulichkeit der Daten bei Weitergabe zu gewährleisten (z.B. Anonymisierung, Pseudonymisierung). Genom-Daten sind aufgrund ihrer Struktur und aufgrund ihrer Reidentifizierung durch bestimmte Personenkonstellationen gesondert zu betrachten. Entsprechende Ausführungen dazu werden später erstellt.

4.3. Aufbewahrung von Daten

(dieser Absatz tritt erst in Kraft, nachdem im Laufe der Aufbau- und Vernetzungsphase der BMBF MII die entsprechende Funktionalität zur Forschungsdatenarchivierung am Giessener DIZ etabliert wurde)

Auf Antrag des Datenempfängers übernimmt DIZ gegen angemessene und mindestens selbstkostendeckende Gebühren die Aufbewahrung einer Kopie übergebener Daten im Auftrag des Datenempfängers für die Maximaldauer der für den jeweiligen Verwendungszweck vorgeschriebenen Regularien.

5. Antrag an UAC

Der Antrag an das UAC auf Datenbereitstellung muss diverse Kernaspekte enthalten, welche in der Geschäftsordnung des UAC („Geschäftsordnung für das Use & Access Committee (UAC) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“) detailliert beschrieben sind.

6. Datennutzungsvertrag

Voraussetzung für die Übergabe der Daten nach Genehmigung des Vorhabens ist der Abschluss eines Datennutzungsvertrags.

Der Vertrag beinhaltet eine Beschreibung sämtlicher vom Datengeber erfolgten Auflagen und Nutzungsbedingungen.

Der Datennutzungsvertrag spezifiziert insbesondere:

- i. Projektbeginn und Projekttende
- ii. die dem Datenempfänger für das Forschungsprojekt zur Verfügung gestellten Daten
- iii. die Regelung von Patenten und anderen Schutzrechten gemäß Ziff. 6, Grundsätze der Datennutzung gemäß Ziff. 8, die Pflicht zur Berichterstattung und Information gemäß Ziff. 9 und zur Rückübermittlung der Ergebnisse gemäß Ziff. 10
- iv. den Nutzungszeitraum und den spätesten Zeitpunkt für die Löschung übergebener Daten

Der Vertrag beinhaltet folgende Verpflichtungen des Datenempfängers:

Der Datenempfänger verpflichtet sich in diesem Vertrag,

- (a) die Angaben im Antrag einzuhalten
- (b) sämtliche im Vertrag beschriebenen Bedingungen und Auflagen für die Nutzung und Speicherung der Daten einzuhalten
- (c) die Nutzung der Daten auf den angegebenen Nutzungszeitraum zu beschränken
- (d) technische und organisatorische Maßnahmen zur angemessenen Sicherung der bereitgestellten Daten vor Missbrauch und Verlust zu treffen, die den gesetzlichen Vorgaben, insbesondere dem Datenschutz entsprechen, insbesondere keinen Versuch der Reidentifikation zu unternehmen sowie das UAC über die dafür etablierte elektronische Plattform unverzüglich bei Verdacht auf

Datenschutzverletzungen (z.B. falls eine Reidentifikation möglich ist oder möglich sein könnte) oder andere Unregelmäßigkeiten bei der Verarbeitung der Daten zu unterrichten,

- (e) die Daten nach Ablauf der vom DIZ gesetzten Nutzungsfrist von allen Datenträgern sicher zu löschen und die Löschung DIZ schriftlich zu bestätigen
- (f) die volle Verantwortung für die Verwendung der Daten zu übernehmen,
- (g) insbesondere weitere Personen, denen er Zugang zu den Daten gewährt, schriftlich auf die Einhaltung der Nutzungsbedingungen gemäß Auflagen DIZ verpflichten, und die Zugangsberechtigungen schriftlich zu dokumentieren,
- (h) keine Kopien von Daten anzulegen und keine Daten an Dritte weiterzugeben
- (i) innerhalb eines Jahres nach Projektende dem DIZ einen Abschlussbericht in elektronischer Form über die dafür etablierte elektronische Plattform zu übermitteln. Im Falle einer Datennutzung zur Erstellung einer wissenschaftlichen Publikation genügt die Vorlage des Publikationsmanuskriptes
- (j) die in Abschnitt 3 genannten Publikationsregeln zu beachten.

Der Datenempfänger stimmt vertraglich zu, dass bei Verstößen gegen die Nutzungsordnung dem Datenempfänger die eingeräumte Nutzungserlaubnis ganz oder teilweise entzogen werden kann.

Der Vertrag soll geeignete Regelungen dafür vorsehen, dass wichtige Ergebnisse und abgeleitete Daten aus Forschungsprojekten, die für die weitere Nutzung des DIZ relevant sind, nach Abschluss eines Forschungsprojektes an das DIZ in geeigneter elektronischer Form übermittelt werden.

Im Vertrag ist zu vereinbaren, dass

- JLU oder UKGM oder seine Beschäftigten keine Gewähr für die Korrektheit der Daten und die Eignung der Daten für den beantragten und genehmigten Zweck übernehmen,
- JLU oder UKGM oder seine Beschäftigten nicht für Schäden jeglicher Art haften, die durch das Arbeiten mit den zur Verfügung gestellten Daten entstehen,
- diese Haftungsbegrenzungen nicht für Vorsatz oder grobe Fahrlässigkeit gelten,
- weder JLU noch UKGM außerhalb der Fälle vorsätzlicher Pflichtverletzungen für mittelbare Schäden haften,
- dass die vertraglichen Haftungsbeschränkungen auch für die gesetzliche Haftung vom JLU und UKGM sowie die persönliche Haftung seiner gesetzlichen Vertreter, Mitarbeiter sowie Erfüllungs- und Verrichtungsgehilfen gelten.

Der Vertrag legt fest, dass der Datenempfänger für alle übermittelten Daten verantwortlich und haftbar ist und dass er für alle durch ihn bei der Nutzung des Datenmaterials verursachten Schäden jeglicher Art haftet, insbesondere solche, die durch unberechtigte Nutzung oder Weitergabe von Daten, und / oder Ergebnissen entstehen.

7. Entscheidungsverfahren

Ein detaillierter Ablauf inklusive geltender Rahmenbedingungen ist in der Geschäftsordnung des UAC ersichtlich. („Geschäftsordnung für das Use & Access Committee (UAC) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“)

8. Berichterstattung und Informationspflicht

Der Datenempfänger hat innerhalb eines Jahres nach Projektende dem UAC einen Abschlussbericht, sowie gegebenenfalls entstandene Publikationen in elektronischer Form über die dafür etablierte elektronische Plattform zu übermitteln.

9. Rückübermittlung von Ergebnissen

Wichtige Ergebnisse und abgeleitete Daten aus Forschungsprojekten, die für die weitere Nutzung des DIZ relevant sind, sollen im Regelfall nach Abschluss eines Forschungsprojektes an das DIZ in geeigneter elektronischer Form übermittelt werden. Weiteres wird in den Nutzungsverträgen geregelt.

10. Änderung der Daten-Zugriffs- und Nutzungsordnung

Die Daten-Zugriffs- und Nutzungsordnung des DIZ Gießen behält ihre Gültigkeit bis das Dekanat des Fachbereichs Medizin der Justus-Liebig-Universität-Gießen und die Geschäftsführung des UKGM / Standort Gießen eine Änderung mit einfacher Mehrheit beschließen. Änderungen benötigen stets eine schriftliche Begründung.

11. Inkrafttreten

Diese Daten-Zugriffs- und Nutzungsordnung tritt zum (mit Verabschiedung der Geschäftsordnung des UAC) in Kraft.

13.2 Geschäftsordnung des Use & Access Committee

Geschäftsordnung für das Use & Access Committee (UAC) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen

Präambel

Im Zuge der Medizininformatik-Initiative des Bundesministeriums für Bildung und Forschung (BMBF) wurde zur Etablierung von einheitlichen Datenzugängen und deren Austausch speziell im Hinblick auf die lokalen Datenintegrationszentren sogenannte Use & Access Committees (UAC) etabliert. Zu diesem Zweck wurde eine entsprechende Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen, Stand 14.10.2020, aufgesetzt.

Die vorliegende Geschäftsordnung des UAC umfasst sämtliche Rahmenbedingungen für Einreichung von Datennutzungsanträgen, für stattfindende Sitzungen des UAC bis hin zur finalen Freigabe und eventuellen Widerrufen der bereitzustellenden Daten.

Definitionen

a) Use & Access Committee

Gemäß des Eckpunktepapiers einer einheitlichen Nutzungsordnung ist jedes Konsortium aufgefordert ein oder mehrere Use & Access Committees (UAC) und deren Zuständigkeit für die Datenintegrationszentren (DIZ) festzulegen. Diese UACs führen die fachliche und inhaltliche Prüfung von Anträgen auf Datennutzung durch und involvieren dabei die lokalen Kompetenzen aus den Bereichen Datenschutz, Ethikkommission und Justitiariaten. Sie führen eine Entscheidung zur Nutzungsfreigabe herbei, wobei die Zustimmung identifizierbarer relevanter Datengeber eingeholt wird.

b) Datenempfänger

Als Datenempfänger werden Personen oder Institutionen bezeichnet, die im Rahmen der BMBF MII bereitgestellte Services nutzen, um Daten zu finden und über diese Daten weitere Informationen im Detail zu erhalten. Diese Personen oder Institutionen können

- Teil bzw. wissenschaftliche Mitarbeiter des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des Universitätsklinikum Gießen und Marburg GmbH (Standort Gießen) sein,
- Teil bzw. wissenschaftliche Mitarbeiter eines der geförderten Konsortien sein, oder
- außerhalb eines der teilnehmenden Konsortien angesiedelt sein

In diesem Kontext wird auch auf die Definition für „Consumer“ in OAIS¹ hingewiesen.

¹ CCSDS Reference Model for an Open Archival Information System (OAIS). Recommended Practice. Magenta Book. 2012. Consultative Committee for Space Data Systems: Recommendation for Space Data System Practices, CCSDS 650.0-M-2, <https://public.ccsds.org/pubs/650x0m2.pdf> (Abruf: 2016-11-23).

c) Datengeber

Als Datengeber werden Personen oder Institutionen bezeichnet, die im Rahmen der BMBF MII Daten zur Nutzung im Forschungs- oder Versorgungskontext zur Verfügung stellen. In diesem Zusammenhang wird auch auf die Definition für „Producer“ in OAIS hingewiesen.

d) Antrag

Anträge stellen die inhaltlichen und formalen Rahmenbedingungen eines Forschungsprojekts dar.

e) Transferstelle

Die Transferstelle ist Teil des DIZ und für die technische und ggf. administrative Umsetzung des Data Sharing verantwortlich.

f) Zentrale Antrags- und Registerstelle (ZARS)

Die Antrags- und Registerstelle ist eine zentrale und neutrale Einrichtung der Begleitstruktur, die koordinative und administrative Aufgaben übernimmt. Sie nimmt Anträge von externen Partnern auf Datennutzung an und reicht diese an die zuständigen UACs bzw. Transferstellen der Datenintegrationszentren weiter. Sie führt ein öffentliches, zentrales Register aller Anträge und daraus resultierender Data-Sharing-Projekte und ein Verzeichnis der lokalen Nutzungs- und Verfahrensordnungen.

g) Datennutzungsvereinbarung

Die Datennutzungsvereinbarung regelt alle rechtlichen und administrativen Aspekte für die an einem Data-Sharing-Projekt beteiligten Partner und bezieht sich inhaltlich auf den gemeinsamen Antrag der beteiligten Partner zu einem Datennutzungsprojekt.

h) Datenintegrationszentrum (DIZ)

Das Datenintegrationszentrum am Standort Gießen wird von dem Fachbereich Medizin der JLU in Kooperation mit dem UKGM / Standort Gießen betrieben. Die durch die Betreiber des DIZ ernannten Vertreter sind in der „Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“, Stand 14.10.2020, unter Punkt 1.1 ersichtlich.

1. Use & Access Committee Struktur

Das UAC am Standort Gießen ist strukturell als Verbund von Mitgliedern des Fachbereichs Medizin der Justus-Liebig-Universität und des UKGM / Standort Gießen anzusehen.

2. Mitglieder

Die Mitglieder des Gießener Use & Access Committees (UAC) werden im Benehmen zwischen dem Dekanat des Fachbereichs Medizin der JLU und der Geschäftsführung UKGM / Standort Gießen benannt.

Das UAC soll Mitglieder enthalten, die sich aus den nachfolgenden Bereichen zusammensetzen (Maximal ein Mitglied pro Bereich).

- Klinik
- Vorklinik
- Klinisch-theoretische Institute
- Forschungsprofessuren
- Zahnmedizin
- Institut für Medizinische Informatik
- Dekanat Fachbereich Medizin
- Geschäftsführung UKGM
- Medizininformatik THM
- Ein durch die Betreiber des DIZ ernannten Vertreter (Definition h)

Der Leiter des Datenintegrationszentrums wird in beratender Funktion in Hinblick auf die technische und strukturelle Umsetzung hinzugezogen.

Des Weiteren benennt das UAC eine/n wissenschaftliche/n Mitarbeiter/in des Fachbereichs Medizin der Justus-Liebig-Universität Gießen, welche/r als UAC-Koordinator tätig ist. Der UAC-Koordinator verfügt über kein Stimmrecht und übernimmt administrative Aufgaben.

3. Neubesetzung und Ausstieg

Im 2-Jahresabstand wird die Zusammensetzung des UAC von den in Punkt 2 genannten Gremien überprüft und es kann eine entsprechende Abberufung bzw. Neubesetzung von Mitgliedern des UAC erfolgen. Bei Austritt einer Person aus dem UAC kann auch unterjährig eine Nachbesetzung erfolgen.

4. Publizität der Mitglieder

Die jeweils aktuelle Aufstellung der UAC-Mitglieder ist stets auf der Homepage einer der Betreiber des Datenintegrationszentrums ersichtlich.

5. Beschlüsse

Beschlüsse des UAC erfordern die einfache Mehrheit aller Mitglieder. Sofern ein stimmberechtigtes Mitglied des UAC bei einer Sitzung verhindert sein sollte, ist es dem Mitglied gestattet, sein Stimmrecht für diese Sitzung auf einen selbsternannten Vertreter zu übertragen. Der Vertreter muss eine Angehörigkeit zum gleichen Fachgebiet besitzen. Sofern keine Vertretung bestimmt wurde, entfällt das Mitglied inklusive Stimmrecht aus der Sitzung. Der Mehrheitsentscheid bleibt bestehen.

Das Gießener UAC führt die fachliche und inhaltliche Prüfung von Zugriffs- und Nutzungsanträgen durch, die an das DiZ gerichtet werden. Es entscheidet über die Anträge, wobei die Zustimmung identifizierbarer relevanter Datengeber eingeholt wird. Es stützt sich bei seinen Entscheidungen u. a. auf vorliegende Voten beteiligter Ethikkommissionen (falls diese involviert sind) und begründet ihre Befürwortung bzw. Ablehnung von Zugriffs- und Nutzungsanträgen. Weiter wird die Leitung des Datenintegrationszentrums zur Beratung bezüglich technischer und struktureller Machbarkeit hinzugezogen.

6. Antragseinreichung

Der Antrag an das UAC auf Datenbereitstellung muss mindestens folgende Angaben enthalten

- Antragsteller / Datennutzer
- Verantwortlicher Projektleiter
- Vertragspartner
- weitere Beteiligte (z. B. Kooperationspartner)
- Vorhabentitel
- Vorhabenziel
- Vorhabenbeschreibung
- wissenschaftlicher Hintergrund
- beabsichtigter Zeitraum der Datennutzung
- Arbeitsprogramm
- Einzelheiten zu den beantragten Daten
- Publikationspolicy, Art, Umfang und Autorenschaft der geplanten Publikationen
- Angaben zur Feasibility
- Zustimmende Bewertung der für das geplante Forschungsvorhaben zuständigen Ethikkommission(en)

Sofern das Antragsdokument die oben genannten Punkte enthält kann der Antrag an das UAC gesendet werden und zählt somit als eingereicht. Zu diesem Zwecke existiert eine elektronische Plattform welche auf der Homepage des UAC, siehe Punkt 4, ersichtlich ist.

7. Antragsverwaltung

Zur Verwaltung der Zugangs- und Nutzungsanträge wird dem UAC eine entsprechende elektronische Plattform zur Verfügung gestellt. Die Plattform ermöglicht eine direkte Übersicht zur Unterscheidung von laufenden und abgeschlossenen Anträgen sowie Neueinreichungen. Weiter soll es möglich sein, Kommentare und bereits vorab Voten für entsprechende Dokumente zu hinterlegen. Ziel ist es, diese Plattform interoperabel und transparent mit analogen elektronischen Plattformen auf MIRACUM-Ebene sowie nationaler Ebene zu verknüpfen, so dass mittelfristig entsprechende Antragsverfahren vollständig elektronisch und ohne jegliche Medienbrüche abgewickelt werden können.

8. Entscheidungsverfahren

- a) Nutzungsanträge werden zunächst durch DIZ einer formalen Prüfung unterworfen. Die Entscheidung über den Antrag erfolgt unter Berücksichtigung der am Gießener Universitätsklinikum und im Fachbereich Medizin der JLU vorliegenden Strukturen durch das Use & Access Committee. Nach der Genehmigung oder Ablehnung der Datennutzung für ein beantragtes Forschungsvorhaben wird die Entscheidung den entsprechenden Antragstellern/Datenempfängern bzw. den anfragenden übergeordneten Gremien/Einrichtungen in MIRACUM bzw. auf nationaler Ebene übermittelt. In letzterem Fall leiten diese Gremien/Einrichtungen die Entscheidung an den Antragssteller weiter.
- b) DIZ und UAC gewährleisten, dass Anträge in einer angemessenen Frist bearbeitet werden.
- c) Wie im NSG Eckpunktepapier einer einheitlichen Nutzungsordnung beschrieben, beruht die Datenbereitstellung immer auf Freiwilligkeit und kann im Einzelfall, unter Berücksichtigung der Interessen der datenerhebenden Stellen verweigert werden (Opt-out). Im Falle der Verweigerung der Datenherausgabe ist diese zu begründen.
- d) DIZ führt ein Register, welches alle eingehenden Daten Zugriffs- und Nutzungsanträge, deren jeweiligen Status sowie die abschließenden Entscheidungen einer genehmigten Datennutzung bzw. ggf. Ablehnung dokumentiert.
- e) Fühlt sich das UAC in Ausnahmesituationen nicht im Stande, über einen Antrag zu entscheiden, so wird eine entsprechende Entscheidung zeitnah in Benehmen mit dem Dekanat des Fachbereichs Medizin und dem UKGM getroffen.
- f) Bei sich stark überschneidenden Forschungsvorhaben soll im Rahmen des Antragsverfahrens unter Berücksichtigung der Interessen aller Beteiligten auf eine Kooperation der beantragenden Projektpartner hingewirkt werden.
- g) Das UAC übernimmt eine Aufsichtsfunktion und kann sich jederzeit über den Stand von Datenfreigabe, Datenbearbeitung und Publikation der von ihm betreuten Vorhaben berichten lassen.

9. Berichterstattung und Informationspflicht

Der Datenempfänger hat innerhalb eines Jahres nach Projektende dem UAC einen Abschlussbericht, sowie gegebenenfalls entstandene Publikationen in elektronischer Form über die dafür etablierte elektronische Plattform zu übermitteln.

10. Rückübermittlung von Ergebnissen

Wichtige Ergebnisse und abgeleitete Daten aus Forschungsprojekten, die für die weitere Nutzung des DIZ relevant sind, sollen im Regelfall nach Abschluss eines Forschungsprojektes an das DIZ in geeigneter elektronischer Form übermittelt werden. Weiteres wird in den Nutzungsverträgen geregelt.

11. Anpassung der Geschäftsordnung

Die Geschäftsordnung des UAC Gießen behält ihre Gültigkeit bis die Mitglieder eine Änderung mit einfacher Mehrheit beschließen. Änderungen dürfen von jedem Mitglied vorgeschlagen werden und benötigen stets eine schriftliche Begründung.

12. Salvatorische Klausel

Sollten einzelne Bestimmungen dieser Geschäftsordnung unwirksam oder undurchführbar sein bzw. unwirksam oder undurchführbar werden, so wird dadurch die Wirksamkeit der Geschäftsordnung im Übrigen nicht berührt. Anstelle der unwirksamen oder undurchführbaren Bestimmung soll diejenige wirksame und durchführbare Regelung treten, deren Wirkung der wissenschaftlichen und medizinischen Zielsetzung möglichst nahekommt, welche ursprünglich mit der unwirksamen bzw. undurchführbaren Bestimmung verfolgt wurde. Die vorstehenden Bestimmungen gelten also entsprechend auch für den Fall, dass sich die Geschäftsordnung als lückenhaft erweist.

13. Inkrafttreten

Diese Geschäftsordnung tritt zum (mit Verabschiedung der Daten-Zugriffs- und Nutzungsordnung) in Kraft.

13.3 Geschäftsordnung für das Datenintegrationszentrum

Geschäftsordnung für das Datenintegrationszentrum (DIZ) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen

Präambel

Im Zuge der Medizininformatik-Initiative (MII) des Bundesministeriums für Bildung und Forschung (BMBF) wurden sogenannte Datenintegrationszentren (DIZ) als lokale Kerninfrastrukturen geschaffen. Am Standort Gießen wird dieses DIZ von dem Fachbereich Medizin der Justus-Liebig-Universität Gießen und dem UKGM / Standort Gießen betrieben.

