

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.



Konzeptionelle Eckpunkte zur Anbindung der Registerdatenbank an künftige europäische Vernetzungsstrukturen (z.B. den EHDS)

Erstellt für das Bundesministerium für Gesundheit



Bundesministerium
für Gesundheit

Version 1.04

Berlin, 20. Januar 2023



Impressum

TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Charlottenstraße 42 | 10117 Berlin

Tel.: 030 – 22 00 247 – 0 | Fax: 030 – 22 00 247 – 99

E-Mail: info@tmf-ev.de | Internet: www.tmf-ev.de

Ansprechpartnerin:

Dr. Anna Niemeyer

Erarbeitet und vorgelegt von:

Dr. Anna Niemeyer, Berlin

Sebastian C. Semler, Berlin

sowie mit Unterstützung von:

Prof. Dr. Juliane Fluck

Berlin, 20.01.2023

Inhalt

1	Einleitung.....	1
2	Vorgehen.....	3
2.1	Literaturrecherche.....	3
2.2	Recherche laufender Aktivitäten und Projekte.....	4
2.3	Durchführung eines Expertiseworkshops	4
3	Begriffsklärung und -bestimmung	5
4	Metadaten in Deutschland	8
5	Metadaten in Europa	11
6	Probleme / Herausforderungen.....	18
7	Verknüpfbarkeit von Metadaten	19
7.1	Interoperabilität von Metadaten / Standardisierung	19
7.2	Zusammenführung von Metadaten	23
7.3	Kuratierung und Pflege von Metadaten.....	24
8	Harmonisierungsinitiativen.....	24
9	Erste Impulse für Handlungsempfehlungen.....	27
10	Anhang.....	29
10.1	Literaturverzeichnis	29
10.2	Tabellenverzeichnis	32
10.3	Abkürzungsverzeichnis und Glossar	33

1 Einleitung

Allgemeine Einführung in das Thema und Anriss der Fragestellung.

Als eine Konsequenz aus der COVID-19-Pandemie hat die EU die Notwendigkeit eines stärkeren EU-Rahmens für die Gesundheitssicherheit festgestellt und am 11.11.2020 die Eckpfeiler zur Schaffung einer europäischen Gesundheitsunion bekannt gegeben mit dem Ziel die Resilienz der EU gegenüber grenzüberschreitenden Gesundheitsgefahren zu erhöhen¹.

Der Europäische Raum für Gesundheitsdaten (EHDS) ist eine der Säulen der neu geschaffenen europäischen Gesundheitsunion und soll die Datenbasis für eine Verbesserung der Gesundheitsversorgung in der EU bereitstellen und dabei Bürgerinnen und Bürgern die Kontrolle über ihre eigenen Gesundheitsdaten ermöglichen. Zentraler Aspekt zur Erreichung dieser Ziele ist die Bereitstellung eines kohärenten, sicheren, zuverlässigen und effizienten Rahmens für die Verwendung von Gesundheitsdaten, die es Forschenden, Innovatoren, öffentlichen Einrichtungen und der Industrie ermöglicht Zugang zu großen Mengen an Gesundheitsdaten von hoher Qualität zu erlangen. Dies kann nur gelingen, wenn die bestehenden Herausforderungen sowohl der technischen und der semantischen Interoperabilität auch für den grenzüberschreitenden Austausch von Gesundheitsdaten bewältigt werden können.

Grundvoraussetzung hierfür sind neben der Einhaltung des Datenschutzes und der Gewährleistung eines hohen Maßes an Cybersicherheit vor allem die Herstellung von Interoperabilität der nationalen Gesundheitsdatenquellen untereinander und grenzüberschreitend.

Im Entwurf der EU-Verordnung vom 03.05.2022 ist u.a. vorgesehen, dass die für den Zugang zu Gesundheitsdaten zuständigen Stellen einen Metadatenkatalog mit einer Liste der verfügbaren Datensätze führen, in dem jeder Datensatz einschließlich: Datenquelle, Datenumfang, Hauptmerkmale und Bedingungen für die Bereitstellung der Daten beschrieben wird. Die nationalen Kataloge sollen mit dem EU-Datensatzkatalog verbunden werden.²

Metadaten stellen eine zentrale Komponente zur Herstellung von Interoperabilität von Datenquellen dar und sind gleichzeitig, gesammelt in Metadatenkatalogen, ein wichtiger Aspekt zur Herstellung der Auffindbarkeit (der Datenidentifizierung) sowie der Klassifizierung, dem Abruf und der Validierung von Datensätzen.