Definitionen

a) Betreiber und deren Vertreter

Die Betreiber des Datenintegrationszentrums (DIZ), der Fachbereich Medizin der Justus-Liebig-Universität Gießen und das UKGM / Standort Gießen ernennen entsprechende Interessenvertreter. Zu den Aufgaben gehört die Vertretung der Betreiber in Bezug auf das DIZ nach innen und außen sowie die Bereitstellung jener Ressourcen (Personal-, Sach- und Investivmittel), welche aus den MIRACUM- bzw. anderweitigen Drittmittel-Fördermitteln oder aus ergänzenden Bereitstellungen des UKGM bzw. des Fachbereichs Medizin resultieren. Weitere Aufgaben und Rahmenbedingungen sind in der „Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“ unter Punkt 1.1 ersichtlich.

c) Use & Access Committee

Gemäß des Eckpunkteapiers einer einheitlichen Nutzungsordnung ist jedes Konsortium aufgefordert ein oder mehrere Use & Access Committees (UAC) und deren Zuständigkeit für die Datenintegrationszentren (DIZ) festzulegen. Diese UACs führen die fachliche und inhaltliche Prüfung von Anträgen auf Datennutzung durch und involvieren dabei die lokalen Kompetenzen aus den Bereichen Datenschutz, Ethikkommission und Justitiariaten. Sie führen eine Entscheidung zur Nutzungsfreigabe herbei, wobei die Zustimmung identifizierbarer relevanter Datengeber eingeholt wird.

d) Konsortium

Das MIRACUM Konsortiums umfasst zehn deutsche Universitätsmedizin-Standorte (Magdeburg, Marburg, Gießen, Frankfurt, Mainz, Mannheim, Erlangen, Greifswald, Dresden und Freiburg). Das Leitthema der Ausschreibung und damit auch des MIRACUM Konsortiums sowie des DIZ ist es, mittels an den Standorten etablierter DIZ, einen verbesserten Datenzugang, eine gemeinsame Datennutzung und (wo nötig auch) Datenaustausch zu ermöglichen.

e) Transferstelle

Die Transferstelle ist Teil des DIZ und für die technische und ggf. administrative Umsetzung des Data Sharing verantwortlich.

1. Struktur

1.1 Leitung

Der Leiter des DIZ steuert und überwacht dessen Arbeit. Er ist gegenüber den im DIZ tätigen Personen in allen Belangen des DIZ inhaltlich weisungsbefugt. Dies gilt unbeschadet disziplinarischer Weisungsbefugnisse, die sich aus den jeweiligen Dienstverhältnissen des Personals des DIZ ergeben.

Der Leiter des DIZ vertritt das DIZ in allen Belangen nach innen und außen. Er führt die Geschäfte des DIZ in Abstimmung mit den Vertretern der Betreiber.

Der DIZ-Leiter wird von den Vertretern der DIZ-Betreiber ernannt. Um die Handlungsfähigkeit der DIZ-Leitung zu gewährleisten existiert für den DIZ-Leiter ein Stellvertreter. Dieser wird ebenfalls von den Vertretern der Betreiber ernannt, jedoch stets in Einvernehmen mit dem DIZ-Leiter.

Die aktuelle Aufstellung der Leitung des DIZ ist in dem Dokument „Leitung des Datenintegrationszentrums (DIZ) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“ ersichtlich.

1.2 Neubesetzung und Ausstieg

Im 2-Jahresabstand wird die Zusammensetzung der DIZ-Leitung von den Vertretern der Betreiber überprüft und es kann eine entsprechende Abberufung bzw. Neubesetzung der Leitung und / oder Stellvertretung erfolgen. Bei Ausscheiden oder kurz- bzw. langfristiger Verhinderung des DIZ-Leiters übernimmt der Stellvertreter kommissarisch die Aufgaben der Leitung des DIZ. Um die volle Handlungsfähigkeit des DIZ zu gewährleisten muss zeitnah nach dem Ausscheiden des DIZ-Leiters ein neuer Leiter von den Vertretern der Betreiber des DIZ ernannt werden. (1.1) Dies gilt ebenfalls für die Position des Stellvertreters. Eine entsprechende Nachbesetzung kann auch unterjährig erfolgen.

1.3 Publizität

Die Leitung des DIZ ist stets auf einer Homepage der Betreiber des DIZ oder des DIZ selbst ersichtlich.

1.4 Transferstelle

Die lokale Transferstelle ist der operative Kern des DIZ. Sie setzt die Beschlüsse des UAC um und bewerkstelligt bzw. veranlasst die darin genehmigte Herausgabe von Daten.

Die lokale Transferstelle berät Antragsteller hinsichtlich Art und Umfang der über das DIZ verfügbaren Daten sowie zu Fragen des Datenschutzes. Die im Zuge der Beratung herausgegebenen Informationen dürfen keine Rückschlüsse auf einzelne Personen erlauben. Art und Umfang der Beratung regeln entsprechende SOPs.

1.5 Forschungsdatenmanagement

Das DIZ dient im Rahmen des Forschungsdatenmanagements als primäre Anlaufstelle für Forscher des Fachbereichs Medizin der Justus-Liebig-Universität Gießen. Neben einer intensiven Beratung bietet das DIZ den Forschern Möglichkeiten zur intelligenten und digitalen Datenspeicherung sowie dem damit verbundenen Datenmanagement, künstlicher Intelligenz, maschinellem Lernen und statistischer Ergebnisaufbereitung.

1.6 Datenintegrationsplattform

Das DIZ stellt Forschern eine Vielzahl an digitalen Werkzeugen zur Verfügung welche innerhalb der Datenintegrationsplattform nutzbar sind. Durch die Nutzung der durch das DIZ bereitgestellten Werkzeuge können Forschungsdaten der Anwender in die Datenlandschaft des Datenintegrationszentrums integriert und im Gesamtkontext genutzt werden. Entsprechende Rahmenbedingungen können ergänzend in dem Dokument „Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen“ eingesehen werden.

Zu der Nutzung der Datenintegrationsplattform bietet das DIZ eine entsprechende Beratung.

2. Nutzung von Daten

2.1 Antrag

Die Einreichung eines Antrags erfolgt anhand definierter Kriterien welche in der Geschäftsordnung für das Use & Access Committee (UAC) des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen ersichtlich sind.

Sofern der Antrag vollständig und sachlich korrekt über die durch das DIZ zur Verfügung gestellte elektronische Plattform eingereicht wurde, werden die Mitglieder des UAC durch die Transferstelle über den Eingang des Antrags informiert.

Anschließend wird der Antrag durch das UAC geprüft.

2.2 Bearbeitung

Nach Prüfung des Antrags durch das UAC wird die Leitung des DIZ zur Beratung in Bezug auf technische und strukturelle Machbarkeit hinzugezogen. Nach positivem Beschluss übernimmt die Transferstelle die weitere Bearbeitung des Antrags. Parallel wird der Antragsteller durch das UAC über den Beschluss informiert. (Geschäftsordnung für das Use & Access Committee (UAC) des UKGM / Standort Gießen und des Fachbereichs Medizin der Justus-Liebig-Universität Gießen, Punkt 8.a.)

2.3 Datenbereitstellung

Die Transferstelle setzt nach positivem Beschluss die entsprechende Anfrage technisch um und stellt die Daten dem Antragsteller zur Verfügung.

2.4 Datennutzung und Ergebnisse

Zu beachten sind die rechtlichen Grundlagen zu Publikationen, Patenten und Schutzrechten welche in der Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen unter Punkt 3 ersichtlich sind.

Des Weiteren gelten die Vorgaben zur Berichterstattung und Informationspflicht sowie die Rückübermittlung von Ergebnissen (Daten-Zugriffs- und Nutzungsordnung für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen, Punkte 8 und 9)

2.5 Langzeitarchivierung

Alle im Rahmen von Punkt 2.4 generierten Ergebnisse unterliegen der am Standort durch das DIZ etablierten Langzeitarchivierung von Forschungsdaten. Das hierzu notwendige Vorgehen ist in einer entsprechenden Standard Operating Procedure (SOP) auf der Homepage des DIZ ersichtlich. Die Dauer der Langzeitarchivierung für Forschungsdaten ist den entsprechenden Datennutzungs- bzw. Datenspeicherungsverträgen zu entnehmen.

2.6 Verfügbarkeit

Das DIZ richtet sich bei der Verfügbarkeit von Daten nach der Verfügbarkeit der innerhalb des GIFONET (Gießener Forschungsnetz) befindlichen Serverinfrastruktur. Prinzipiell gewährt das DIZ keine hundert prozentige Verfügbarkeit der Daten und orientiert sich bei dem zeitlichen Aspekt an den in Punkt 2.5 definierten Zeitraum.

3. Organisatorisches

3.1 Arbeitsabläufe

Die Arbeitsabläufe des DIZ unterliegen den Grundsätzen der guten wissenschaftlichen Praxis. Sie werden in Standard Operating Procedures (SOPs) niedergelegt, die in ihrer jeweils aktuellen Version auf der Homepage des DIZ veröffentlicht werden.

3.2 Personenidentifikation

Daten werden von den DIZ ausschließlich nach UAC-Vorgabe und in anonymisierter oder pseudonymisierter Form herausgegeben. Ist der Empfänger der Daten die föderierte Transferstelle, so erfolgt die Pseudonymisierung unter Nutzung der föderierten Treuhandstelle. Dadurch wird sichergestellt, dass die zugewiesenen Pseudonyme für eine Zusammenführung mit Daten anderer DIZ nutzbar sind.

4. Nutzung und Bereitstellung von Diensten

4.1 Vergütung

Die bereitgestellten Dienste des DIZ und die damit verbundene Aufwandsentschädigung bei deren Nutzung ist stets der aktuellsten Version des Leistungskataloges des DIZ zu entnehmen.

5. Sonstiges

5.1 Anpassung der Geschäftsordnung

Die Geschäftsordnung des DIZ Gießen behält ihre Gültigkeit bis die DIZ-Leitung und die Vertreter der Betreiber des DIZ eine Änderung mit einfacher Mehrheit beschließen. Änderungen benötigen stets eine schriftliche Begründung.

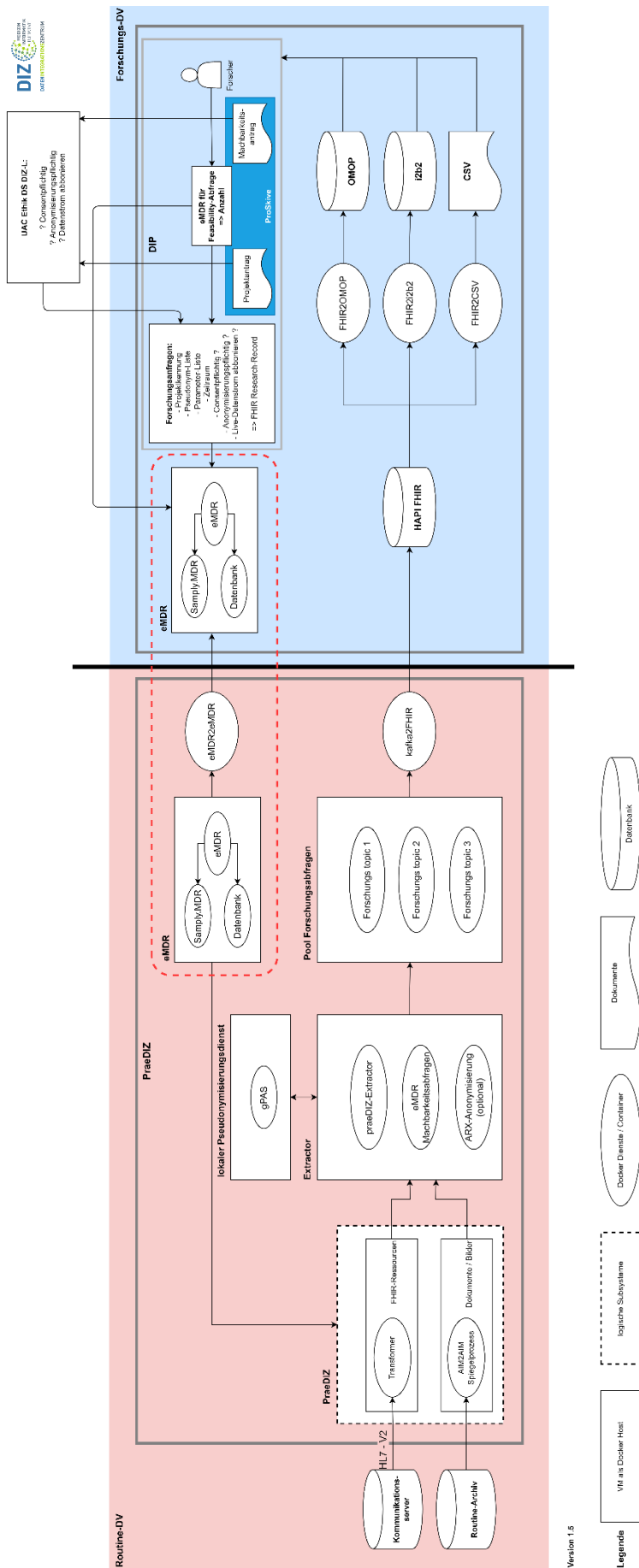
5.2 Salvatorische Klausel

Sollten einzelne Bestimmungen dieser Geschäftsordnung unwirksam oder undurchführbar sein bzw. unwirksam oder undurchführbar werden, so wird dadurch die Wirksamkeit der Geschäftsordnung im Übrigen nicht berührt. Anstelle der unwirksamen oder undurchführbaren Bestimmung soll diejenige wirksame und durchführbare Regelung treten, deren Wirkung der wissenschaftlichen und medizinischen Zielsetzung möglichst nahekommt, welche ursprünglich mit der unwirksamen bzw. undurchführbaren Bestimmung verfolgt wurde. Die vorstehenden Bestimmungen gelten also entsprechend auch für den Fall, dass sich die Geschäftsordnung als lückenhaft erweist.

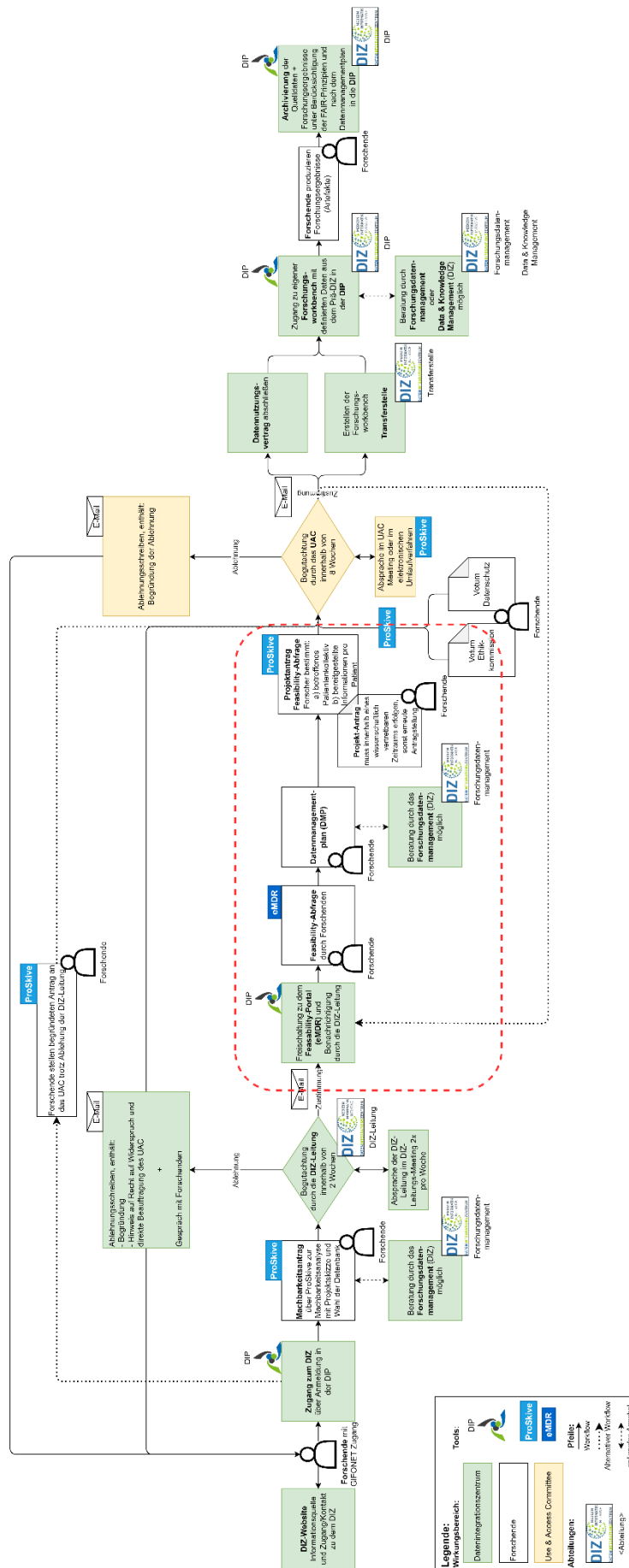
5.3 Inkrafttreten

Diese Geschäftsordnung tritt zum (nach Verabschiedung der Daten-Zugriffs- und Nutzungsordnung sowie der Geschäftsordnung des UAC und der Benennung der Vertreter der Betreiber des DIZ) in Kraft.

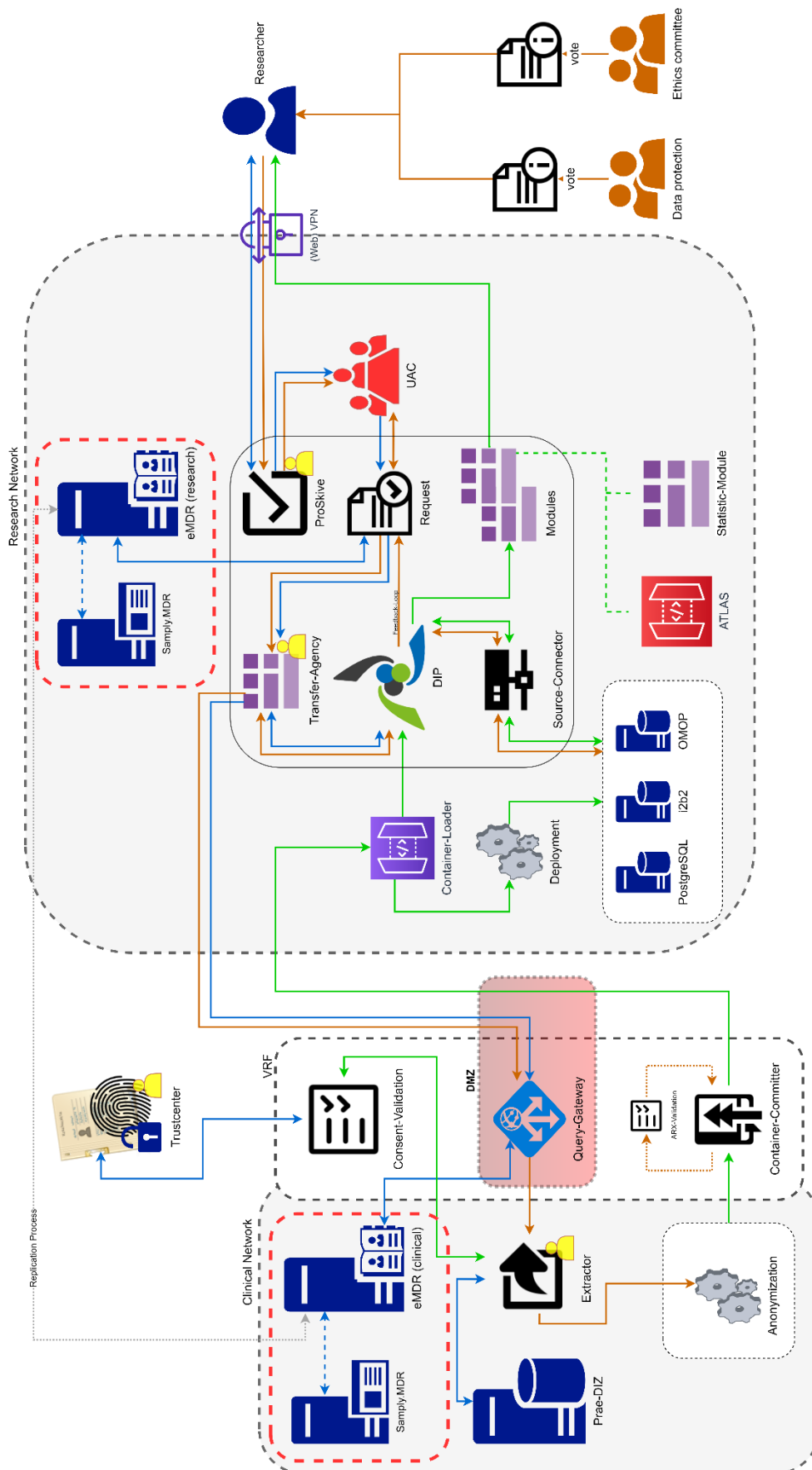
13.4 Technischer Ablauf des Datenintegrationszentrums



13.5 Organisatorischer Ablauf des Datenintegrationszentrums



13.6 Systemschicht des Datenintegrationszentrums



13.7 IT-Infrastruktur für das Datenintegrationszentrum

IT-Infrastruktur für das Datenintegrationszentrum des Fachbereichs Medizin der Justus-Liebig-Universität Gießen und des UKGM / Standort Gießen

	Name, Vorname / Beschreibung	Bereich
Autor	Eberhardt, Florian	DIZ-Leitung
Ziel	Beschreibung der Infrastruktur des DIZ Gießen	
Zielgruppe	Systemadministration DIZ Gießen	
Gültigkeitsbereich	DIZ Gießen	

Historie

Version	Datum	Status	Bearbeiter
V0.1	15.12.2020	erstellt	Eberhardt, Florian
V0.2	17.12.2020	review	Heller, Nick
V1.0	29.04.2021	Freigegeben	Michel-Backofen, Achim

Definitionen

a) Systemadministratoren

Unter der Bezeichnung Systemadministratoren verstehen sich die von den DIZ-Leitern benannten Personen, welche für die Administration der Serverlandschaft des DIZ zuständig sind. Hierbei spielen die Anschaffung, Integration und Verwaltung der Server eine große Rolle, sowie die gesamte Verwaltung der virtualisierten Serverumgebung. Weiter werden alle innerhalb des DIZ benötigten Systemzugänge über die Systemadministratoren beantragt und eingerichtet.