¹ <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0724&from=EN>

² https://health.ec.europa.eu/document/download/756d7c59-8641-42a5-94d0-2215f97ec7e5_en

Ein wichtiger Aspekt der Forschung sind die von Wilkinson et al. ([Wilkinson et al., 2016](#)) vorgeschlagenen Prinzipien der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit, die sog. FAIR-Principles, die mit ihren Anforderungen an die Strukturierung und Aufbereitung von Daten und Metadaten als Leitlinien zu verstehen sind.

Die eindeutige und präzise Definition von Metadaten ist von entscheidender Bedeutung und rückt zunehmend in den Fokus aktiver (Register-)Forschung ([Dugas et al., 2019](#); [Dugas et al., 2014](#); [Kersloot, Jacobsen, Groenen, & Rakov, 2018](#); [Lopprich, Jones, Meinecke, Goldschmidt, & Knaup, 2014](#); [Nicholson & Perego, 2020](#); [Scott, 2011](#); [Sinaci & Laleci Erturkmen, 2013](#); [Stausberg & Harkener, 2021](#); [Stausberg et al., 2022](#); [Stöhr, Günther, & Majeed, 2021](#); [Ulrich et al., 2022](#)). Allerdings werden hier aufgrund der Vielfalt der Definitionen, Formate, Standards und Kontexte häufig verschiedene Interpretationen des Begriffes Metadaten verwendet ([Ulrich et al., 2022](#)). Eine einheitliche, standardisierte Definition der Metadaten zur Beschreibung medizinischer Dokumentation gehört daher zu den dringendsten Aufgaben der Standardisierung im Gesundheitswesen und bildet eine der Grundlagen für den grenzüberschreitenden Austausch von Gesundheitsdaten.

Für den geplanten europäischen Gesundheitsdatenraum entwickelt TEHDAS, die gemeinsame Aktion für den Europäischen Gesundheitsdatenraum mit Vertretenden aus 25 Staaten, europäische Grundsätze für die Sekundärnutzung von Gesundheitsdaten, einschließlich der Aspekte Governance, Infrastruktur und Interoperabilität³. Darüber hinaus wird in einem von der Kommission geförderten Projekt mit dem Namen EHDS2Pilot⁴ an konkreten use cases überprüft, wie die EHDS-Verordnung konkret umgesetzt werden kann. Auch hier spielen die Metadaten sowie Interoperabilitätsfragen eine große Rolle. Das Projekt ist im Oktober 2021 gestartet, Ergebnisse liegen daher noch nicht vor. An den hier erarbeiteten Konzepten werden sich die europäischen Staaten orientieren, um eine Anbindung an den europäischen Gesundheitsdatenraum sicherstellen zu können.

In Deutschland gibt es bereits eine Reihe von Interoperabilitätsinitiativen, die sich mit den Herausforderungen der Harmonisierung von Gesundheitsdaten auseinandersetzen und an Konzepten zur Bewältigung arbeiten. Das Interop Council, ein Expertise-Gremium für Interoperabilität im deutschen Gesundheitswesen, das bei der gematik angesiedelt ist, hat im Jahr 2022 eine fragebogenbasierte Kartierung von Interoperabilitätsinitiativen mit Fokus auf FHIR und Terminologien vorgenommen. An dieser nicht-repräsentativen Befragung haben vier Register

³ <https://tehdas.eu/>

⁴ <https://www.ehds2pilot.eu/>

teilgenommen (von 85 teilnehmenden Organisationen)⁵. Im Fokus stehen in der Routineversorgung erhobene Daten in Bezug auf die Verwendung technischer Standards (FHIR) oder Terminologien. Metadaten bzw. Metadatenkataloge oder -Repositorien werden nicht adressiert. Die Medizininformatik Initiative (MII) hat mit dem Ziel, Daten aus der klinischen Versorgung einrichtungsübergreifend verfügbar zu machen bundesweit Datenintegrationszentren an den universitären Standorten aufgebaut und eine groß angelegte Interoperabilitätsinitiative gestartet. Die Datenintegrationszentren haben die Aufgabe, die Konzepte und Lösungen zu entwickeln und umzusetzen, die eine standortübergreifende gemeinsame Nutzung von klinischen Daten zu Forschungszwecken ermöglichen. Ein wesentlicher Grundbaustein sind neben der Konsentierung von einheitlichen Standards und standardisierten Datensätzen auch die automatisierte Harmonisierung von Daten aus der Versorgung mittels IT-basierter Diensten. Die in der Arbeitsgruppe Interoperabilität der MII identifizierten Themen und Fragen zu Metadaten werden in der Taskforce Metadaten bearbeitet⁶.

Die umfassende Beschreibung der verwendeten Datenelemente und Wertebereiche stellen eine der zentralen Aufgaben beim Aufbau von Registern dar. Im Betrieb von Registern ist die Weiterentwicklung der Datenelemente ein wichtiger zu verankernder Prozess zur Sicherstellung der Auswertbarkeit und Nutzbarkeit von Registerdaten. Eine Übersicht zu Metadaten-Repositorien im nationalen und europäischen Umfeld fehlt bislang.

2 Vorgehen

2.1 Literaturrecherche

Die Recherche relevanter Datenquellen erfolgte in erster Linie in PubMed und Google Scholar, weitere relevante Publikationen wurden mittels Desktoprecherche und Hinweisen aus dem Kreis der Expertinnen und Experten ergänzt. Die thematische Textanalyse der ausgewählten Literaturstellen erfolgte orientierend und kursorisch. Die Ergebnisse werden im Kontext in den einzelnen Kapiteln aufgeführt.

Sofern es sich um publizierte Buch- oder Journalbeiträge handelt, werden diese als Literaturstelle im Text in runden Klammern dargestellt und finden sich in alphabetischer Reihenfolge im Literaturverzeichnis (vgl. Kapitel 10.1). Juristische Verweise auf Gesetzestexte und -Initiativen etc. werden in Fußnoten angegeben. Webseiten und Referenzen auf Dokumente von Institutionen werden in Fußnoten (mit Link) zitiert. Auf Internetseiten zu Projekten (sofern keine Publikation vorliegt, dann wie oben beschrieben) wird ebenfalls in Fußnoten verwiesen.

⁵ InteropCouncil Arbeitskreis Kartierung der Interoperabilitätsinitiativen, <https://www.ina.gematik.de/fileadmin/Dokumente/Anhang - Auswertung der Umfrage des Arbeitskreises Kartierung von Interoperabilitätsinitiativen.pdf>

⁶ Z.B. Medizininformatik Initiative <https://www.medizininformatik-initiative.de/de/start>

2.2 Recherche laufender Aktivitäten und Projekte

Parallel zu der unter 2.1 beschriebenen Literaturrecherche erfolgte die Recherche laufender Aktivitäten und Projekte sowie aktueller Kataloge. Hier wurden ebenfalls die Kontakte und Vernetzungen der TMF sowohl auf nationaler als auch auf internationaler Ebene, zum einen als Ausgangspunkt der Recherche als auch als Quelle für weitere über die Ergebnisse der Literaturrecherche hinausgehende Erkenntnisse genutzt.

2.3 Durchführung eines Expertiseworkshops

Über die Literaturrecherche hinaus wurde gezielt externe Expertise in Form eines Workshops eingebunden. Tabelle 1 zeigt die Teilnehmenden des Expertiseworkshops am 18.11.2022.

Tabelle 1: Teilnehmende Einrichtungen Expertiseworkshop 18.11.2022

Name	Thema / Schwerpunkt
Prof. Dr. med. Dipl.-Inf. Martin Dugas	Direktor Institut für Medizinische Informatik, Universität Heidelberg
Prof. Dr. Juliane Fluck	AG Informationsmanagement, Institut für Geodäsie und Geoinformation, Universität Bonn Leitung Wissensmanagement, ZB MED - Informationszentrum Lebenswissenschaften
Dipl.-Inf. Matthias Löbe	Institut für Medizinische Informatik, Statistik und Epidemiologie, Medizinische Fakultät, Universität Leipzig
	DESAIC, FA für Anästhesiologie, Institut für Medizinische Informatik, Universität Heidelberg
	Direktor Institut für Medizinische Informatik – IMI, Universität Frankfurt wiss. Leitung Datenintegrationszentrum (DIZ)
	AG-Leitung Arbeitsgruppe „Data Exploitation & Integration“, Institut für Medizinische Informatik – IMI, Universität Frankfurt
	FAHSAI, Institut für Medizinische Informatik, Statistik und Epidemiologie, Medizinische Fakultät, Universität Leipzig