1. Infrastruktur – Physisch

Das DIZ-Gießen verfügt über zwei HP ProLiant DL380 Gen9 Server, welche mittels VMware Enterprise in einem Clusterverbund betrieben werden.

Spezifikationen der Server:

Intel® Xeon® CPU E5-2690 v4 @ 2.30GHz

2 Sockets mit je 14 Cores

RAM gesamt: 512 GB

Als Speicher für das Cluster stehen dem DIZ derzeit 10 TB zur Verfügung, welcher voraussichtlich im Q1 2021 um 100 TB erweitert wird. Der Speicher wird in einem RAID-Verbund betrieben und ist somit gegen Ausfälle geschützt.

Infrastrukturell sind die Server in den Räumlichkeiten der Abteilung für klinische und administrative Datenverarbeitung (AKAD) des UKGM Standort Gießen installiert. Hierbei übernimmt das AKAD die Installation der physischen Server, die Einbindung der Server in das Gießener Forschungsnetz (GIFONET) und die Administration der Firewall des GIFONET, welche direkten Einfluss auf die Server nimmt.

2. Infrastruktur – Virtuell

Das Cluster des DIZ-Gießen ist voll virtualisiert und wird unter der VMware ESXi Version 6.7.0 betrieben. Die VMs werden hierbei von den Administratoren des DIZ angelegt und verwaltet. Die Registrierung der VM innerhalb des GIFONET sowie die Freigabe innerhalb der Firewall wird in enger Kooperation mit dem AKAD durchgeführt.

2.1 Groupware

Zur Unterstützung der digitalen Zusammenarbeit der DIZ-Mitarbeiter wurden diverse Groupware-Dienste aufgesetzt. Hierbei spielen auch Systeme zur Quellcode- und Systemdokumentation eine große Rolle.

MIRgitlab dient der Bereitstellung eines GitLabs zur Quellcodeverwaltung und Dokumentation. Sämtliche innerhalb des DIZ, der Use-Cases und Begleitprojekten entwickelte Komponenten werden dort zentral verwaltet.

MIRredmine stellt ein Redmine bereit um sämtliche Arbeiten innerhalb des DIZ planen und dokumentieren zu können. Hierbei werden Meilensteine, Tickets und ein Wiki für jedes Projekt erstellt und kontinuierlich aktualisiert. Die gesamte Systemadministration ist ebenfalls innerhalb von Redmine abgebildet.

Die VM *MIRgroupware* stellt primär die Chatsoftware Zulip zur Verfügung um einen direkten und schnellen Austausch unter den DIZ-Mitarbeitern zu ermöglichen. Sämtliche Projekte werden in Channel organisiert und können bei Bedarf durchsucht werden. Weiter stellt die VM eine Calibre-Web Instanz bereit um eine zentrale PDF und eBook Datenbank zu gewährleisten.

MIRnextcloud bildet die zentrale Groupware VM ab, mit derer diverse Kalender für die DIZ-Mitarbeiter (caldav), ein Kontaktspeicher (carddav) sowie eine zentrale Cloud-Lösung (webdav) zur Verfügung gestellt wird.

2.2 DIZ-Kerndienste

Das DIZ betreibt derzeit drei VMs welche den direkten Kontakt zu Forscherinnen und Forschern herstellen soll. *MIRACUM* stellt eine standortinterne Informationsseite zum MIRACUM-Konsortium bereit, inklusive der standortspezifischen Projekte und Lösungen. Passend zu dieser Website existiert die VM *DIZ*, welche die Außendarstellung des Gießener DIZ bereitstellt. Hierbei werden Informationen über die Mitarbeiter des DIZ, die Geschäftsordnungen, Projekte und das Antragsverfahren zur Datennutzung übersichtlich und leicht verständlich bereitgestellt.

Die Datenintegrationsplattform (DIP) stellt zukünftig den zentralen Einstiegspunkt in das DIZ dar. Dabei wird die DIP abstrakt als eine Art Betriebssystem des DIZ angesehen. Die entsprechende VM *DIP* dient der Bereitstellung der Plattform.

2.3 Use-Cases

Sämtliche Use-Cases werden stets mit separaten virtuellen Maschinen versorgt, welche durch die entsprechend verantwortlichen Personen verwaltet werden. Beispielsweise MIRmosaic, MIRrecruit, MIRsoda für Use Case 1 und MIRdatashield, MIRdistcomp, MIRnlp für den Use Case 2

2.4 Datenintegration

Für sämtliche Aspekte der Datenintegration werden ebenfalls separate VMs bereitgestellt wie MIRkafka, MIRretl, MIRi2b2 und MIRomop.

2.5 Sonstige

Weiter verfügt das DIZ über diverse Test-VMs und VMs welche nicht zu den oben genannten Bereichen zugeordnet werden können. Dies betrifft beispielsweise die virtuellen Maschinen MIRdocker (FAS), MIRauth (interne Authentifizierung), MIRproskive (PPM) und MIRproskivedev (Testumgebung für das PPM-Deployment)

3. Absicherung der Infrastruktur

Die Infrastruktur ist physisch durch Sicherheitsmechanismen des AKAD abgesichert, da die Server sich in den Räumlichkeiten des GIFONET befinden. Der Zugriff auf den vSphere zur Administration des Clusters kann nur kabelgebunden über das GIFONET, unter Verwendung separater Administratoren-Zugänge, welche durch das Active-Directory des AKAD verwaltet werden, erfolgen. Hierbei wird ein Zugriff zu einem Citrix-Gateway benötigt, welches ebenfalls durch separate Zugänge abgesichert wird. Weiter spielt die von dem AKAD betriebene Firewall eine große Rolle, welche keinen Zugriff von außerhalb des GIFONET auf die VMs zulässt. Eine Freischaltung von benötigten Ports nach Außen erfolgt stets in enger Zusammenarbeit mit Mitarbeitern des AKAD. Die Authentifizierung auf den virtuellen Maschinen ist in den nachfolgenden Unterkapiteln beschrieben.

3.1 Authentifizierung

Die Authentifizierung an Systemen erfolgt aktuell über Test-Accounts, welche je nach Bedarf angelegt wurden. Im nächsten Schritt wird zum produktiven Einsatz der Systeme ein lokales LDAP für das DIZ aufgesetzt und an das im Zuge der MIRACOLIX-Tool Etablierung eingesetzte

Keycloak angebunden. Serverseitig (VM) erfolgt der Zugriff aktuell per SSH, welches nur kabelgebunden aus dem GIFONET möglich ist. Die Zugänge für die Benutzer auf den VMs werden durch die Systemadministratoren angelegt und müssen aus Sicht des Passwortes von den Benutzern geändert werden. Hierbei liegt die lokale SOP zur Authentifizierung zu Grunde welche Vorgaben zur Passwortgenerierung definiert.

Im nächsten Schritt werden sämtliche VMs auf eine reine SSH-Schlüssel Authentifizierung umgestellt. Der Vorteil hierbei ist der erhöhte Grad an Sicherheit und die Vereinheitlichung von Zugangsdaten. Somit können zukünftig sehr flexibel Benutzerzugänge auf den Systemen hinzugefügt und entfernt werden, indem die Schlüssel innerhalb des SSH-Hostspeichers angepasst werden.

3.2 Administrativer Zugriff

Der Zugriff auf eine VM erfolgt wie bei jedem Benutzer über ein Passwort welches nach definierten Richtlinien gesetzt wurde. Die Passwörter für administrativen Zugriff (sudo-User) werden automatisch mittels KeePass 2 generiert und in einer zentralen Datenbank abgelegt. Zugriff auf diese Datenbank haben ausschließlich die benannten Systemadministratoren. Zusätzlich ist aus Sicherheitsgründen auf allen Systemen der root-Benutzer deaktiviert.

Zukünftig wird auch an dieser Stelle auf eine SSH-Schlüssel-Authentifizierung zurückgegriffen. Weiter wird eine Datenbank zur Verwaltung der öffentlichen SSH-Schlüssel aller Mitarbeiter gepflegt, wodurch den Systemadministratoren ermöglicht wird schnell und flexibel Einträge und somit Zugänge einzurichten bzw. zu entfernen.

Die benannten Systemadministratoren sind in dem Ernennungsdokument („DIZ-Ernennung_Systemadministration“) benannt.

4. Überwachung und Verwaltung der Infrastruktur

Ein wichtiger Aspekt ist die Verwaltung der VMs und das damit einhergehende Monitoring. Aufgrund der unterschiedlichen Betriebssysteme, jedoch der gleichen zu Grunde liegenden Basis Linux, wird zukünftig jede VM mittels dem Tool Cockpit an eine zentrale Monitoring-VM in Form eines Knotens angebunden. Zugriff auf die zentrale VM erhalten lediglich die benannten Systemadministratoren. Mittels Cockpit lässt sich die Auslastung eines jeden angebotenen Servers überwachen. Weiter bietet die Software die Möglichkeit Log-Dateien einzusehen und somit Fehler auf den Systemen schnellstmöglich zu identifizieren.

4.1 Serveraktualität

Derzeit werden in regelmäßigen und definierten Zeiträumen sämtliche Systeme mittels Updates versorgt. Hierbei ist zu beachten, dass alle systemrelevanten Sicherheitsupdates automatisch über die Distributions-Repositoryn bezogen und installiert werden. Die nächsten Schritte sehen vor, dass sämtliche Systeme mittels nullmailer und apticron automatisiert E-Mails an die Systemadministratoren versenden, sobald Updates auf den Systemen verfügbar sind. Dadurch wird ermöglicht, dass bei Bedarf bestimmte Updates unmittelbar ausgelassen oder durchgeführt werden können. Zusätzlich dazu wird nach erfolgreicher Umstellung der Server auf die SSH-Schlüssel-Authentifizierung eine Cluster Shell eingerichtet, was das Updaten aller VMs mittels einer SSH-Konsole ermöglicht. Diese Prozesse können ebenfalls automatisiert und mittels Cron-Jobs in definierten Abständen durchgeführt werden. Aus

Sicherheitsgründen und um Inkompatibilitäten auf den Systemen durch Updates zu vermeiden wird deshalb keine Automatisierung angestrebt, wodurch stets eine manuelle Prüfung und Durchführung der Updates erfolgen muss.

4.2 Serversicherungen

Im DIZ Gießen werden gezielte VMs täglich inkrementell gesichert und die Sicherungen bis zu einem Jahr vorgehalten. Die zu sichernden VMs beziehen sich auf die Kern-VMs des DIZ wie beispielsweise alle Groupware-VMs, ETL-VMs, das DIP und alle weiteren Datenrepositorien. Die Backups werden auf einem, vom DIZ-VM-Speicher unabhängigen, RAID-Verbund gesichert um eine höhere Datensicherheit zu gewährleisten.

13.8 Datenschutzkonzept DIZ Gießen



Datenschutzkonzept DIZ Gießen

Datenschutzkonzept DIZ Gießen

1.0, 14.07.2022
Dr. Achim Michel-Backofen

Änderungshistorie

Version	Datum	Bearbeitungsart/Betroffene Abschnitte	Bearbeiter
0.1	11.10.2021	Erster Entwurf	Dr. Achim Michel-Backofen
0.2	27.10.2021	Anpassungen	Dr. Achim Michel-Backofen
0.3	20.05.2022	Anpassungen	Dr. Achim Michel-Backofen
0.3	14.07.2022	Review	Dr. Achim Michel-Backofen
1.0	14.07.2022	Freigabe	Dr. Achim Michel-Backofen

Inhaltsverzeichnis

1. Übersicht und Grundlagen	4
2. Verantwortlichkeiten und Organisationsstruktur	7
3. Zweck der Verarbeitung und rechtliche Grundlage	7
4. Kreis der Betroffenen	7
5. Daten	7
6. Systeme und Prozesse	8
7. Personengruppen mit Zugriff auf die Daten	9
8. Wahrung von Betroffenenrechten	9
8.1 Recht auf Information der Betroffenen gemäß (Art. 13 und 14 EU-DSGVO)	9
8.2 Recht auf Berichtigung (Art. 16 EU-DSGVO)	9
8.3 Recht auf Löschung (Art. 17 EU-DSGVO)	10
8.4 Recht auf Auskunft (Art. 15 EU-DSGVO)	10
8.5 Recht auf Einschränkung der Verarbeitung / „Sperrung“ (Art. 18 EU-DSGVO)	10
8.6 Recht auf Datenübertragbarkeit (Art. 20 EU-DSGVO)	10
8.7 Recht auf Widerruf (Art. 7 Abs. 3 EU-DSGVO)	10
8.8 Widerspruchsrecht (Art. 21 EU-DSGVO)	11
9. Technische und organisatorische Maßnahmen	11
9.1 Zugangskontrolle	11
9.2 Datenträgerkontrolle	11
9.3 Speicherkontrolle	11
9.4 Benutzerkontrolle	12
9.5 Zugriffskontrolle	12
9.6 Übertragungskontrolle	12
9.7 Eingabekontrolle	12
9.8 Transportkontrolle	12
9.9 Wiederherstellbarkeit	12
9.10 Zuverlässigkeit	13
9.11 Datenintegrität	13
9.12 Auftragskontrolle	13
9.13 Verfügbarkeitskontrolle	13
9.14 Trennbarkeit	13
10. Sensibilisierung / Schulungskonzept	13
11. Datenschutzvorfälle und Anfragen	13
Abkürzungs- und Fremdwörterverzeichnis	14

Referenzen	16
Mitgeltende Unterlagen	17

Die technische Infrastruktur des DIZ Gießen besteht aus zwei logisch und physikalisch getrennten Bereichen:

- Dem sogenannten Synedra-praeDIZ, das die routinemäßig anfallenden Daten aus den klinischen Informationssystemen zusammenführt, in FHIR (Fast Healthcare Interoperability Resources) Datenobjekte umwandelt und mittels normierter Klassifikationen (LOINC, SNOMED) semantisch anreichert. Das Synedra-praeDIZ ist technisch ein integraler Teil der Routine-DV des UKGM und für Forscher nicht zugänglich.
- Einem zentralen Extraktionsprozess der jede Forschungsanfrage als separaten Auftrag verarbeitet und nur die in der Forschungsanfrage festgelegten Datenelemente der festgelegten Patientenkohorte extrahiert und in pseudonymisierter/anonymisierter Form der Datenintegrationsplattform (DIP) des DIZ zur Verfügung stellt. Der Extraktionsprozess ist der einzige berechtigte Nutzer des Synedra-praeDIZ.
- Der Datenintegrationsplattform des DIZ die getrennt für jede Forschungsanfrage die in der Machbarkeitsanfrage definierten und vom Extraktionsprozess gelieferten Forschungsdaten rollenbasiert verwaltet und in pseudonymisierter/anonymisierter Form dem Forscher zur Verfügung stellt.

Diese Architektur folgt dem Prinzip des „Data Protection by Design and Default“.

Die Pseudonymisierung/Anonymisierung der medizinischen Daten und die Trennung von identifizierenden Daten (IDAT) und medizinischen Daten (MDAT) findet mit Hilfe des Datenextraktionsprozesses statt. Dieser fragt die ebenfalls im Schutzbereich der Routine-DV laufende Instanz des Pseudonymisierungsdienstes ab und ist der einzige Teil der Architektur der Zugriff auf den Pseudonymisierungsdienst hat. Organisatorisch läuft der Pseudonymisierungsdienst im Verantwortungsbereich der lokalen Treuhandstelle.

Für jede Forschungsanfrage wird in der Datenextraktionsplattform eine für den Forscher geeignete Werkzeugumgebung, die sogenannte Forschungsworkbench aufgebaut. Diese enthält immer einen FHIR-Server als Datendrehscheibe und wahlweise die vom Forscher im Datennutzungsvertrag genannten weiteren Analysewerkzeuge (zum Beispiel i2b2 oder OMOP). Die Kopplung des Datenextraktionsprozesses mit der Forschungsworkbench findet direkt über den dort integrierten FHIR-Server statt. Die Fähigkeit eines Datenexports per HL7-FHIR-Ressourcen ist ein Kernbestandteil der Architektur der Medizininformatik Initiative (MII) und somit auch des DIZ-Gießen [8]. Über diesen wird die Forschungsworkbench mit genau den Daten versorgt, die im Rahmen der Machbarkeits-Abfrage definiert wurden.

Das Datenschutzkonzept von MIRACUM bringt die Auswertung zu den Standorten und belässt die pseudonymisierten Daten im Standort. Dabei werden folgende Kerndatenschutzprinzipien umgesetzt:

- **Privacy-by-Design und Privacy-by-Default**
Alle Daten sind pseudonymisiert
- **Datenminimierung**
Es werden nur die für eine Forschungsanfrage definierten Parameter und nur die in der zur Forschungsanfrage gehörenden Kohorte definierten Patienten übertragen

Die Pseudonymisierung umfasst folgende Schritte:

- **Erzeugen eines Master-Patient-Identifiers mit Hilfe von E-PIX (Master-Patient-Index)**
Hierbei wird die klinikumsinterne Patienten-ID mit einer neu erzeugten Master-Patient-ID vereinigt [9]
- **Erzeugen des forschungsanfragenspezifischen Pseudonyms bei der gPAS Instanz der lokalen Treuhandstelle**
Mit Hilfe der Master-Patienten-ID kann bei der lokalen Treuhandstelle das forschungsanfragenspezifische Pseudonym erzeugt/abgefragt werden. Dieses wird für die weitere Datenaufbereitung in der internen gPAS pseudonym Datenbank gespeichert [10].

Die lokalen technischen Komponenten (E-PIX, gICS und gPAS) dieser Pseudonymisierungsstrecke sind integraler Teil der Synedra-praeDIZ Architektur und daher im selben Routine-Netzwerksegment wie das Synedra-praeDIZ untergebracht und von außerhalb nicht zugänglich.

Der FHIR Server der Forschungsworkbench enthält wie beschrieben nur die für die Forschungsanfrage definierte Patientenkohorte in pseudonymisierter Form mit den auf die Forschungsanfrage reduzierten Daten. Daher darf der forschungsanfragenspezifische Datenpool im Gifonet-Bereich des DIZ unter dem Management der Datenintegrationsplattform (DIP) stehen.

Abbildung 2 zeigt schematisch das Zusammenwirken aller eben genannten Komponenten.

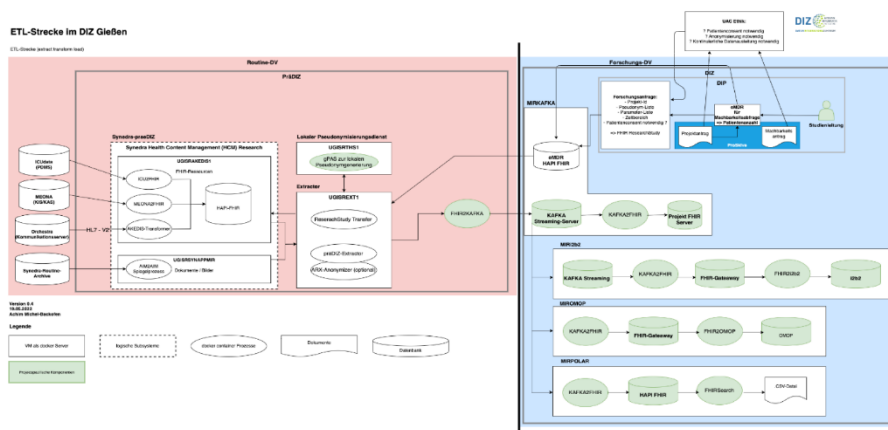


Abbildung 2: ETL-Architektur des DIZ Gießen, [11], 2022.

2. Verantwortlichkeiten und Organisationsstruktur

Die externe organisatorische Einbindung des DIZ Gießen in das UKGM Standort Gießen und den Fachbereich Medizin ist in „DIZ-Organigramm-extern“ [12] beschrieben. Die interne organisatorische Struktur des DIZ Giessen ist in „Organigramm-intern“ [13] beschrieben. Die prinzipiellen Leistungen des DIZ Gießen aufgrund der Forschungsanfragen unterstützt werden sind im DIZ-Leistungskatalog [14] festgelegt.