Die folgenden Fragen wurden dabei erörtert:

- Welche Arten von Metadatenkatalogen und -services gibt es in Deutschland?
- Wie sieht die Nutzung dieser Metadatenkataloge und -services aus und welche Herausforderungen ergeben sich für die Nutzung?
- Welche Form der Harmonisierung ist bereits gelungen – in Deutschland und in Europa?
- Welche nennenswerten Initiativen gibt es auch im europäischen Umfeld hierzu?
- Welche Wünsche und Handlungsempfehlungen ergeben sich für die bessere Nutzung von Metadatenkatalogen und -repositories?
- Welche weiteren Fragen und Anregungen sollten berücksichtigt werden?

Mit Zustimmung der Teilnehmenden wurde das Expertisegespräch aufgezeichnet. Die Aufzeichnung wird nach Abnahme des vorliegenden Berichts vernichtet.

3 Begriffsklärung und -bestimmung

Im Folgenden werden die im Gutachten verwendeten Begrifflichkeiten und Konstrukte erläutert und definiert. Da die internationale Literatur diese Begriffe nicht einheitlich verwendet, ist eine eindeutige Abgrenzung nicht immer einfach.

a) Metadaten

Nach einer allgemeinen Definition von Metadaten sind diese strukturierte Daten, die andere Daten beschreiben und Informationen über ihre Merkmale enthalten ([ISO/IEC, 2013](#)). Bei der Unterscheidung zwischen Daten und Metadaten bestimmt der zu erfüllende Zweck die Ebene und die Art der Daten, die das Ergebnis wiedergeben. Metadaten bilden eine spezifische Teilmenge der Dokumentationsdaten und dienen in erster Linie dazu, die Daten auffindbar zu machen, auch in bibliothekarischen Auskunftssystemen. Um sie maschinenlesbar zu machen, zum Beispiel in Semantic-Web-Anwendungen, werden sie häufig im XML-Format gespeichert.

Dabei sind verschiedene Arten von Metadaten zu differenzieren.

Strukturelle Metadaten

- Strukturelle Metadaten enthalten Angaben zu einem bestimmten Objekt oder einer Ressource. Sie enthalten auch Informationen über Klassifikationen und den Aufbau. Sie bestehen aus Definitionen von Merkmalen, Wertebereichen und Abhängigkeiten. Synonym zu Merkmal werden auch die Begriffe Datenelement, Item oder Variable verwendet.

Deskriptive Metadaten

- Beschreibende Metadaten dienen der Identifizierung von Daten und liefern grundlegende Informationen zu bestimmten Aspekten des betrachteten Datums.

Erhaltungsmetadaten

- Bewahrende oder auch Retentionsmetadaten dienen der Erhaltung eines Objekts und geben Auskunft darüber, wie ein bestimmtes digitales Objekt bzw. eine bestimmte Datei gepflegt und verwahrt werden kann.
- Erhaltungsmetadaten gewährleisten die Integrität.
- Diese Informationen können entscheidende Angaben enthalten, die erforderlich sind, damit ein System mit einer bestimmten Datei kommunizieren oder interagieren kann.

Herkunftsmetadaten / Provenienz

- Provenienzmeteradaten beschreiben die Herkunft eines Datenobjektes sowie die erfolgten Veränderungen.

Nutzungsmetadaten

- Beschreiben die Frequenz, Art und Zeitpunkt der Nutzung eines Datenobjektes.

Administrative Metadaten

- Beschreiben die Art von Anweisungen, Richtlinien und Einschränkungen, die für ein Datenobjekt gelten. Mithilfe dieser Daten können Admins den Zugriff auf Dateien je nach den Qualifikationen der Nutzenden einschränken.
- Dienen hauptsächlich der Verwaltung, d. h. der Steuerung und Vereinfachung komplexer Daten.

Interoperable Metadaten

- Metadaten, aus unterschiedlichen Quellen sind nicht notwendigerweise interoperabel. Hierfür sind Standards für Metadaten erforderlich, die neben den Aspekten Semantik und Syntax auch ein Datenmodell, d.h. die Struktur der Metadaten umfassen.