Der allgemeine Geschäftsbetrieb des DIZ Gießen ist in der „DIZ Geschäftsordnung“ [15] geregelt.

Die konkreten Personen sind in den Dokumenten „10.6.DIZ-Leitung“ [16] und „0.5_Personalstamm-DIZ-Gi“ [17] benannt.

Für den Betrieb des Synedra-praeDIZ ist die Transferstelle im DIZ Gießen verantwortlich.

Der verantwortliche Mitarbeiter für die Transferstelle ist derzeit:

Dr. Jochen Zohner
Datenintegrationszentrum/Transferstelle
Institut für Medizinische Informatik
Heinrich Buff Ring 6
35392 Gießen
Jochen.Zohner@akad.med.uni-giessen.de

Der Mitarbeiter der Transferstelle überwacht die Funktionalität des Synedra-praeDIZ auf Ebene des Datentransfers.

Der technische Betrieb und das Monitoring der technischen Infrastruktur des praeDIZ erfolgt durch Mitarbeiter der Abteilung für klinische und administrative Datenverarbeitung (AKAD) in deren abgesicherten und zugangskontrollierten Räumlichkeiten.

3. Zweck der Verarbeitung und rechtliche Grundlage

Das Synedra-praeDIZ Gießen und die Datenintegrationsplattform sind Teil des Datenintegrationszentrums Gießen. [1]. Die Teilnahme von Fachbereich Medizin und UKGM Standort Gießen ist in der Datenzugriffs- und Nutzungsordnung [1] sowie in der Geschäftsordnung [4] des Datenintegrationszentrums geregelt. Diese beschreiben in allgemeiner Form auch den Zweck der Verarbeitung im Rahmen der Aufgaben des Datenintegrationszentrums.

4. Kreis der Betroffenen

Der Kreis der betroffenen Personen betrifft grundsätzlich alle Patienten des UKGM Standort Gießen [1]. Diese Patienten stehen alle für eine Machbarkeitsabfrage als erste Stufe einer Forschungsanfrage zur Verfügung.

5. Daten

Prinzipiell stehen die Daten der folgenden primären Datenquellen des UKGM Standort Gießen für eine Machbarkeitsabfrage als erste Stufe einer Forschungsanfrage zur Verfügung.

- Synedra Universalarchiv
- MEONA KIS/KAS
- ICUdata PDMS

- SWISSLAB LIS

6. Systeme und Prozesse

eingesetzte IT-Komponenten und Anwendungen

- **Synedra-praeDIZ**
Das Synedra-praeDIZ enthält den nicht pseudonymisierten Spiegel der klinischen Routedaten der strategisch wichtigen klinischen Systeme (MEONA, Synedra-Universalarchiv, ICUdata und Swisslab). Hierzu empfängt es Daten durch direkte Spiegelung über den **AIM2AIM-Prozess** aus dem Synedra-Universalarchiv sowie mittels des **AKEDIS-FHIR-Hub** über den Kommunikationsserver Orchestra. Im Synedra-praeDIZ sind sowohl IDAT als auch MDAT in Form von FHIR-Objekten gespeichert. Hierbei werden Dokumente in der Synedra-Datenbank und FHIR-Ressourcen der strukturierten medizinischen Daten sowie der Patientenbewegungen in einer integrierten HAPI-FHIR-Instanz gespeichert.
- **Master-Patient-Index (MPI)**
Der Master-Patient-Index in Form der Software E-PIX verwaltet die Zusammenführung heterogener Patienten- und Fall-ID's zu einer gemeinsamen Master-Patient-ID. Diese wird dann im weiteren Verlauf der Pseudonymisierung als Ausgangspunkt verwendet. [9]
- **Pseudonymverwaltung**
Die Pseudonymverwaltung ist in Form der Software gPAS realisiert und wird zur Erzeugung forschungsvorhabenspezifischer Pseudonymkreise verwendet. Der praeDIZ-Extraktor arbeitet forschungsvorhabenspezifisch und versieht jeden Forschungsvorhaben-Datenpool mit Pseudonymen eines eigenen Pseudonymkreises. [10]
- **Consentverwaltung**
Die Consentverwaltung ist in Form der Software gICS realisiert und wird zur Konfiguration, Bereitstellung und Verwaltung der Module des eingesetzten Broad Consent verwendet. [19]
- **Dispatcher**
Der Dispatcher stellt die REST-Schnittstelle zu den Komponenten E-PIX, gPAS und gICS bereit.
- **praeDIZ-Extraktor**
der praeDIZ-Extraktor dient zur Überleitung pseudonymisierter Daten an den forschungsanfragenspezifischen Teil der DIZ Architektur. Er beinhaltet folgende Arbeitsschritte
 - Umwandeln der Pseudonyme der angefragten Projektkohorte in die zugehörige Master-Patient-ID mit Hilfe einer gPAS Abfrage
 - Umwandeln der Master-Patient-ID in die im Synedra-praeDIZ verwendete Klinikums-Patienten-ID des Patienten mit Hilfe einer E-PIX Abfrage
 - Extraktion der benötigten Daten eines Patienten mit Hilfe der im Synedra-praeDIZ verwendeten Patienten-ID und der forschungsanfragenspezifischen Konzeptliste (LOINC-, SNOMED-Codes)
 - Ersetzen aller identifizierenden Patientendaten (IDAT) durch das forschungsanfragenspezifische Pseudonym
 - Senden der Daten an die für die Forschungsanfrage zuständige FHIR-Serverinstanz

Die eben aufgelisteten Komponenten laufen unter der Docker-Umgebung folgender virtueller Maschinen:

- UGIRSYPNAPP10 10.129.35.18
Auf dieser VM läuft die Dokumentenarchitektur des Synedra-praeDIZ
- UGIRAKEDIS1 10.129.33.117

- Auf dieser VM läuft der AKEDIS-FHIR-HUB mit dem zugehörigen HAPI-FHIR-Server
- UGISRTHS1 (IP noch nicht benannt)
 - Auf dieser VM laufen die Module die E-PIX, gPAS, gICS und der Dispatcher
- UGISREXT1 (IP noch nicht benannt)

Auf dieser VM läuft der praeDIZ-Extraktor

7. Personengruppen mit Zugriff auf die Daten

Die kritischste Stelle der Architektur des DIZ Gießen ist das Synedra-praeDIZ. Dieses enthält den nicht pseudonymisierten Spiegel der klinischen Routedaten der strategisch wichtigen klinischen Systeme.

Im DIZ Gießen besitzt nur der Mitarbeiter der Transferstelle Zugriff auf die Daten des Synedra-praeDIZ.

Dieser ist für die höheren Ebenen des technischen Transfers der Daten (vor allem Datenintegrität und Qualitätsprüfungen verantwortlich). Diese finden mit dem pseudonymisierten Output der Datenextraktion statt. Die verwendeten Qualitätssicherungsalgorithmen und -werkzeuge sind so gestaltet, dass sie die Identität der Patienten nicht offenbaren.

Das Synedra-praeDIZ als Quelle der medizinischen Daten hat nur einen technischen Benutzer, den praeDIZ Extraktorprozess der Daten anhand einer Patientenliste und Datenelementliste aus dem Synedra-praeDIZ entnimmt und in pseudonymisierter Form den projektspezifischen Weiterverarbeitungsketten zur Verfügung stellt.

8. Wahrung von Betroffenenrechten

Bei Forschungsprojekten, die die Einholung eines Patientenconsent erfordern (z. B.: Broad Consent der MII) [19], wird dieser extern in einer Treuhandstelle verwaltet. Diese erzeugt und verwaltet auch die projektspezifischen Pseudonyme. Der Patient kann im Datenintegrationszentrum – Gießen (DIZ) eine Anfrage stellen und erfahren in welchen Forschungsprojekten seine Daten verwendet werden. Die Anfrage kann telefonisch oder per E-Mail gestellt werden. Beide Informationen sind auf der über das Internet frei zugänglichen Webseite des DIZ-Gießen (<https://diz.med.uni-giessen.de/>) einzusehen. Hier können die Betroffenen auch viele weitere Informationen zum Ablauf von datengestützten Forschungsprojekten und zum Aufbau des DIZ-Gießen und seiner organisatorischen Verankerung im Universitätsklinikum Gießen bekommen.

8.1 Recht auf Information der Betroffenen gemäß (Art. 13 und 14 EU-DSGVO)

Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Datenelemente eines Betroffenen die in einem Forschungsprojekt benutzt werden sind in der Definition des Forschungsprojektes eindeutig maschinell hinterlegt. Daher kann den Betroffenen eine detaillierte Liste aller verwendeten Daten für jedes Forschungsprojekt zugänglich gemacht werden.

8.2 Recht auf Berichtigung (Art. 16 EU-DSGVO)

Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher ist eine Korrektur von fehlerhaften Daten jederzeit durch Änderungen im klinischen Produktivsystem möglich. Nach Freigabe des Datenschützers kann der Transferstellenmitarbeiter diese Protokolle zur Verfügung stellen.

8.3 Recht auf Löschung (Art. 17 EU-DSGVO)

Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher kann jedes Datenelement aus einem Forschungsprojekt wieder entfernt werden. Natürlich können vom Forscher produzierte und veröffentlichte Ergebnisse hierdurch nicht ungeschehen gemacht werden.

8.4 Recht auf Auskunft (Art. 15 EU-DSGVO)

Für jedes Forschungsprojekt wird protokolliert, welche Datenelemente aus dem Katalog aller verfügbaren Datenelementen in dem Forschungsprojekt verwendet werden. Jedes aus dem Synedra-praeDIZ exportierte Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher kann von der Transferstelle über jedes in ein Forschungsprojekt exportierte Datenelement exakt Auskunft gegeben werden.

8.5 Recht auf Einschränkung der Verarbeitung / „Sperrung“ (Art. 18 EU-DSGVO)

Für jedes Forschungsprojekt wird protokolliert, welche Datenelemente aus dem Katalog aller verfügbaren Datenelementen in dem Forschungsprojekt verwendet werden. Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher ist auch der Export von Daten in ein Forschungsprojekt auf atomarer Ebene beschränkbar. Dies findet schon bei Definition und Genehmigung von Forschungsanfragen durch das UAC Anwendung.

8.6 Recht auf Datenübertragbarkeit (Art. 20 EU-DSGVO)

Für jedes Forschungsprojekt wird protokolliert, welche Datenelemente aus dem Katalog aller verfügbaren Datenelementen in dem Forschungsprojekt verwendet werden. Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher kann auf Anfrage eines Betroffenen durch die Transferstelle eine komplette Auflistung aller in Forschungsprojekten herausgegebenen Informationen eines Betroffenen erstellt werden.

8.7 Recht auf Widerruf (Art. 7 Abs. 3 EU-DSGVO)

Die Transferstelle des DIZ Gießen nimmt Anfragen zum Widerruf der Herausgabe von Daten eines Betroffenen entgegen. Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher kann ein Widerruf für jedes Forschungsprojekt durch Löschung der herausgegebenen Daten umgesetzt werden. Es ist aber natürlich unmöglich bereits veröffentlichte Forschungsergebnisse ungeschehen zu machen.

Um ein bereits gegebenen Consent für ein Forschungsvorhaben zurückzuziehen, muss sich der Betroffene selbst an die Treuhandstelle wenden. Deren Kontakt ist auf DIZ Webseite einsehbar

(<https://diz.med.uni-giessen.de/>). Wenn ein solcher Consent von der Treuhandstelle zurückgezogen wird, werden automatisch alle weiteren Datenexporte für diesen Betroffenen im DIZ Gießen gesperrt.

8.8 Widerspruchsrecht (Art. 21 EU-DSGVO)

Für jedes Forschungsprojekt wird protokolliert, welche Datenelemente aus dem Katalog aller verfügbaren Datenelementen in dem Forschungsprojekt verwendet werden. Jedes aus dem Synedra-praeDIZ exportierbare Datenelement besitzt eine eindeutige ID, die es erlaubt sein Lebenszyklus von der Quelle im klinischen Produktivsystem bis zum Export in eine Forschungsdatenbank exakt nachzuverfolgen. Alle Änderungen an einem Datenelement sind versioniert und werden lückenlos protokolliert. Daher kann auch jedem Datenelement widersprochen werden und eine Korrektur oder Löschung bis in die Zieldatenbanken der Forschungsvorhaben umgesetzt werden. Selbstverständlich können bereits veröffentlichte Forschungsergebnisse nicht ungeschehen gemacht werden. Der Betroffene kann aber einsehen in welchen Forschungsvorhaben seine Daten konkret verwendet wurden.

9. Technische und organisatorische Maßnahmen

9.1 Zugangskontrolle

Sowohl die technische Infrastruktur des Synedra-praeDIZ als auch des DIZ werden in der geschützten Umgebung des AKAD ausgeführt. Die technische Infrastruktur ist komplett redundant ausgelegt. Alle Rechnerräume sind mit einem elektronischen Zugangskontrollsystem ausgestattet. Nur Mitarbeiter der Arbeitsgruppe Systemtechnik im AKAD verfügen über eine Zugangsberechtigung zu den Räumen.

Alle Räume des DIZ Gießen sind mit einer elektronischen Schließanlage ausgestattet. Die Mitarbeiter des DIZ-Gießen besitzen nur Zugang zu der Etage und zu Ihrem eigenen Büro. Der Zugang zu den Applikationen des DIZ ist für jeden Mitarbeiter rollenbezogen über ein Active Directory System gesteuert, das nur von der Administrationsgruppe des DIZ [20] verwaltet wird. Des Weiteren sind die Bestandteile der DIZ-Infrastruktur nur über spezielle Webprogramme erreichbar, deren Zugang rollenbasiert über die DIP gesteuert wird. Diese Rechte werden ebenfalls von der Administrationsgruppe des DIZ verwaltet. Alle Administrationsschritte sind in SOPs festgelegt und die erteilten Rechte sind im Antragsformular dokumentiert, welche durch die DIZ-Leitung abgezeichnet werden.

9.2 Datenträgerkontrolle

Daten zu Forschungsanfragen werden nur auf den Servern der Synedra-praeDIZ- und DIZ-Infrastruktur abgelegt. Es gibt keine auf Forschungsanfragen bezogene Daten auf den Workstations der Mitarbeiter. Das Kopieren von Daten zu Forschungsanfragen auf mobile Datenträger ist per SOP untersagt. Die Speicherinfrastruktur des Synedra-praeDIZ und des DIZ sind redundant und hoch verfügbar ausgelegt. Die Daten von Forschungsanfragen werden nach Abschluss des Forschungsprojektes vollautomatisch nach FAIR-Prinzipien archiviert.

9.3 Speicherkontrolle

Daten zu Forschungsanfragen werden nur auf den Servern der Synedra-praeDIZ- und DIZ-Infrastruktur abgelegt. Es gibt keine forschungsanfragenbezogenen Daten auf den Workstations der Mitarbeiter. Das Kopieren von Daten zu Forschungsanfragen auf Mobicdatenträger ist per SOP untersagt. Die Speicherinfrastruktur des Synedra-praeDIZ und des DIZ ist redundant und hoch verfügbar ausgelegt. Die Daten von Forschungsanfragen werden nach Abschluss des Forschungsprojektes vollautomatisch nach FAIR-Prinzipien archiviert.

9.4 Benutzerkontrolle

Alle Mitarbeiter des DIZ-Gießen sind in der Personalliste aufgeführt [16]. Nur diese Mitarbeiter haben auf Basis eines elektronischen Zugangskontrollsystems Zutritt zur Eingangstür und ihrem eigenen Büro. Alle Arbeitsstationen der Mitarbeiter sind über ein Active Directory Zugangskontrolliert und nur über ein benutzerspezifisches Login zu verwenden. Die Verwendung von Gruppen-Logins ist untersagt. Die Vergabe von Logins und rollengesteuerten Zugriffsrecht ist in einer SOP geregelt [21] und wird in Formularen festgehalten die von der DIZ-Leitung gegengezeichnet werden.

9.5 Zugriffskontrolle

Der Zugriff auf nicht-pseudonymisierte Patientendaten des Synedra-praeDIZ ist für Mitarbeiter des DIZ Gießen nicht möglich. Der Zugriff auf pseudonymisierte Daten von Forschungsvorhaben ist nur über die Werkzeuge der DIP möglich. Das Nutzungsverhalten in der DIP wird rollengesteuert kontrolliert. Die Rollenzuweisung erfolgt nach Vorgabe einer SOP [21] und wird in Formularen dokumentiert die von der DIZ-Leitung gegengezeichnet werden. Alle Aktionen der DIP werden protokolliert. Auf diese Protokolle haben nur die Mitarbeiter der Systemadministrationsgruppe des DIZ Zugriff.

9.6 Übertragungskontrolle

Alle Zwischenschritte der Datenextraktion und Bereitstellung, von Synedra-praeDIZ bis zur Bereitstellung der Forschungsdaten in der DIP, werden protokolliert und vom Transferstellenmitarbeiter auf Fehlermeldungen kontrolliert. Jeder Extraktionsvorgang erzeugt ein Datenqualitätsprotokoll, das ebenfalls vom Transferstellenmitarbeiter überprüft wird. In einer SOP ist festgelegt [22], [23], [24] das Probleme, die mit diesen Werkzeugen aufgedeckt werden, auf den wöchentlich stattfindenden Besprechungen der DIZ-Leitung diskutiert werden.

9.7 Eingabekontrolle

Daten für Forschungsanfragen können nur in den datenliefernden Subsystemen in der Patientenversorgung entstehen (SAP, Meona, ICUdata, etc.). Alle diese Systeme verfügen über Eingabekontrollen und Integritätsprüfungen der eingegebenen Daten. Falls im Rahmen der Forschungsanfrage Probleme bei der Datenqualität festgestellt werden, werden diese an den Betreiber des zuständigen klinischen Subsystems gemeldet, sodass sie an der Entstehungsquelle korrigiert werden können. Da jedes im DIZ verwendbare Datenelement mittels der eindeutigen Business-ID von der Quelle bis zur Ausgabe verfolgt werden kann, können Fehler ganz gezielt beseitigt werden.

9.8 Transportkontrolle

Daten von Forschungsanfragen werden innerhalb der Synedra-praeDIZ- und DIZ-Infrastruktur nur auf elektronischem Wege transportiert. Dieser Transportweg wird komplett protokolliert und auftretende Fehler werden kontrolliert. Lösungsmöglichkeiten für solche Fehler werden im Rahmen der DIZ-Leitungsbesprechungen gemeinsam mit der Transferstelle des DIZ gesucht. Für Probleme, die jenseits der Datenintegrationsplattform durch Aktionen des Forschers entstehen ist das DIZ-Gießen nicht verantwortlich. Dies ist im Datennutzungsvertrag, den der Forscher unterzeichnen muss, festgeschrieben [5], [6], [7].

9.9 Wiederherstellbarkeit

Von allen im Synedra-praeDIZ und im DIZ vorhanden Daten werden regelmäßig Backups erstellt. Dies ist in einem entsprechenden Formular [25] festgelegt und wird von der Systemadministrationsgruppe im DIZ überwacht und protokolliert. Die Wiederherstellung wird periodisch getestet.

9.10 Zuverlässigkeit

Alle Komponenten des Synedra-praeDIZ und des DIZ laufen auf einer redundanten und hochverfügbaren virtuellen Plattform (VM-Ware, Docker). Die Datenspeicherung erfolgt in einem NET-APP storage cluster. Dies alles wird in KRITIS zertifizierten Rechnerräumen betrieben. Dadurch ist eine redundante und ausfallsichere Stromversorgung garantiert. Für alle verwendeten Komponenten existieren Wartungsverträge mit den Lieferanten.

9.11 Datenintegrität

Alle Komponenten des Synedra-praeDIZ und des DIZ laufen auf einer redundanten und hochverfügbaren virtuellen Plattform (VM-Ware, Docker). Die Datenspeicherung erfolgt in einem NET-APP storage cluster. Dies alles wird in KRITIS zertifizierten Rechnerräumen betrieben. Die Datenintegrität wird über den gesamten Weg vom klinischen Subsystem über das Synedra-praeDIZ bis zur Zieldatenbank des Forschungsvorhabens in der Datenintegrationsplattform geprüft und protokolliert. Fehlerquellen werden analysiert und eliminiert. Dadurch ist die Datenintegrität von der Quelle bis zum Ziel kontinuierlich gewährleistet.

9.12 Auftragskontrolle

Eine Auftragskontrolle ist nicht notwendig da keine externen Dienstleister eingebunden sind.

9.13 Verfügbarkeitskontrolle

Die Verfügbarkeit des Synedra-praeDIZ und DIZ wird kontinuierlich von der Systemadministrationsgruppe des DIZ und des AKAD überwacht und gewährleistet.

9.14 Trennbarkeit

Die Verarbeitung von identifizierenden (I-DAT) und medizinischen (M-DAT) Daten ist organisatorisch und technisch klar getrennt. I-DAT werden nur im Synedra-praeDIZ von Mitarbeitern der klinischen Routine (AKAD) betreut. Im eigentlichen DIZ werden nur pseudonymisierte M-DAT verarbeitet. Die Verbindung von I-DAT und Pseudonymen findet nur in der externen Treuhandstelle durch unabhängige Mitarbeiter der Treuhandstelle statt. Die zugehörigen technischen Infrastrukturen sind räumlich getrennt.

10. Sensibilisierung / Schulungskonzept

Der für den Datenschutz verantwortliche Mitarbeiter des DIZ führt regelmäßig (mindestens einmal jährlich) eine Schulung der DIZ-Mitarbeiter zu aktuellen Fragen des Datenschutzes durch. Dies ist in einer SOP festgelegt [26]. Die Anwesenheit der Mitarbeiter wird protokolliert und von der DIZ-Leitung gegengezeichnet.

11. Datenschutzvorfälle und Anfragen

In einer SOP [27] ist festgelegt, dass alle Datenschutzvorfälle und Anfragen zum Datenschutz an den für Datenschutz benannten Mitarbeiter des DIZ-Gießen zu richten sind. Dieser führt ein Protokoll und informiert ggf. den Datenschutzbeauftragten des UKGM Standort Gießen sowie die DIZ-Leitung. Die Protokolle werden von der DIZ-Leitung vierteljährlich geprüft. Dies wird ebenfalls protokolliert. Gegebenenfalls erfolgt eine Meldung an die Vertreter der Betreiber des DIZ sowie den Datenschutzbeauftragten des UKGM Standort Gießen.

Abkürzungs- und Fremdwörterverzeichnis

AKAD	Abteilung für klinische und administrative Datenverarbeitung
Broad Consent	Breite Einwilligung
DV	Datenverarbeitung
DIS	Data Integration System
DIZ	Datenintegrationszentrum
DIP	Datenintegrationsplattform
EDC	Electronic Data Capture System
E-PIX	Master-Patient-Index
ETL	Extract-Transform-Load-Prozess
FAIR	Findable, Accessible, Interoperable, and Re-usable
FHIR	Fast Healthcare Interoperability Resources
gICS	Generic Informed Consent Service
gPAS	A generic pseudonym administration service
HL7	Health Level Seven International
ICUData PDMS	Online-Dokumentationssystem für die Intensivmedizin
IDAT	Identifizierende Daten
JLU	Justus-Liebig-Universität
KRITIS	Kritische Infrastruktur
LOINC	Logical Observation Identifiers Names and Codes
MDAT	Medizinische Daten
Meona	Krankenhausinformationssystem
MII	Medizininformatik-Initiative
MIRACUM	Medical Informatics in Research and Care in University Medicine
SNOMED	Systematisierte Nomenklatur der Medizin
SOP	Standard Operating Procedure
SWISSLAB LIS	Laborinformationssystem
UAC	Use & Access Committee
UKGM	Universitätsklinikum Gießen und Marburg

Abbildungsverzeichnis

Abbildung 1: Ablauf Forschungsvorhaben im DIZ Gießen [2], 2022.	4
Abbildung 2: ETL-Architektur des DIZ Gießen, [11], 2022.	6

Referenzen

- [1] Dokument.10.1.Daten-Zugriffs- und Nutzungsordnung_V1.0_2021-04-29
- [2] Dokument.0.13_Prozessbeschreibung-Forschungsvorhaben_V1.1_2021-04-29
- [3] Dokument.1.2_DIP-Handbuch_2_V1.0_2021-04-09
- [4] Dokument.10.3.UAC-GO_V1.0_2021-04-29
- [5] Datennutzungsvertrag_JLU_2021-04-29
- [6] Datennutzungsvertrag_UKGM_JLU_2021-04-29
- [7] MII_NutzV_v1.3_MII_Web
- [8] Medizininformatik Initiative. Interoperabilität in der MII [Internet]. [cited 2021 Oct 10]. Available from: URL: <https://www.medizininformatik-initiative.de/de/zusammenarbeit/arbeitsgruppe-interoperabilitaet>
- [9] Unabhängige Treuhandstelle Universitätsmedizin Greifswald. Unser Record-Linkage E-PIX [Internet]. [cited 2022 Mar 10]. Available from: URL: <https://www.ths-greifswald.de/forscher/e-pix/>
- [10] Unabhängige Treuhandstelle Universitätsmedizin Greifswald. Unser Pseudonym-Management gPAS [Internet]. [cited 2022 Mar 10]. Available from: URL: <https://www.ths-greifswald.de/forscher/gpas/>
- [11] Prozess.3.5_ETL-Strecke-DIZ_V0.4_2022-05-19
- [12] Dokument.0.3_DIZ-Organigramm-extern_V1.0_2021-03-29
- [13] Dokument.0.2_DIZ-Organigramm-intern_V1.0_2021-03-29
- [14] Dokument.10.7.DIZ-Leistungskatalog_V1.0_2021-04-29
- [15] Dokument.10.2.DIZ-GO_V1.0_2021-04-29
- [16] Dokument.10.6.DIZ-Leitung
- [17] Dokument.0.5_Personalstamm-DIZ-Gi_V1.0_2020-12-17
- [18] Unabhängige Treuhandstelle Universitätsmedizin Greifswald. Unser Einwilligungs-Management gICS [Internet]. [cited 2022 Mar 10]. Available from: URL: <https://www.ths-greifswald.de/forscher/gics/>
- [19] 2022-03-22_FB11-Biobank-PATIENTEN_V3-1_mit-Anhang-DZL_clean
- [20] DIZ-Ernennung_Systemadministration
- [21] SOP.3.9_Erstellung-Zugänge-und-Zuweisung-Systemrechte_V2.0_2022-04-05
- [22] SOP.7.4_Validierung-Datenextrakte_V1.0_2021-04-29
- [23] SOP.7.2_SOP-Datenanforderung_V1.0_2021-03-31
- [24] SOP.7.3_SOP-Datenextraktion_V1.0_2021-03-31

- [25] Dokument.3.1_IT-Infrastruktur_V1.0_2021-04-29
- [26] SOP.0.5_SOP-Personalschulung_V1.0_2020-12-18
- [27] SOP.2.8_Umgang_mit_Fragen_zum_Datenschutz_V1.0_2021-03-29

Mitgeltende Unterlagen

1. Dokument.10.1.Daten-Zugriffs- und Nutzungsordnung_V1.0_2021-04-29
2. Dokument.0.13_Prozessbeschreibung-Forschungsvorhaben_V1.1_2021-04-29
3. Dokument.1.2_DIP-Handbuch_2_V1.0_2021-04-09
4. Dokument.10.3.UAC-GO_V1.0_2021-04-29
5. Datennutzungsvertrag_JLU_2021-04-29
6. Datennutzungsvertrag_UKGM_JLU_2021-04-29
7. MII_NutzV_v1.3_MII_Web
8. Prozess.3.5_ETL-Strecke-DIZ_V0.4_2022-05-19
9. Dokument.0.3_DIZ-Organigramm-extern_V1.0_2021-03-29
10. Dokument.0.2_DIZ-Organigramm-intern_V1.0_2021-03-29
11. Dokument.10.7.DIZ-Leistungskatalog_V1.0_2021-04-29
12. Dokument.10.2.DIZ-GO_V1.0_2021-04-29
13. Dokument.10.6.DIZ-Leitung
14. Dokument.0.5_Personalstamm-DIZ-Gi_V1.0_2020-12-17
15. 2022-03-22_FB11-Biobank-PATIENTEN_V3-1_mit-Anhang-DZL_clean
16. DIZ-Ernennung_Systemadministration
17. SOP.3.9_Erstellung-Zugänge-und-Zuweisung-Systemrechte_V2.0_2022-04-05
18. SOP.7.4_Validierung-Datenextrakte_V1.0_2021-04-29
19. SOP.7.2_SOP-Datenanforderung_V1.0_2021-03-31
20. SOP.7.3_SOP-Datenextraktion_V1.0_2021-03-31
21. Dokument.3.1_IT-Infrastruktur_V1.0_2021-04-29
22. SOP.0.5_SOP-Personalschulung_V1.0_2020-12-18
23. SOP.2.8_Umgang_mit_Fragen_zum_Datenschutz_V1.0_2021-03-29

Lastenheft

Entwicklung eines Systems zur
Durchführung von
Machbarkeitsabfragen im
Forschungskontext

1 Einleitung

Das Datenintegrationszentrum Gießen benötigt ein neues System zur Durchführung von Machbarkeitsabfragen im Forschungskontext. Das System soll in der Lage sein, alle relevanten rechtlichen und organisatorischen Anforderungen zu erfüllen und sich nahtlos in die Infrastruktur und die technischen sowie organisatorischen Abläufe des Datenintegrationszentrums einfügen. Das System muss über Kommunikationsschnittstellen verfügen, um eine maximale Interoperabilität mit anderen Systemen zu gewährleisten und ein durch das BMBF beauftragtes Audit eines Wirtschaftsprüfungsunternehmens bestehen.

2 Anforderungen

2.1 Rechtliche Aspekte

Das System muss den rechtlichen Anforderungen entsprechen, die durch die DSGVO und das Sozialgesetzbuch V (SGB V) vorgegeben sind. Es muss eine detaillierte Behandlung der rechtlichen Aspekte zur Wahrung des Datenschutzes sowie der Rechte der Patienten und Probanden gewährleisten. Es muss ein rechtliches fundamentales Konstrukt generiert und etabliert werden, welches die organisatorischen Aspekte des Datenintegrationszentrums, der Begutachtung und Freigabe von Forschungsanfragen sowie die allgemeinen Richtlinien und Vorgaben zur Nutzung von Daten im Kontext der medizinischen Informatik abdeckt. Die Rahmenbedingungen sind in Form von Ordnungen am Standort zu etablieren.

2.2 Metadatenrepositorium

Das System muss in der Lage sein, klinische Daten aus verschiedenen Quellen zu integrieren und zu verwalten. Hierbei muss die Semantik im Kontext der Klassifikationen und Terminologien gewährleistet sein. Ein neues und zentrales Metadaten-Repositorium wird benötigt, um diese Aspekte in einer Basisvariante abzudecken. Das vorhandene klinische Routinearchiv Synedra.AIM kann nicht für das Vorhaben verwendet werden.

2.3 Infrastruktur und Abläufe

Das System muss sich nahtlos in die Infrastruktur und die technischen sowie organisatorischen Abläufe des Datenintegrationszentrums einfügen. Es muss über Kommunikationsschnittstellen verfügen, um eine maximale Interoperabilität mit anderen Systemen zu gewährleisten und eine hart gekapselte Insellösung zu vermeiden.

2.4 Sicherheitsanforderungen

Das System muss die erforderlichen Sicherheitsmaßnahmen umsetzen, um die Vertraulichkeit, Integrität und Verfügbarkeit der Daten zu gewährleisten. Es muss in der Lage sein, sich an neue Anforderungen im Kontext des Datenschutzes, der Datensicherheit und der standortspezifischen rechtlich-organisatorischen Aspekte anzupassen.

3 Spezifikationen

3.1 Funktionale Anforderungen

- Das System muss in der Lage sein, klinische Daten aus verschiedenen Quellen zu integrieren und zu verwalten.
- Das System muss die Semantik im Kontext der Klassifikationen und Terminologien gewährleisten.
- Das System muss in der Lage sein, Forschungsanfragen zu begutachten und freizugeben.
- Das System muss über Kommunikationsschnittstellen verfügen, um eine maximale Interoperabilität mit anderen Systemen zu gewährleisten.
- Das System muss die erforderlichen Sicherheitsmaßnahmen umsetzen, um die Vertraulichkeit, Integrität und Verfügbarkeit der Daten zu gewährleisten.

3.2 Nicht-funktionale Anforderungen

- Das System muss den rechtlichen Anforderungen entsprechen, die durch die DSGVO und das SGB V vorgegeben sind.
- Das System muss einem durch das BMBF beauftragten Audit eines Wirtschaftsprüfungunternehmens bestehen.
- Das System muss in der Lage sein, sich an neue Anforderungen im Kontext des Datenschutzes, der Datensicherheit und der standortspezifischen rechtlich-organisatorischen Aspekte anzupassen.
- Das System muss sich nahtlos in die Infrastruktur und die technischen sowie organisatorischen Abläufe des Datenintegrationszentrums einfügen.
- Das System muss über eine benutzerfreundliche Oberfläche verfügen, die eine einfache Bedienung ermöglicht.

3.3 Qualitätsanforderungen

- Das System muss eine hohe Zuverlässigkeit und Verfügbarkeit gewährleisten.
- Das System muss eine hohe Performance und Skalierbarkeit gewährleisten.
- Das System muss eine hohe Sicherheit und Integrität der Daten gewährleisten.

4 Zeitplan

- Analyse der Anforderungen: 2 Monate
- Konzeptentwicklung: 3 Monate
- Implementierung: 13 Monate
- Integration und Test: 4 Monat
- Schulungen: 2 Monate

Pflichtenheft

eMDR und rechtlich
organisatorische
Ordnungen/Grundlagen

Version 1.4

Autor des Dokuments	Florian Eberhardt	Datum	29.07.2021
---------------------	-------------------	-------	------------

Historie der Dokumentversionen

Version	Datum	Autor	Änderungsgrund / Bemerkungen
1.0	17.05.2019	Florian Eberhardt	Erstellung
1.1	31.05.2019	Florian Eberhardt	Anforderungen durch PI und Projektbeteiligte ergänzt
1.2	27.08.2019	Florian Eberhardt	Anforderungen an das eMDR ergänzt
1.3	30.08.2019	Florian Eberhardt	Finalisierung des Dokuments mit kleinen Korrekturen
1.4	29.07.2021	Florian Eberhardt	Kleinere Anpassungen auf Basis des durchgeführten Audits

Inhaltsverzeichnis

Historie der Dokumentversionen	2
Inhaltsverzeichnis	2
1 Einleitung	5
1.1 Allgemeines	5
1.1.1 Zweck und Ziel dieses Dokuments	5
1.1.2 Projektbezug	5
1.1.3 Abkürzungen	6
1.2 Beteiligte	6
2 Konzept und Rahmenbedingungen	7
2.1 Ziele des Anbieters	7
2.2 Ziele und Nutzen des Anwenders	7
2.3 Benutzer / Zielgruppe	7
2.4 Systemvoraussetzungen	7
2.5 Ressourcen	8
3 Beschreibung der Anforderungen	9
3.1 Anforderung 1	9
3.1.1 Beschreibung	9
3.1.2 Wechselwirkungen	9
3.1.3 Umsetzung	10
3.2 Anforderung 2	10

3.2.1	Beschreibung	10
3.2.2	Wechselwirkungen	11
3.2.3	Umsetzung	11
3.3	Anforderung 3	12
3.3.1	Beschreibung	12
3.3.2	Wechselwirkungen	12
3.3.3	Umsetzung	12
3.4	Anforderung 4	13
3.4.1	Beschreibung	13
3.4.2	Wechselwirkungen	13
3.4.3	Umsetzung	13
3.5	Anforderung 5	14
3.5.1	Beschreibung	14
3.5.2	Wechselwirkungen	14
3.5.3	Umsetzung	14
3.6	Anforderung 6	15
3.6.1	Beschreibung	15
3.6.2	Wechselwirkungen	15
3.6.3	Umsetzung	15
3.7	Anforderung 7	16
3.7.1	Beschreibung	16
3.7.2	Wechselwirkungen	16
3.7.3	Umsetzung	16
3.8	Anforderung 8	17
3.8.1	Beschreibung	17
3.8.2	Wechselwirkungen	17
3.8.3	Umsetzung	17
3.9	Anforderung 9	18
3.9.1	Beschreibung	18
3.9.2	Wechselwirkungen	18
3.9.3	Umsetzung	18
3.10	Anforderung 10	19

3.10.1	Beschreibung	19
3.10.2	Wechselwirkungen	19
3.10.3	Umsetzung	19
3.11	Anforderung 11	19
3.11.1	Beschreibung	19
3.11.2	Wechselwirkungen	20
3.11.3	Umsetzung	20
3.12	Anforderung 12	20
3.12.1	Beschreibung	20
3.12.2	Wechselwirkungen	20
3.12.3	Umsetzung	21
3.13	Anforderung 13	21
3.13.1	Beschreibung	21
3.13.2	Wechselwirkungen	21
3.13.3	Umsetzung	22
3.14	Anforderung 14	22
3.14.1	Beschreibung	22
3.14.2	Wechselwirkungen	22
3.14.3	Umsetzung	22
3.15	Anforderung 15	23
3.15.1	Beschreibung	23
3.15.2	Wechselwirkungen	23
3.15.3	Umsetzung	23
3.16	Anforderung 16	24
3.16.1	Beschreibung	24
3.16.2	Wechselwirkungen	24
3.16.3	Umsetzung	24
4	Abnahmekriterien	25
5	Anhang / Ressourcen	25
5.1	Mitgeltende Dokumente	25

1 Einleitung

1.1 Allgemeines

1.1.1 Zweck und Ziel dieses Dokuments

Dieses Pflichtenheft definiert die Anforderungen an die zu erstellenden rechtlichen und organisatorischen Ordnungen sowie der technischen Umsetzungen eines erweiterten Metadaten-Repositorys, im folgenden eMDR genannt. Die Umsetzung folgt im Zuge der Dissertation von Florian Eberhardt.

1.1.2 Projektbezug

Die zu erstellenden Ordnungen und das zu entwickelnde System stellen im Rahmen des Aufbaus des Gießener Datenintegrationszentrums (DIZ) ein essenzielles Erfolgskriterium dar.

Das Datenintegrationszentrum am Standort Gießen wird im Zuge der Medizininformatikinitiative und einer damit einhergehenden Förderung des Standortes aufgebaut und etabliert. Gießen ist hierbei Teil des MIRACUM-Konsortiums und kann sich an bereitgestellten Softwarelösungen der einzelnen Standorte und Partner bedienen. Die sogenannten MIRACOLIX-Tools stellen hierbei ein OpenSource-Toolset dar, welches im Zuge von dezidierten Arbeitspaketen an den einzelnen Standorten installiert und getestet werden muss.

Aufgrund der speziellen Anforderung des Standortes wird eine Eigenentwicklung benötigt, welche sich jedoch nahtlos in die bestehenden Infrastrukturen eingliedert und bereits vorhandene Systeme mittels Schnittstellen ansprechen und nutzen kann. Ziel ist es eine homogene Systemlandschaft zu schaffen, welche das Ziel verfolgt die Forschung am Standort Gießen technisch zu unterstützen und Machbarkeitsabfragen nicht nur zu vereinfachen, sondern auch abteilungsübergreifend und somit standortweit zu ermöglichen. Die zu erstellenden Ordnungen bieten hierbei die rechtliche Grundlage und Handlungsfähigkeit des DIZ, des Use- & Access-Committee (UAC), sowie die Absicherung von Forschenden im Zuge ihrer Projektdurchführung mittels Daten aus Forschung und Routine, welche im DIZ zusammengeführt wurden.

1.1.3 Abkürzungen

Abkürzung	Bedeutung
DIZ	Datenintegrationszentrum
MII	Medizininformatikinitiative
eMDR	Erweitertes Metadatenrepositorium
UAC	Use- & Access-Committee
DZNO	Datenzugriffs- und Nutzungsordnung
MDR	Metadatenrepositorium
JSON	JavaScript Object Notation
DIP	Datenintegrationsplattform
JLU	Justus-Liebig-Universität Gießen
UKGM-Gi	Universitätsklinikum Gießen und Marburg GmbH Standort Gießen
QM	Qualitätsmanagement
ETL	Extract Transform Load
PI	Principal Investigators
GIFONET	Gießener Forschungsnetz
MIG	Medical Informatics Group

1.2 Beteiligte

Rolle	Name
Betreuer	Prof. Dr. rer. physiol. Keywan Sohrabi
Betreuer	Herr Prof. Dr. biol. hom. Kurt Marquardt
Betreuer	Herr Dr. med. Achim Michel-Backofen
Umsetzung	Florian Eberhardt, M. Sc.

2 Konzept und Rahmenbedingungen

2.1 Ziele des Anbieters

Ziel ist die Erstellung rechtlich organisatorischer Nutzungs- und Geschäftsordnungen sowie die Bereitstellung einer webbasierten Anwendung zur Durchführung von Machbarkeitsabfragen auf den Datenbestand des DIZ. Ein benutzerfreundlicher Aufbau und eine intuitive Bedienung der Komponenten und Funktionen sollen hierbei im Vordergrund stehen.

2.2 Ziele und Nutzen des Anwenders

Der Anwender hat die Möglichkeit nach einem vereinheitlichten Login und einer Freigabe durch das lokale UAC Machbarkeitsabfragen auf den Datenbestand durchzuführen. Das System selbst kann im Hintergrund über Schnittstellen angesprochen und mit Daten bespielt werden. Weiter können einfache Mengenoperationen auf die extrahierten Informationen durchgeführt werden.

Das System soll dem Anwender nicht nur Informationen über das Vorkommen, sondern auch die Semantik der extrahierten Daten bereitstellen.

2.3 Benutzer / Zielgruppe

Die Zielgruppe wird primär durch Forschende des Fachbereichs 11 Medizin der Justus-Liebig-Universität Gießen repräsentiert. Ebenso betroffen sind Mitarbeiter des Universitätsklinikums Gießen und Marburg GmbH am Standort Gießen, welche im Zuge eines Projektantrages an das UAC eine Freigabe erhalten haben.

2.4 Systemvoraussetzungen

Das zu entwickelnde eMDR wird unter dem Webserver Nginx in aktuellster Version und der Python Version 3.7 mit dem Framework Flask betrieben. Das Frontend wird mittels Vue.js in Version 2 realisiert. Eine Speicherung der Informationen erfolgt über eine PostgreSQL-Datenbank in Version 12.11. Eine Synchronisation von Datenbeständen unter den benötigten Systeminstanzen erfordert mehrere bereitgestellte virtuelle Serverumgebungen und der

Installation von Airbyte. Zu verwenden sind Serverinstanzen mit einem Betriebssystem Ubuntu Server in mindestens der Version 16.04.