Auch die mit Metadaten verbundenen Begrifflichkeiten wie die Definition von Matching, Mapping und Transformation von Metadaten fallen in der Literatur sehr unterschiedlich aus. Dieses heterogene Verständnis führt zu Misinterpretationen und behindert die korrekte Nutzung der Metadaten sowie der entsprechenden Dateninstanz. Einen Vorschlag zur Einführung einer konsistenten Systematik zur Definition von Metadaten beschreibt eine aktuelle Arbeit eines Konsortiums von 10 Experten. Diese haben nach einer systematischen Literaturrecherche eine Konsultationsphase durchlaufen, an deren Ende harmonisierte Definitionen für Metadaten im Sinne einer metadatengestützten Datenintegration standen. Daraus wurde ein neues Schema für die Klassifizierung von Komponenten für angereicherte Metadatenobjekte entwickelt, das sich explizit auf den Erstellungskontext von Metadaten konzentriert. Diese harmonisierten Definitionen und das neue Schema ermöglichen ein verbessertes gegenseitiges Verständnis der Metadaten und ihres Kontexts ([Ulrich et al., 2022](#)).

Die Standardisierung des Metadatenvokabulars ist notwendig, um die Auffindbarkeit der Daten zu verbessern und Interoperabilität zu gewährleisten. Dies wird durch die Verknüpfung der Metadaten gewährleistet. Außerdem ermöglichen Standards eine einheitliche Beschreibung ähnlicher Datensätze in Bezug auf Inhalt und Struktur.

b) Metadatenstandards

Metadatenstandards enthalten eine definierte Auswahl von Informationen, die zum Auffinden und Identifizieren dieser Daten notwendig sind. Eine Wiederverwendbarkeit der Daten ist damit nicht unbedingt gewährleistet. Zu den gängigsten bibliographischen interdisziplinären Metadatenstandards gehören Dublin Core, DataCite Metadata Schema und MARC21.

Da jede wissenschaftliche Gemeinschaft ihre eigenen Anforderungen hat, werden auch verschiedene disziplinspezifische Metadatenstandards entwickelt. So wird beispielsweise in den Sozial- und Wirtschaftswissenschaften häufig der Standard der Data Documentation Initiative (DDI) verwendet, während in den Naturwissenschaften das ICAT-Schema oder das Crystallographic Information Framework zum Einsatz kommen. Eine Übersicht zu Metadatenstandards kategorisiert nach Themen bzw. Fachgebieten und nutzenden Einrichtungen sowie vorhandene Zuordnungen zwischen verschiedenen Schemata findet sich auf den Seiten der Research Data Alliance (RDA)⁷.

c) Metadaten-Repositories

Ein Metadata Repository (MDR) ist eine zentrale Datenbank oder ein Register von Datenelementen, mit dem diese auch verwaltet werden können. Das MDR enthält Metadaten zu Datenelementen, d.h. Merkmale, Wertemengen, Kodierungen, Verwendungskontexte (z.B. ihre Komposition in eCRFs), Abhängigkeiten etc.

Ein MDR hilft, die Erstellung des Dokumentationskonzepts zu beschleunigen und die Vergleichbarkeit der Merkmale zu verbessern. Es ermöglicht die Recherche nach Metadaten, die Harmonisierung von Metadaten und die Bereitstellung von Metadatenstandards.

Metadaten-Kataloge, -Repositories, -Registries oder Data-Dictionaries werden häufig synonym verwendet und speichern Metadaten zentral; zunehmend basieren Metadaten-Repositories auf dem ISO 11179 Standard ([ISO/IEC, 2013](#)).

Die Verwendung von Metadaten-Repositories unterstützen also

- (1) die Standardisierung durch Wiederverwendung bestehender Datenelemente,
- (2) die Formulierung von Anfragen auf der Grundlage der jeweiligen Datenelemente und
- (3) die Zusammenführung von Daten unterschiedlicher Datenquellen bei Verwendung gleicher Datenelemente.

d) Metadaten Service

Metadaten Services sind softwarebasierte Dienste, Prozeduren, Algorithmen und Routinen, die erfasste und dokumentierte medizinische Daten semantisch beschreiben können.