2.5 Ressourcen

Zur Implementierung werden diverse Python Pakete verwendet, welche beispielsweise konkrete Funktionen im Bereich des Datenmanagements, der Datenbankanbindung und Schnittstellen bereitstellen. Namentlich zu nennen sind hier die wichtigen Pakete python-dotenv, jsonschema, Requests, psycopg2 und plotly. Weiter wird zur Verwendung einer Container-Virtualisierung die Software Docker benötigt. Zur grafischen Darstellung der Ergebnisse wird die Software Plotly benötigt. Weiter erfolgt die Nutzung von Keycloak und der hauseigenen Datenintegrationsplattform (DIP).

3 Beschreibung der Anforderungen

Es folgt eine Auflistung der einzelnen Anforderungen und der Plan für deren Umsetzung.

3.1 Anforderung 1

Nr. / ID	ORG_R01	Nichttechnischer Titel	Rechtlich organisatorische Ordnungen erstellen
----------	---------	------------------------	--

3.1.1 Beschreibung

Zum Aufbau und Etablierung des DIZ am Standort Gießen werden unterschiedliche Ordnungen benötigt, welche diverse Bereiche im Kontext der Forschung, Antragsstellung und Nutzung von Daten am Standort abdecken. Zu erstellen sind drei Ordnungen, welche nachfolgend definiert sind:

1. Zur Beantragung, Verwaltung und Herausgabe von Daten wird am Standort Gießen eine **Datenzugriffs- und Nutzungsordnung** benötigt, welche die Grundlage der zukünftigen Antragsstellung darstellt. Weiter definiert die Ordnung welche Rechte und Pflichten ein Antragssteller im Kontext von Datenprojekten besitzt.
2. Zur Verwaltung und Freigabe von Forschungsanträgen benötigt der Standort, auch bedingt durch die Zugehörigkeit zu einem MII-Konsortium, ein Use- & Access-Committee. Die rechtliche Grundlage zur Handlungsfähigkeit des Gremiums wird durch eine entsprechende **Geschäftsordnung des Use- & Access-Committee** gewährleistet.
3. Die Aufgaben, sowie Rechte und Pflichten des Datenintegrationszentrums müssen schriftlich und transparent dargestellt werden. Hierbei müssen die Aufgaben, interne Strukturen/Organisation und die Tätigkeitsschwerpunkte bei gestellten Forschungs- und Projektanfragen klar definiert sein. Diese Grundlage muss durch die **Geschäftsordnung des Datenintegrationszentrums** repräsentiert werden.

3.1.2 Wechselwirkungen

Direkte Wechselwirkungen gibt es im Kontext der drei zu erstellenden Ordnungen, der Datenzugriffs- und Nutzungsordnung, der Geschäftsordnung des Use- & Access-Committee, sowie der Geschäftsordnung des Datenintegrationszentrums. Die Inhalte dürfen sich untereinander weder einschränken, noch darf es bei einer direkten Gegenüberstellung thematisch zu unterschiedlichen Auslegungen im Kontext der Geltungsbereiche kommen.

Eine indirekte Wechselwirkung kann durch die, innerhalb der MII zentral definierten Nutzungsordnung dargestellt werden. So dürfen bei standortübergreifenden Abfragen, welche auf der zentralen Nutzungsordnung basieren, keine lokalen Strukturen diesem Vorhaben

widersprechen. Ebenso muss die Datenhoheit weiterhin beim Standort verbleiben und ein zentraler Antrag kann die lokalen Genehmigungsstrukturen nicht umgehen.

3.1.3 Umsetzung

Die Umsetzung der Anforderung findet in enger Zusammenarbeit mit den Projektbetreuern und den Beteiligten des Aufbaus des DIZ statt. Weiter werden über die Projektbetreuer die standortspezifischen Interessen gewahrt und die generierten Inhalte mit die unterschiedlichen Forschungs- und Rechtsabteilungen der beteiligten Institutionen abgestimmt. Die benötigten Vorgaben an die Ordnungen wurden im Vorfeld grob erfasst und werden anschließend im Detail ausgearbeitet.

Die Erzeugung der Dokumente erfolgt im Rahmen des Qualitätsmanagements des DIZ und weist eine Dokumentenhistorie und Dokumentenverwaltung mittels der Versionsverwaltung Git auf.

Der Stand der Datenzugriffs- und Nutzungsordnung wird nach erster Erzeugung den Projektleitern und rechtlichen Vertretern zur Durchsicht übermittelt. Bei daraus folgenden Änderungsvorschlägen werden diese eingearbeitet und erneut zur Prüfung vorgelegt. Die Wiedervorlage wird über das Qualitätsmanagement des DIZ organisiert und abgewickelt.

3.2 Anforderung 2

Nr. / ID	EMDR_R02	Nichttechnischer Titel	Backend des eMDR
----------	----------	------------------------	------------------

3.2.1 Beschreibung

Das eMDR stellt ein eigenständiges System dar und benötigt die Betrachtung der nachfolgenden Aspekte, um ein Grundgerüst des Systems zu erstellen. Zu Beginn erfolgt die Erzeugung des Backend:

- **Entwicklung (Sprache/Werkzeug):** Python 3 unter Verwendung des Frameworks Flask
- **Datenbasis:** PostgreSQL-Datenbank
- **Authentifizierung:** Die Authentifizierung wird primär in das Frontend, beziehungsweise Keycloak ausgelagert
- **Infrastruktur:** Nginx Webserver
- **Schnittstellen:** RESTful-API zum Frontend, zu angebundenen Systemen
- **Dokumentation:** Einheitliche und ausführliche Dokumentation des Codes, der Funktionen, der verwendeten Infrastruktur und des Gesamtkontextes

3.2.2 Wechselwirkungen

Das Backend selbst hat eine direkte Wechselwirkung mit dem Frontend. Die bereitgestellten Daten müssen von dem Frontend interpretiert und weiterverarbeitet werden können. Ebenso müssen die bereitgestellten Schnittstellen zu angebundenen Systemen konform sein.

3.2.3 Umsetzung

Das Backend wird mittels der Entwicklungsumgebung PyCharm unter Python Version 3 entwickelt. Hierbei wird zur Definition und Bereitstellung von Schnittstellen das Framework Flask verwendet. Beide Komponenten sind technisch stark etabliert und für den Einsatz in Kombination mit unterschiedlichen Datenbasen geeignet. Als Datenbasis wird eine PostgreSQL-Datenbank verwendet, welche mittels entsprechenden Python-Tools direkt angesprochen wird.

Flask stellt ein sehr leichtgewichtiges Framework da, welches auch die benötigte Performance und Kompatibilität bereitstellt. Die Python Anwendung lässt sich mit ihren zu definierenden Schnittstellen nahtlos in die DIZ-Infrastruktur integrieren. Zur Bereitstellung wird ein Nginx-Webserver konfiguriert, welcher die Python Umgebung (venv) bereitstellt. Die Bereitstellung erfolgt unter einer virtuellen Maschine innerhalb der DIZ-Infrastruktur unter Verwendung von Ubuntu Server 16.04 oder höher als Betriebssystem.

Die Authentifizierung von Benutzern und Systemen wird komplett in das Frontend ausgelagert. Die Dokumentation des Systems erfolgt über den Quellcode, sowie einer Dokumentation innerhalb der hausinternen GitLab-Instanz des DIZ.

Funktional stellt das Backend die Kommunikation mit der Datenbasis, dem Frontend und angebundenen Diensten bereit. Weiter erfolgen im Backend aufgrund der gegebenen Performance sämtliche, eventuell benötigten, Berechnungen und Datenextraktionen sowie Anpassungen.

Die Datenbasis selbst wird auf Grundlage eines zu definierenden Kerndatensatzes generiert, um die für Machbarkeitsabfragen benötigten Informationen, strukturiert und geschwindigkeitsorientiert, abspeichern zu können. Hierbei werden separate Sichten, sogenannte Views, eingesetzt um konkrete Datenbereiche innerhalb der Datenbank vorzuhalten und längere Abfragezeiten zu vermeiden.

3.3 Anforderung 3

Nr. / ID	EMDR_R03	Nichttechnischer Titel	Frontend des eMDR
----------	----------	------------------------	-------------------

3.3.1 Beschreibung

Das eMDR stellt ein eigenständiges und getrenntes System unter dem Aspekt Backend und Frontend dar. Zur Erstellung des Frontend und somit dem Einstiegspunkt der Nutzer müssen folgende Aspekte berücksichtigt werden:

- **Entwicklung (Sprache/Werkzeug):** Vue.js unter Verwendung von Vuetify
- **Datenbasis:** Entfällt, da diese über das Backend abgebildet wird
- **Authentifizierung:** Keycloak-Instanz des DIZ (Tokenbasiert)
- **Infrastruktur:** Nginx Webserver
- **Schnittstellen:** RESTful-API zum Backend, zu angebotenen Systemen
- **Dokumentation:** Einheitliche und ausführliche Dokumentation des Codes, der Funktionen, der verwendeten Infrastruktur und des Gesamtkontextes

3.3.2 Wechselwirkungen

Das Frontend selbst hat eine direkte Wechselwirkung mit dem Backend. Die bereitgestellten Daten müssen von dem Frontend interpretiert und weiterverarbeitet werden können. Ebenso muss die Token-basierte Authentifizierung im Kontext des Backend verarbeitet werden können um einen sicheren Informationsaustausch unter angebotenen Systemen zu ermöglichen.

3.3.3 Umsetzung

Das Frontend wird mittels der Entwicklungsumgebung Visual Studio Code und unter Vue.js mit dem Material Design Framework Vuetify entwickelt. Es werden benötigte Schnittstellen zu dem Backend mittels Axios entwickelt, wodurch eine Kommunikation zwischen dem Frontend und dem Backend final erfolgen kann.

Das primäre Ziel stellt die Benutzerfreundlichkeit und Vereinfachung der Datendarstellung im Kontext von Machbarkeitsabfragen dar. Hierbei gilt es ein schlichtes Design und eine simple Benutzerführung umzusetzen. Es wird aus Sicht der Oberfläche eine Menüleiste erstellt, welche die wesentlichen Punkte einer Machbarkeitsabfrage unmittelbar abdeckt. Die einfache Suche nach Vorkommen benötigt hierbei lediglich einen Code und eine dazugehörige Terminologie / Klassifikation, welche in der ersten Version mittels ICD-10, OPS-301 und LOINC abgebildet werden.

Die Anzeige von Resultaten wird mittels der Darstellung des konkreten Vorkommens implementiert. Zusätzlich dazu erfolgt die Darstellung der Semantik und dem hierarchischen

Aufbau beziehungsweise der Position des abgefragten Kodes innerhalb der korrespondierenden Terminologie / Klassifikation.

Die Authentifizierung an dem System soll über die Anbindung bereits vorhandener Authentifizierungsdienste des DIZ erfolgen. Hierzu verfügt Vue.js bereits über diverse Kompatibilitäten wodurch es eine Grundlage zu einer solchen Anbindung bietet.

Die Dokumentation des Systems erfolgt über den Quellcode, sowie einer Dokumentation innerhalb der hausinternen GitLab-Instanz des DIZ. Die Bereitstellung des Frontend erfolgt mittels einem Nginx-Webserver, welcher innerhalb der DIZ-Infrastruktur, unter Verwendung von Ubuntu Server 16.04 oder höher als Betriebssystem, zur Verfügung gestellt wird.

3.4 Anforderung 4

Nr. / ID	EMDR_R04	Nichttechnischer Titel	Anforderungen an Ordnungen
----------	----------	------------------------	----------------------------

3.4.1 Beschreibung

Basierend auf den zu erstellenden Ordnungen erfolgen direkte Anforderungen an Systeme für Machbarkeitsabfragen und somit auch für das eMDR. Die Anforderungen müssen in konkret und unkonkret unterschieden werden, wohingegen letzteres eine Rücksprache mit den beteiligten Verantwortlichen mit sich bringt.

3.4.2 Wechselwirkungen

Es können Wechselwirkungen neuer benötigter Funktionen auf bereits existente Funktionen entstehen, welche erst absehbar sind, nachdem die Ordnungen erstellt wurden.

3.4.3 Umsetzung

Die aus den Ordnungen resultierenden Anforderungen an das eMDR werden separat dokumentiert und mit der Leitung des DIZ besprochen. Die Umsetzung der Anforderungen erfolgt während der eMDR Entwicklung und kann unterschiedlichste Bereiche des Frontend und Backend betreffen.

Technische Änderungen werden direkt implementiert. Organisatorische Änderungen werden zunächst mit den bestehenden Prozessen innerhalb des DIZ ab- und im weiteren Verlauf angeglichen. Zusätzlich hierzu erfolgt eine Zusammenarbeit mit dem Qualitätsmanagement des DIZ, um konkrete Brüche innerhalb der bereits etablierten Abläufe zu vermeiden.

Sofern datenschutzrechtliche Aspekte betroffen sind, werden diese mit dem Ansprechpartner für Datenschutz innerhalb des DIZ besprochen. Sollte es keine Einigung oder es zu weiterführenden Fragen beziehungsweise Probleme kommen, wird der zuständige Datenschutzbeauftragte der Institution (JLU und / oder UKGM-GI) kontaktiert.

Sofern gewünschte Änderungen aus technischer Sicht nicht umsetzbar sein sollten, erfolgt eine Analyse und Lösungsfindung mit den Projektbetreuern und Principal Investigators (PIs) des MIRACUM-Projektes am Standort Gießen.

3.5 Anforderung 5

Nr. / ID	EMDR_R05	Nichttechnischer Titel	Synchronisation etablieren
----------	----------	------------------------	----------------------------

3.5.1 Beschreibung

Die Synchronisation von Daten zwischen unterschiedlichen Datenbanken ist zwingend erforderlich um einheitliche Datenbestände im Routine- und Forschungsumfeld zu erhalten. Die Synchronisation selbst muss in konkreten Intervallen und stets transparent erfolgen.

3.5.2 Wechselwirkungen

Es können Wechselwirkungen zwischen den unterschiedlichen Datenbanken entstehen, sofern die Synchronisation nicht erfolgreich verläuft und die Systeme daraufhin bei Abfragen unterschiedliche Ergebnisse liefern.

3.5.3 Umsetzung

Die Synchronisation, welche über mindestens zwei Datenbanken erfolgen soll wird mittels einer Middleware umgesetzt. Diese Middleware wird durch die Software Airbyte repräsentiert, welche im Kontext von Datenintegrationen (ETL-Stecken) bereits etabliert ist. Die Software gliedert sich in Quell- und Ziel-Konnektoren auf und bietet einen entsprechenden PostgreSQL-Konnektor für beide dieser Bereiche.

Die Software wird auf einer separaten virtuellen Maschine mit mindestens Ubuntu Server 16.04 installiert und benötigt Zugriff auf den Port 5432 der jeweiligen Datenbankinstanzen. Diese Freigabe erfolgt durch eine Anpassung der korrespondierenden Firewalls.

Airbyte selbst wird für eine Vollsynchronisation konfiguriert, wodurch die Datenbestände auch Prioritäten erhalten können und permanent den gleichen Datenbestand vorweisen. Das

Synchronisationsintervall wird auf nächtlich definiert, wodurch tagsüber der Netzwerkverkehr reduziert wird und die Daten zur Dokumentation tagesaktuell verwendet werden können.

Es werden für die Synchronisationen stets Protokolle erstellt, welche im Zuge einer späteren, eventuell benötigten Fehleranalyse als Grundlage verwendet werden können.

3.6 Anforderung 6

Nr. / ID	EMDR_R06	Nichttechnischer Titel	Systemgrafiken erzeugen
----------	----------	------------------------	-------------------------

3.6.1 Beschreibung

Im Zuge des durch die MII beauftragten Audits werden detaillierte Systemgrafiken benötigt, welche die Beziehung zwischen den zu erstellenden Ordnungen und dem eMDR im Gesamtkontext des DIZ, sowohl technisch als auch organisatorisch, darstellen. Die Darstellungen müssen in enger Zusammenarbeit mit den innerhalb des DIZ beteiligten Einheiten erfolgen um einen hohen Detailgrad zu erreichen.

3.6.2 Wechselwirkungen

Es entstehen keine Wechselwirkungen, da die Grafiken initial erstellt werden und keinen Einfluss auf bereits bestehende Prozesse haben. Die bereits bestehenden Prozesse müssen in die Grafiken als IST-Zustand aufgenommen werden.

3.6.3 Umsetzung

Die Grafiken zur Eingliederung der Ordnungen und des eMDR in die organisatorische und technische Landschaft des DIZ erfolgt unter Verwendung der Software Draw.io. Für die Erstellung werden die nachfolgenden DIZ internen Einheiten hinzugezogen:

1. Systemadministration: Zur Erstellung einer konkreten Übersicht über die virtuelle Systemstruktur und die Kommunikationswege unter den einzelnen Diensten. Weiter können geplante infrastrukturelle Änderungen bereits erfasst und als zukünftige oder optionale Wege dargestellt werden.
2. IT-Sicherheitsbeauftragter: Im Zuge der Definition von Systemschnittstellen und dem Datenaustausch über Systeme und Netzwerke hinweg, können innerhalb der Grafiken infrastrukturelle Trennungen und kritische Punkte markiert werden.

3. Forschungsdatenmanagement: Zur Abbildung des bisherigen Prozesses der Datenintegration und Extraktion von Daten für Machbarkeitsanalysen, um den aktuellen Stand als Ist-Zustand aufnehmen zu können.
4. Qualitätsmanagement: Zur Erstellung des Ist- und Soll-Zustandes aus organisatorischer Sicht des DIZ, im Kontext von Projektanfragen und der dazugehörigen Machbarkeitsabfragen. Weiter existieren bereits Regelungen zur technischen Datenausleitung und Identifikation von Datenbereichen (ebenfalls Machbarkeitsabfragen), welche ebenfalls in die Systemgrafiken aufgenommen werden.

Die erstellten Grafiken werden dem QM übermittelt und können als Grundlage für das Audit genutzt werden.

3.7 Anforderung 7

Nr. / ID	EMDR_R07	Nichttechnischer Titel	Datenimport-Schnittstellen implementieren
----------	----------	------------------------	---

3.7.1 Beschreibung

Das eMDR benötigt Schnittstellen zum Import von Datensätzen. Der Import muss sich auf standortspezifische Formate beziehen und stellt initial den Import von FHIR und CSV (§21-Datensatz) dar. Durch die Freigabe der §21-Datensätze für die Forschungsvorhaben des MIRACUM-Projektes können diese Daten als initialer Datenbestand angesehen werden und müssen somit als ersten Test importiert werden.

3.7.2 Wechselwirkungen

Basierend auf den zu erstellenden Schnittstellen können keine Wechselwirkungen zu anderen Systemen entstehen. Sofern eine Schnittstelle nicht korrekt funktioniert, hat dies keinen Einfluss auf angebundene Systeme, sondern nur auf die eigene Funktionsweise des eMDR.

3.7.3 Umsetzung

Zunächst erfolgt eine Implementierung einer Import-Funktion für CSV-Daten. Diese wird unter Zuhilfenahme von entsprechenden Python-Paketen zum Einlesen von Dateien entwickelt. Die Datei enthält hierbei die Struktur der §21-Datensatz Ausleitungen und wird in dieser Form eingelesen, transformiert und in die Datenbasis integriert.

Die Integration von FHIR-Nachrichten erfolgt über separate Schnittstellen und der engen Zusammenarbeit mit Mitarbeitern des DIZ, welche konkrete Testdatensätze, Strukturen und Informationen zu Besonderheiten liefern können. Weiter erfolgt die Entwicklung eines Bundle-Importers, welche FHIR-Bundles integrieren kann. Der gesamte Importprozess wird mit einem Log unterstützt, welcher bei einer eventuell benötigten Fehleranalyse zur Hilfe gezogen werden kann.

3.8 Anforderung 8

Nr. / ID	EMDR_R08	Nichttechnischer Titel	Anbindung an zentrale Authentifizierungssysteme
----------	----------	------------------------	---

3.8.1 Beschreibung

Zur Vermeidung von Insellösungen muss das eMDR sich in die Authentifizierungsstruktur des DIZ eingliedern und die verfügbaren Dienste nutzen. Im Falle des DIZ wird dies durch Keycloak repräsentiert, welches an die Authentifizierungsdienste des Gießener Forschungsnetzes (GIFONET) angebunden ist und somit einen Login für jeden Forscher im GIFONET ermöglicht. Die Anbindung muss entweder direkt an das Active-Directory oder an Keycloak erfolgen.

3.8.2 Wechselwirkungen

Sofern eine Anbindung nicht erfolgreich sein sollte besteht eine Wechselwirkung darin, dass es zukünftig keine Single-Sign-On Möglichkeiten im Kontext des eMDR geben kann und dies einen Eingriff in den Ablauf der Benutzer, innerhalb des DIZ, darstellt.

3.8.3 Umsetzung

Die Anbindung des eMDR an den zentralen Authentifizierungsdienst des DIZ Keycloak erfolgt primär über das Frontend. Hierbei erfolgt eine Token-basierte Authentifizierung, welche durch eine Implementierung direkt im Frontend ermöglicht wird. Um eine Authentifizierung überhaupt zu ermöglichen muss das eMDR als separater Client innerhalb des DIZ-Realms (Domäne des DIZ) der Keycloak-Instanz aufgenommen werden. Im Anschluss erfolgt eine Konfiguration dieses Clients aus Sicht der benötigten Redirect-Parameter und zugelassenen Benutzern.

Da das Keycloak des DIZ bereits an das Active-Directory des GIFONET angebunden ist, wird dieser sogenannte Identity-Provider auch für den eMDR-Client verwendet. Das Ergebnis ist somit die Anbindung des eMDR an Keycloak und gleichzeitig an das Active-Directory um einen einheitlichen Login für Forschende zu gewährleisten.

Aus technischer Sicht benötigt das eMDR eine verfügbare Kommunikationsstecke zu der internen Keycloak Instanz, welche über eine Anpassung der Firewall gewährleistet wird. Aus organisatorischer Sicht erfolgt noch eine Rücksprache mit dem DIZ internen IT-Sicherheitsbeauftragten und dem Datenschutzbeauftragten, da es sich hierbei um Benutzerdaten und der Bereitstellung von Ergebnissen aus Machbarkeitsanalysen handelt.

3.9 Anforderung 9

Nr. / ID	EMDR_R09	Nichttechnischer Titel	Anbindung an Datenintegrationsplattform
----------	----------	------------------------	---

3.9.1 Beschreibung

Zur Etablierung eines technisch fortschrittlichen und aus organisatorischer Sicht benutzerfreundlichen Ablaufs ist eine Anbindung des eMDR an die Datenintegrationsplattform (DIP) erforderlich. Die innerhalb des eMDR durchgeführten Machbarkeitsabfragen sollen in Form von Datensätzen benutzerzentriert in die DIP integriert werden können.

3.9.2 Wechselwirkungen

Es existieren keine Wechselwirkungen zwischen den beiden Systemen. Die Systeme können auch ohne Anbindung unabhängig voneinander betrieben werden, eine Anbindung hat keinen negativen oder ungewünschten Einfluss auf die Funktionsweisen der beiden Plattformen.

3.9.3 Umsetzung

Die Anbindung an die Datenintegrationsplattform erfolgt in enger Zusammenarbeit mit dem DIP-Team innerhalb des DIZ. Aus Sicht des eMDR erfolgt die Implementierung einer Schnittstelle und dazugehörigen Funktionen um das Ergebnis einer Machbarkeitsabfrage strukturiert als JSON-Datei an eine vom DIP bereitgestellte Schnittstelle zu senden. Der Entwicklungsaufwand aus Sicht der DIP stellt Anpassungen des Frontend und des Backend dar, welcher durch das DIP-Team übernommen wird.

Das Frontend des eMDR wird insofern angepasst, als dass eine erfolgreiche Machbarkeitsabfrage über einen zusätzlichen Menüpunkt direkt an die DIP gesendet werden kann. Im Hintergrund erfolgt über das verfügbare Keycloak-Token eine Authentifizierung an der DIP und die Abfrage nach verfügbaren Domänen und Datensätzen, in denen die Resultate der Machbarkeitsabfrage integriert werden können.

3.10 Anforderung 10

Nr. / ID	EMDR_R10	Nichttechnischer Titel	Mengenoperationen auf Daten implementieren
----------	----------	------------------------	--

3.10.1 Beschreibung

Eine simple Machbarkeitsabfrage ist nach initialem Feedback bereits ausreichend, soll jedoch durch einfache Mengenoperationen auf mehrere Abfragen ausgebaut werden. Hierbei sollen zunächst die Grundfunktionalitäten bereitgestellt werden, welche durch die Verwendung von Mengenoperationen auf zwei Suchkriterien realisiert werden.

3.10.2 Wechselwirkungen

Aufgrund der Tatsache, dass es sich hierbei um eine zusätzliche und ergänzende Funktionalität handelt, existieren keine Wechselwirkungen zu dem eMDR selbst. Da es sich um Ergebnisse der Machbarkeitsanalyse handelt, können höchstens Wechselwirkungen im anschließenden Projektantrag und der finalen Extraktion der Daten auftreten. Diese müssen bei einem ersten Auftreten gesondert betrachtet und behandelt werden.

3.10.3 Umsetzung

Die Umsetzung dieser Anforderung erfolgt im Frontend und kann auch über die Oberfläche ausgewählt werden. Hierbei kann der Benutzer zwei Suchparameter inklusive Kataloge und die Art der Mengenoperation auswählen und die Suche starten. Implementiert werden hierbei die Optionen der Vereinigung, Differenz und des Durchschnitts von Daten. Entsprechend dazu erfolgt die Implementierung der Ergebnisdarstellung, da für beide Parameter die semantische Bedeutung und Hierarchie angezeigt werden sollen.

3.11 Anforderung 11

Nr. / ID	EMDR_R11	Nichttechnischer Titel	Grafische Darstellungsmöglichkeit implementieren
----------	----------	------------------------	--

3.11.1 Beschreibung

Abseits der dargestellten Informationen einer Machbarkeitsabfrage soll zukünftig, bei komplexeren Auswertungen, die Möglichkeit einer grafischen Darstellung der Ergebnisse bestehen. Diese Darstellung soll möglichst dynamisch sein und die Forschenden in ihrer Auswertung nicht einschränken. Ebenso muss die Funktion gegeben sein die grafischen Auswertungen exportieren zu können.

3.11.2 Wechselwirkungen

Aufgrund der Tatsache, dass es sich hierbei um eine zusätzliche und ergänzende Funktionalität handelt, existieren keine Wechselwirkungen zu dem eMDR selbst. Da es sich um grafische Ergebnisse der Machbarkeitsanalyse handelt, können höchstens Wechselwirkungen im anschließenden Projektantrag und der finalen Extraktion der Daten auftreten. Diese müssen bei einem ersten Auftreten gesondert betrachtet und behandelt werden.

3.11.3 Umsetzung

Die Implementierung einer grafischen Darstellung erfolgt mittels der Verwendung von Plotly. Die Software bietet eine ideale Grundlage im Python-Kontext zur Darstellung von Daten in unterschiedlichen Formen und unterschiedlichen Diagrammtypen. Weiter existiert die Möglichkeit Diagramme dynamisch zu vergrößern, Ausschnitte zu erzeugen und die Grafik in gängige Bildformate zu exportieren.

Die Umsetzung erfolgt seitens des Backends und übermittelt dem Frontend ein fertiges JSON, welches direkt von Plotly interpretiert und dargestellt werden kann.

3.12 Anforderung 12

Nr. / ID	EMDR_R12	Nichttechnischer Titel	Abfragemöglichkeit für Transferstelle bereitstellen
----------	----------	------------------------	---

3.12.1 Beschreibung

Ein innerhalb eines Projektantrages, an das UAC und das DIZ, definiertes Ergebnis einer Machbarkeitsabfrage, über das eMDR, benötigt bei Bedarf eine interne Prüfung. Diese Funktion soll ebenfalls über das eMDR bereitgestellt werden und soll einen exklusiven Zugang für Transferstellenmitarbeiter vorhalten. Die Ergebnisse, welche im Projektantrag angegeben wurden, sollen über die neue Funktion gegengeprüft werden.

3.12.2 Wechselwirkungen

Das Ergebnis der internen Prüfung durch die Transferstelle hat direkten Einfluss auf den Projektantrag. Bei einer manipulierten Machbarkeitsabfrage kann der Antrag entsprechend abgelehnt werden. Dies erfordert ein konkretes Handeln der DIZ-Leitung und des UAC. Ein unauffälliges Prüfungsergebnis hat keine sichtbaren Wechselwirkungen.

3.12.3 Umsetzung

Die Implementierung dieser Funktionalität erfolgt über einen neuen Menüpunkt im Frontend, wodurch einem Transferstellenmitarbeiter eine separate Abfrageoption in der gleichen Anwendung zur Verfügung steht. Zur Umsetzung einer Prüffunktionalität erfolgt eine Anpassung im Backend, sodass jede Abfrage des eMDR mit einer eindeutigen ID (UUID-4) versehen wird. Weiter wird jede fertige Abfrage, welche mit einer ID versehen ist nicht nur an das Frontend gesendet, sondern auch intern in einer Historie gespeichert. Die ID gilt es bei einem Projektantrag anzugeben, wodurch sie letzten Endes und im weiteren Verlauf des Projektantrages der Transferstelle übermittelt wird.

Die Mitarbeiter der Transferstelle können dann über den Menüpunkt innerhalb des eMDR die ID eingeben und bekommen den Datensatz aus der Historie angezeigt. Hierbei ist ersichtlich wann die Abfrage durchgeführt wurde und mit welchem Resultat.

Sollte das Resultat nicht übereinstimmen weist dies auf eine Manipulation des Ergebnisses hin und kann dem UAC und der DIZ-Leitung mitgeteilt werden.

3.13 Anforderung 13

Nr. / ID	EMDR_R13	Nichttechnischer Titel	Systemtests durchführen
----------	----------	------------------------	-------------------------

3.13.1 Beschreibung

Das System muss aus Sicht einzelner Funktionen, der Schnittstellen und der Integration in den technischen und organisatorischen Ablauf des DIZ getestet werden. Zunächst soll die Integration von Testdaten erfolgen um weitere Tests durchzuführen. Das Testen von Funktionalitäten und dem Frontend soll nicht nur aus technischer Sicht, sondern auch aus Anwendersicht erfolgen.

3.13.2 Wechselwirkungen

Die Tests selbst haben keine Wechselwirkung auf angebundene Systeme oder dem eMDR selbst. Fehler, die im nächsten Schritt gegebenenfalls aus den Tests resultieren, müssen zunächst besprochen und anschließend in Form eines Implementierungsaufwandes behoben werden.

3.13.3 Umsetzung

Das Testen von Importschnittstellen und der internen Datenverarbeitung wird durch das testweise Einleiten von Daten aus dem §21-Datensatz durchgeführt. Ebenso werden testweise FHIR-Nachrichten importiert, um die Strecken von Datenimport bis zur Datenbank ausgiebig testen zu können.

Sämtliche Funktionen des Backend werden mittels Unit-Tests durchgetestet. Bei eventuell identifizierbaren Fehlern werden diese direkt behoben und erneuten Tests unterzogen.

Weitere Tests erfolgen über die Benutzeroberfläche des eMDR, welche durch Anwender des DIZ durchgeführt werden. Hierbei werden konkrete Abfragen durchgeführt und Eingabevalidierungen geprüft.

3.14 Anforderung 14

Nr. / ID	EMDR_R14	Nichttechnischer Titel	Containerbasierte Bereitstellung gewährleisten
----------	----------	------------------------	--

3.14.1 Beschreibung

Die Bereitstellung des eMDR soll, aus Sicht des technologischen Fortschritts, containerbasiert erfolgen. Dies erfordert eine Umstellung der initial etablierten Infrastruktur aus Sicht des Backends, Frontends und der Datenbank. Aufgrund einer Vorgabe der bereits existenten DIZ-Systemlandschaft ist die Verwendung von Docker zwingend erforderlich.

3.14.2 Wechselwirkungen

Die Umstellung der Systemlandschaft des eMDR hat intern starke Wechselwirkung unter den Komponenten, sobald diese in unterschiedlichen Containern mit unterschiedlichen Netzwerkkonfigurationen betrieben werden. Sofern die Konfigurationen ideal aufeinander abgestimmt sind und die Systeme in den Containern separiert betrieben werden können, sind die zukünftigen Wechselwirkungen durch die schnelle und systemunabhängige Bereitstellung des Systems geringer. Dies ist auf die reduzierte Zeit bei Systemausfällen zurückzuführen und die somit höhere Verfügbarkeit im Gesamtkontext des Forschungsablaufes innerhalb des DIZ und am Standort.

3.14.3 Umsetzung

Das eMDR wird in drei separate Container unterteilt. Das Backend wird mittels einem selbst definierten Ubuntu 20.04 Container mit installiertem Python 3 und Flask umgesetzt,

wohingegen für das Frontend ein bereits existentes Image aus des Docker-Hub genutzt werden kann und nur geringfügige Anpassungen benötigt. Die Datenbank wird mittels dem offiziellen PostgreSQL-Docker-Image in der neuesten Version bereitgestellt.

Die Container selbst werden in einem Docker-Compose-Verbund verwaltet und in einem separaten Subnetz auf dem Docker-Host betrieben. Hierzu wird ein Docker-Host benötigt, welcher innerhalb der DIZ-Infrastruktur unter mindestens Ubuntu Server 16.04 bereitgestellt wird.

3.15 Anforderung 15

Nr. / ID	EMDR_R15	Nichttechnischer Titel	IT-Grundschutz des BSI wahren
----------	----------	------------------------	-------------------------------

3.15.1 Beschreibung

Zur Gewährleistung der IT-Sicherheit in der IT-Infrastruktur des DIZ und somit auch dem GIFONET, muss das IT-Grundschutzkompendium des BSI analysiert und die einzelnen Inhalte auf die Entwicklung des eMDR angepasst werden. Hierbei gelten die relevanten Module des Kompendiums zu beachten, welche sich auf die containerbasierte Bereitstellung beziehen.

3.15.2 Wechselwirkungen

Im Falle der Anwendung von Vorgaben aus dem IT-Grundschutzkompendium resultiert ein weiterer Entwicklungsaufwand für das eMDR. Sofern hierbei Vorgaben zu Schnittstellen und der Kommunikation zwischen angebundenen Systemen entstehen, können dadurch Auswirkungen auf deren Funktionsweise entstehen, falls die Anforderungen nicht umgesetzt werden können.

3.15.3 Umsetzung

Die Umsetzung des BSI IT-Grundschutzes wird gewährleistet, indem die einzelnen Module des Kompendiums auf deren Anwendbarkeit im Kontext des eMDR zunächst analysiert und im Anschluss angewendet werden. Hierbei gilt es die Vorgaben der einzelnen Module stets vollständig zu erfüllen, da nur dann die Module als erfüllt angesehen werden können.

Die Umsetzung erfolgt direkt im Sinne einer Anpassung des Quellcodes oder der infrastrukturellen Grundlage des eMDR.

3.16 Anforderung 16

Nr. / ID	EMDR_R16	Nichttechnischer Titel	Integration in die DIZ-Infrastruktur
----------	----------	------------------------	--------------------------------------

3.16.1 Beschreibung

Das eMDR benötigt für die Anbindung an das Samplly.MDR zur Abfrage des semantischen Kontextes eine volle Integration in die Infrastruktur des DIZ. Ebenso kann der definierte Ablauf eines Projektantrages nur erfolgen, sofern das eMDR innerhalb des GIFONET etabliert und an interne Strukturen wie das Authentifizierungssystem angebunden ist. Ein weiterer Bestandteil wird durch die Anbindung an Wartungs- und Backupsysteme repräsentiert, welche für den technisch optimalen Betrieb des eMDR erforderlich sind.

3.16.2 Wechselwirkungen

Durch die Integration und Anbindung können, durch zuvor durchgeführte Tests, keine Wechselwirkungen mit anderen bereits etablierten Systemen wahrgenommen werden. Je nach verwendetem Backupsystem können die Daten des eMDR aufgrund der containerbasierten Bereitstellung, im Falle eines Systemfehlers und einem einhergehenden Datenverlustes, die Daten in einem unterschiedlich hohen Grad wiederhergestellt werden. Diese Wechselwirkung kann jedoch aufgrund der Synchronisation der Datenbasen ignoriert werden.

3.16.3 Umsetzung

Durch die Installation des eMDR auf einem Docker-Host innerhalb der DIZ-Infrastruktur existiert bereits eine Grundlage für die Anbindung an zusätzliche Systeme. Das Samplly.MDR wird in Kooperation mit der Medical Informatics Group Frankfurt (MIG) angebunden, indem der Standort Frankfurt als Bereitsteller des Samplly.MDR eine Schnittstellendokumentation bereitstellt. Die Anbindung wird dann über das Backend des eMDR erfolgen, indem die semantischen Informationen bei einer Machbarkeitsabfrage direkt abgerufen werden.

Durch die Verwendung von Keycloak und dessen Anbindung existiert eine volle Integration in die Authentifizierungsstrukturen des DIZ.

4 Abnahmekriterien

Die Abnahmekriterien ergeben sich aus den Anforderungen. Alle im vorherigen Kapitel aufgelisteten Anforderungen sind gleichzeitig die Abnahmekriterien.

5 Anhang / Ressourcen

5.1 Mitgeltende Dokumente

Dokument	Version
Lastenheft	1.0
Projektplan	1.0

Nr.	Vorgangsmodus	Vorgangname	Dauer	Anfang	Ende
1		Rechtliche und organisatorische Ordnungen	1023 Tage	Die 01.01.19	Don 01.12.22
2		Erfassung der Anforderungen	109 Tage	Die 01.01.19	Fre 31.05.19
3		Sichtung bereits vorhandener Dokumente	64 Tage	Die 01.01.19	Fre 29.03.19
4		Anforderungen durch PI und Projektbeteiligte	45 Tage	Mon 01.04.19	Fre 31.05.19
5		Erstellung erster Entwurf	282 Tage	Mon 03.06.19	Die 30.06.20
6		Datenzugriffs- und Nutzungsordnung	65 Tage	Mon 03.06.19	Fre 30.08.19
7		Geschäftsordnung des Datenintegrationszentrums	65 Tage	Mon 02.09.19	Fre 29.11.19
8		Geschäftsordnung des Use- & Access-Committee	65 Tage	Mon 02.12.19	Fre 28.02.20
9		Interne Rückmeldung zu ersten Fassungen	44 Tage	Mon 02.03.20	Don 30.04.20
10		Erste Überarbeitung auf Basis interner Rückmeldung	43 Tage	Fre 01.05.20	Die 30.06.20
11		Erste Rückmeldung der Gremien	248 Tage	Mit 01.07.20	Fre 11.06.21
12		Rückmeldung der Gremien über Vertreter der Betreiber einholen	163 Tage	Mit 01.07.20	Fre 12.02.21
13		Überarbeitung der Ordnungen auf Basis der Rückmeldung	85 Tage	Mon 15.02.21	Fre 11.06.21
14		Überarbeitung auf Basis des Audits	194 Tage	Die 15.06.21	Fre 11.03.22
15		Rückmeldung auf Basis des Audits mit Projektverantwortlichen analysieren	19 Tage	Die 15.06.21	Fre 09.07.21
16		Überarbeitung der Ordnungen auf Basis der Analyse	20 Tage	Mon 12.07.21	Fre 06.08.21
17		Finale Überprüfung der Gremien	155 Tage	Mon 09.08.21	Fre 11.03.22
18		Freigabe der Verantwortlichen und Inkrafttreten der Ordnungen	189 Tage	Mon 14.03.22	Don 01.12.22
19		Erweitertes Metadatenrepositorium (eMDR)	1023 Tage	Die 01.01.19	Don 01.12.22
20		Analyse am Standort verfügbarer Systeme	24 Tage	Die 01.01.19	Fre 01.02.19
21		Analyse der im MIRACUM-Projekt verfügbarer Systeme	40 Tage	Mon 04.02.19	Fre 29.03.19
22		Installation und Testung vorhandener Systeme	65 Tage	Mon 01.04.19	Fre 28.06.19

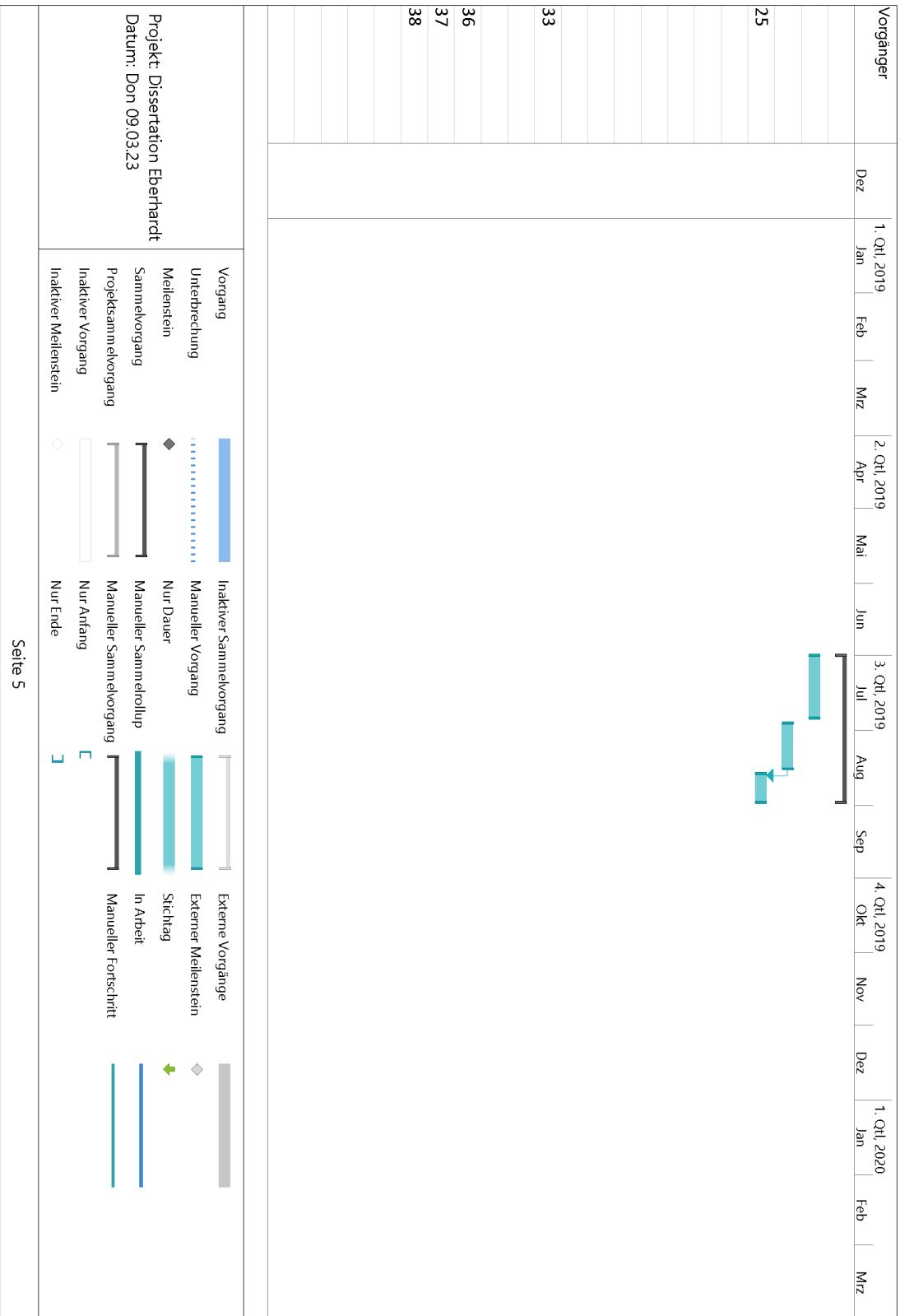
Vorgang	Inaktiver Sammelvorgang	Manueller Vorgang	Externe Vorgänge
Unterbrechung			
Milenstein			
Sammelvorgang			
Projektsammelvorgang			
Inaktiver Vorgang			
Inaktiver Meilenstein			

Inaktiver Sammelvorgang Externe Vorgänge
 Manueller Vorgang Externer Meilenstein
 Nur Dauer Stichtag
 Manueller Sammelrolup In Arbeit
 Manueller Sammelvorgang Manueller Fortschritt
 Nur Anfang Nur Ende
 Inaktiver Meilenstein Nur Ende

Projekt: Dissertation Eberhardt
Datum: Don 09.03.23

Nr.	Vorgangsmodus	Vorgangname	Dauer	Anfang	Ende
45	➔	CSV Import-Schnittstelle	10 Tage	Mon 02.08.21	Fre 13.08.21
46	➔	FHIR Schnittstelle	35 Tage	Mon 16.08.21	Fre 01.10.21
47	➔	Anbindung an zentrale Authentifizierungssysteme des DIZ	35 Tage	Mon 04.10.21	Fre 19.11.21
48	➔	Anbindung an die Datenintegrationsplattform	55 Tage	Mon 22.11.21	Fre 04.02.22
49	➔	Mengenergebnisse auf Suchergebnisse	10 Tage	Mon 07.02.22	Fre 18.02.22
50	➔	Grafische Darstellung von Ergebnissen	10 Tage	Mon 21.02.22	Fre 04.03.22
51	➔	Separate Abfragemöglichkeit für die DIZ-interne Transferstelle	15 Tage	Mon 07.03.22	Fre 25.03.22
52	➔	Systemtests	195 Tage	Mon 30.08.21	Fre 27.05.22
53	➔	Einleitung von Testdaten mittels CSV und FHIR Schnittstelle	30 Tage	Mon 30.08.21	Fre 08.10.21
54	➔	Tests zur Synchronisation der Datenbanken	5 Tage	Mon 28.03.22	Fre 01.04.22
55	➔	Unit- und sonstige Tests zur Prüfung der eMDR-Funktionalitäten	40 Tage	Mon 04.04.22	Fre 27.05.22
56	➔	Infrastrukturelle Anpassungen	60 Tage	Mon 30.05.22	Fre 19.08.22
57	➔	Umstellung des eMDR auf containerbasierte Bereitstellung / Infrastruktur	45 Tage	Mon 30.05.22	Fre 29.07.22
58	➔	Abgleich der infrastrukturellen Umsetzung mit den Anforderungen des BS/IT-Grundschul15 Tage	15 Tage	Mon 01.08.22	Fre 19.08.22
59	➔	Abnahme des Systems durch die DIZ-Leitung und Systemadministration	30 Tage	Mon 22.08.22	Fre 30.09.22
60	➔	Technische Integration des Systems in die IT-Landschaft	19 Tage	Mon 03.10.22	Don 27.10.22
61	➔	Analyse der Verbesserungs- und Erweiterungsmöglichkeiten	25 Tage	Fre 28.10.22	Don 01.12.22

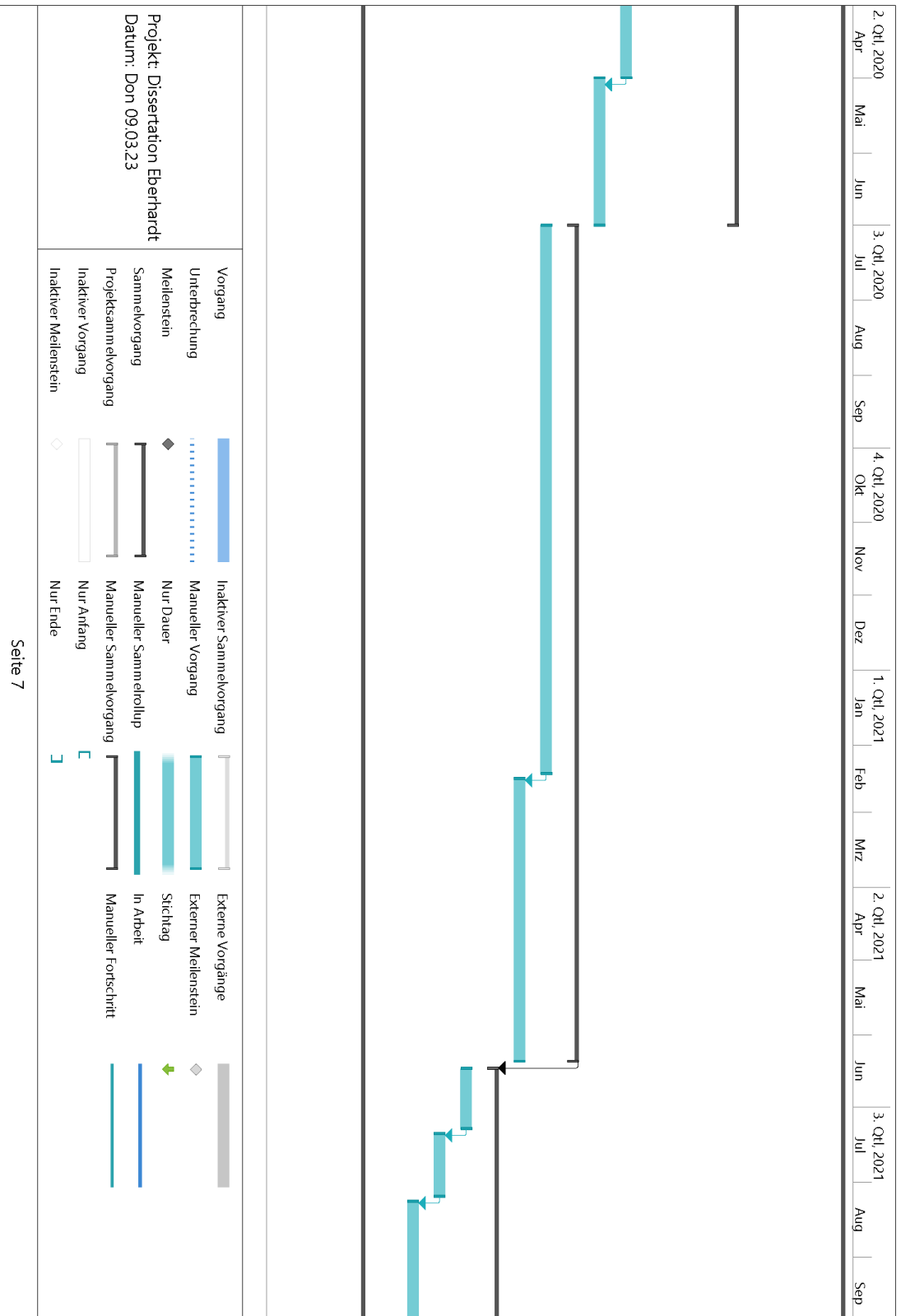


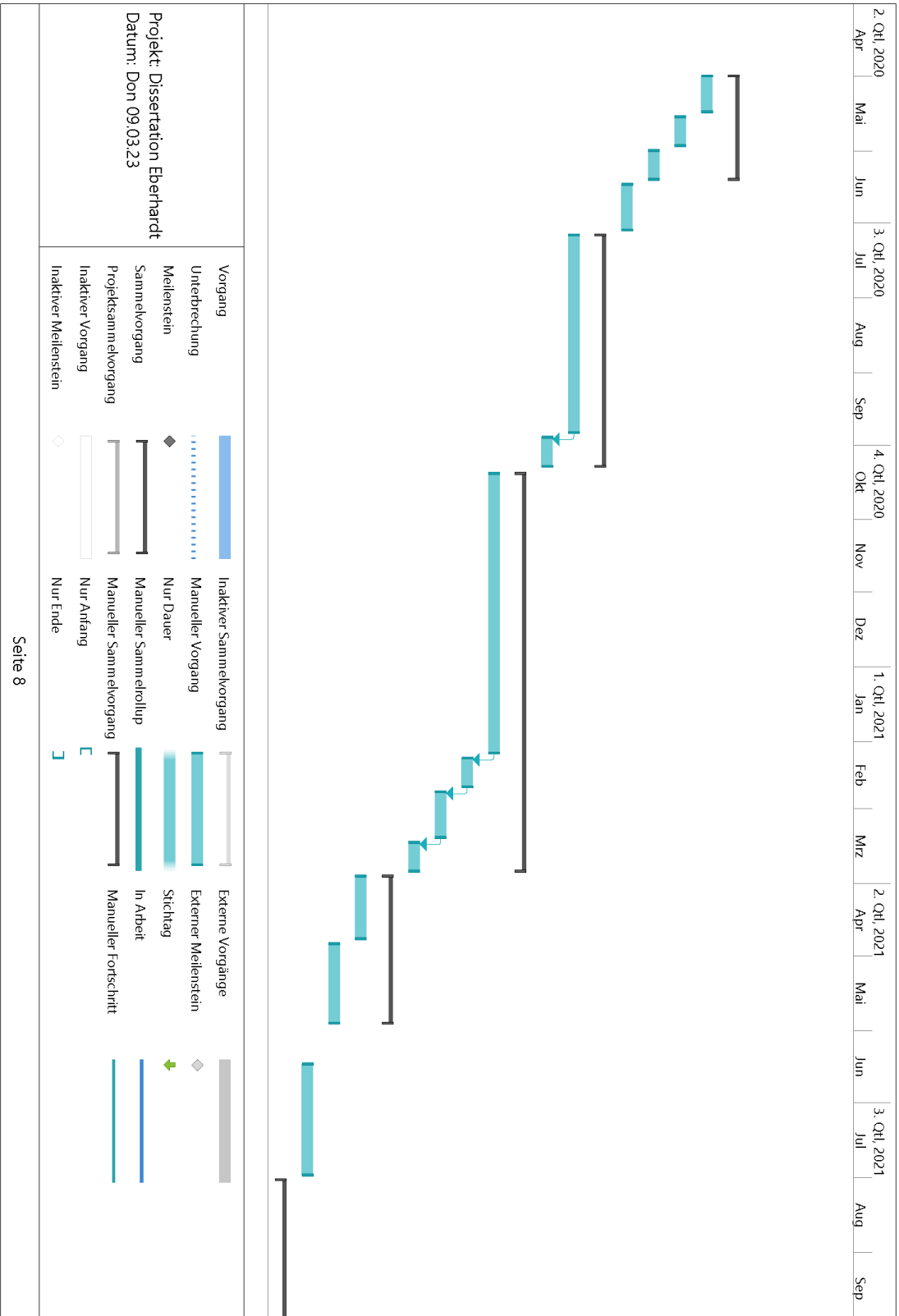


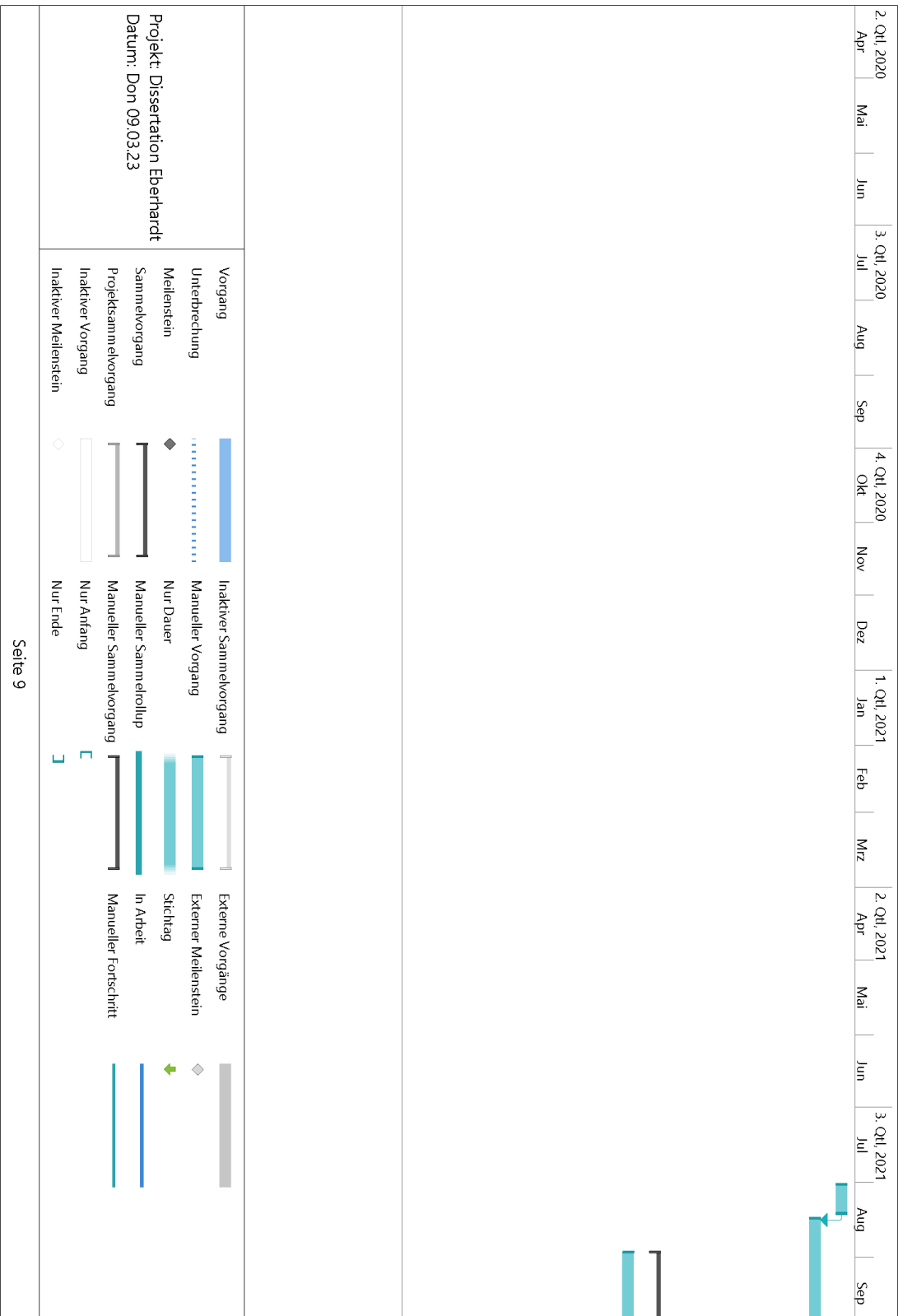
Vorgänger	1. QH, 2019			2. QH, 2019			3. QH, 2019			4. QH, 2019			1. QH, 2020		
	Dez	Jan	Feb	Apr	Mai	Jun	Jul	Aug	Sep	Ok	Nov	Dez	Jan	Feb	Miz
45															
46															
47															
48															
49															
50															
56															
59															
60															

<p>Projekt: Dissertation Eberhardt Datum: Don 09.03.23</p> <p>Vorgang</p> <p>Unterbrechung</p> <p>Meilenstein</p> <p>Sammelvorgang</p> <p>Projektsammelvorgang</p> <p>Inaktiver Vorgang</p> <p>Inaktiver Meilenstein</p>	<p>Inaktiver Sammelvorgang</p> <p>Manneller Vorgang</p> <p>Nur Dauer</p> <p>Manneller Sammelrollup</p> <p>Externe Vorgänge</p> <p>Externer Meilenstein</p> <p>Stichlag</p> <p>In Arbeit</p> <p>Manneller Fortschritt</p> <p>In Arbeit</p> <p>Nur Anfang</p> <p>Nur Ende</p>
--	---

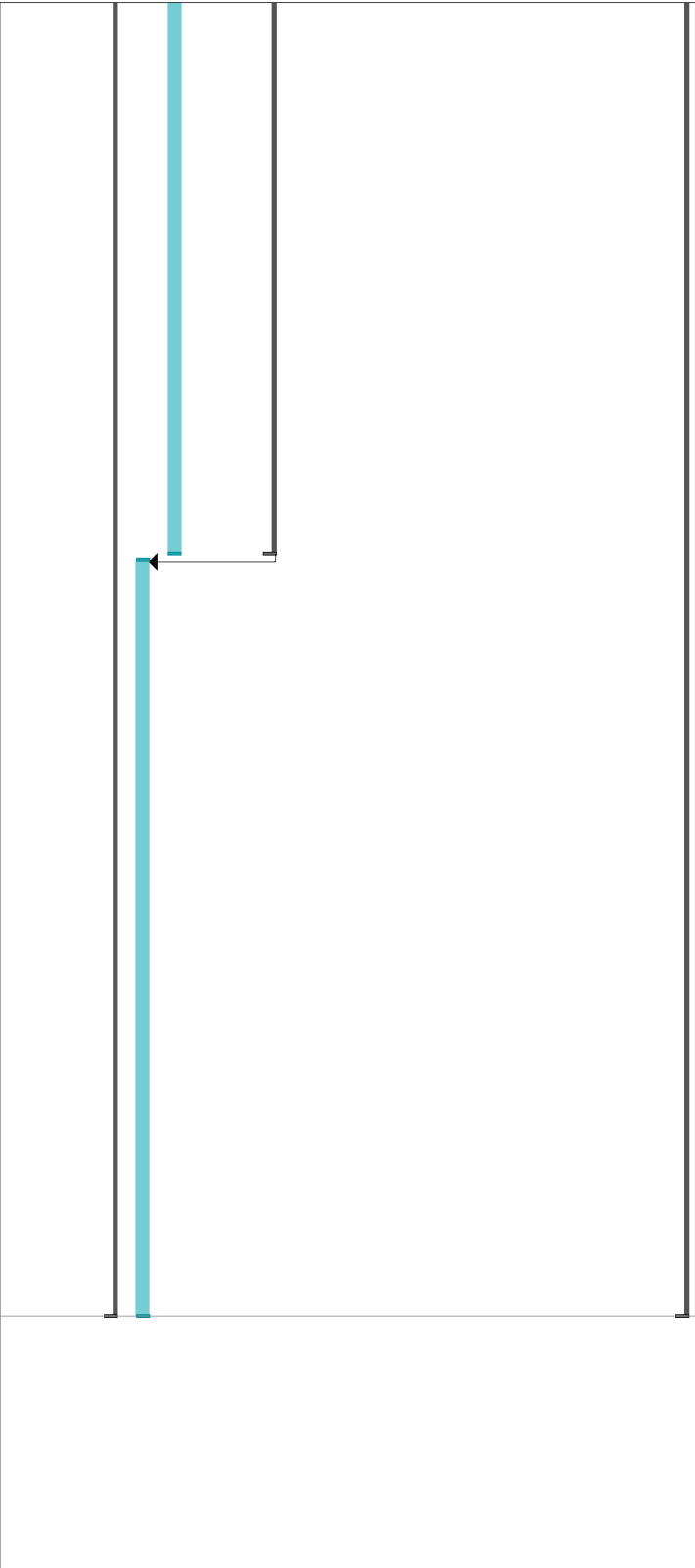
Seite 6







Sep	4. Qtl 2021	Nov	1. Qtl 2022	2. Qtl 2022	3. Qtl 2022	4. Qtl 2022	Nov	1. Qtl 2023
	Oktober		Januar	April	Juli	Oktober	Dezember	Januar



Vorgang	Inaktiver Sammelvorgang	Externe Vorgänge
Unterbrechung	Manueller Vorgang	Externer Meilenstein
Meilenstein	Nur Dauer	Stichtag
Sammelvorgang	Manueller Sammlerollup	In Arbeit
Projektsammelvorgang	Manueller Sammelvorgang	Manueller Fortschritt
Inaktiver Vorgang	Nur Anfang	
Inaktiver Meilenstein	Nur Ende	

Projekt: Dissertation Eberhardt
Datum: Don 09.03.23