Metadaten Services werden für die Zusammenführung und gemeinsame Nutzung von Daten aus unterschiedlichen Datenquellen ebenso benötigt wie für die Kuratierung und Analyse. ([Lobe et al., 2019](#))

⁷ <https://rdamsc.bath.ac.uk>

e) Terminologien

Terminologien sind Vokabularien, die explizit erlauben, enthaltene Termini auf Begriffe zu beziehen (Synonymie) und auf Basis der Begriffsbeziehungen ein hierarchisches Begriffsordnungssystem im Sinne einer Taxonomie anzubieten. Sie stellen neutrale polyhierarchische Begriffshierarchien basierend auf (formal repräsentierten) Begriffsbedeutungen dar und sind durch Komposition neuer Begriffe erweiterbar (Postkoordination) ([Rienhoff & Semler, 2015, Seite 42](#)).

f) Terminologie Services

Zur Herstellung von semantischer Interoperabilität in verteilten Systemen werden Terminologie Services auf Terminologieservern, die verschiedene Terminologien in standardisierter Form zur Verfügung stellen, eingesetzt. Die Services ermöglichen die Synchronisierung von lokaler Semantik mit global vereinbarter Semantik und helfen diese aktuell zu halten. Die bereitgestellten Services können einerseits Dienste für den manuellen oder automatischen Abruf von Vokabularen, andererseits für die automatisierte Synchronisation dezentral verwalteter Vokabulare genutzt werden. Über eine Web-Anbindung können Vokabulare veröffentlicht werden: Primärsysteme können ihre lokalen Terminologien via Webservices synchronisieren und pflegen, und der Zugriff auf Terminologien kann über eine Weboberfläche ermöglicht werden. Voraussetzung ist, dass die Daten eine hohe Strukturierung aufweisen und sich automatisch verarbeiten lassen. ([De Moor et al., 2015](#); [Lobe et al., 2019](#); [Sasse, Darms, & Fluck, 2022](#))

4 Metadaten in Deutschland

In Deutschland gibt es verschiedene Initiativen und Projekte, die sich den Aufbau von Metadaten-Repositories zum Ziel gesetzt haben und von denen einige bereits seit Jahren umfangreiche Sammlungen von Metadaten aufgebaut haben.

Tabelle 2: Metadatenkataloge und -repositories in Deutschland

Bezeichnung	Kennzeichen	Referenzen
CoMetaR	<ul style="list-style-type: none"> • Deutsches Zentrum für Lungenforschung • Datenintegration seit 2013 • Aktuell 44 Register/Datensammlungen im zentralen Data Warehouse • 3.474 annotierte Begriffe • i2b2 Abfrageoberfläche für Forschende • Collaborative Metadata Repository (CoMetaR) 	https://data.dzl.de/cometar/ (Stöhr et al., 2021)
MDM-Portal	<ul style="list-style-type: none"> • 22.500 medizinische Formulare im CDISC Operational Data Model (ODM) Format • insgesamt über 500.000 Datenelemente, über 90% semantisch annotiert • 18 verschiedene Download-Formate, u.a. CDISC Operational Data Model (ODM), FHIR-, Statistik- und Office-Formate 	www.medical-data-models.org (Dugas et al., 2019 ; Dugas, Meidt, Neuhaus, Storck, & Varghese, 2016 ; Dugas, Neuhaus, et al., 2016)

Bezeichnung	Kennzeichen	Referenzen
NFDI4Health	<ul style="list-style-type: none"> Einsatz von SEEK NFDI4Health Task Force COVID-19 Metadata Schema V2_0 Publikation im German Central Health Study Hub COVID-19 	https://www.nfdi4health.de https://repository.publisso.de/re-source/fri:6431357 (Fluck et al., 2021; Lindstädt & Shutsko, 2022)
PaDaWaN (Patient Data Warehouse Navigator)	<ul style="list-style-type: none"> GUI des Data Warehouse am UKW Der PaDaWaN ist im Intranet der Universitätsklinik Würzburg verfügbar und nicht öffentlich zugänglich. 	https://www.informatik.uni-wuerzburg.de/is/open-source-tools/padawan-data-query-tool/ (Dietrich et al., 2018; Fette et al., 2017)
Samply.MDR	<ul style="list-style-type: none"> MDR-Nutzung in der German Biobank Node (GBN) Samply.MDR-basiertes ETL in ADOPT BBMRI-ERIC MDR-Nutzung in MIRACUM (i2b2, li2b2, Data Quality) 	https://bitbucket.org/medicalinformatics/workspace/projects/META-DAT (Kadioglu et al., 2018)

a) CoMetaR

Das frei zugängliche Collaborative Metadata Repository (CoMetaR) bietet Informationen zu Phänotypen und Spezimen. Zum Einsatz kommt eine zentrale Metadaten-Infrastruktur mit harmonisierter Ontologie. Die in lokalen Registern und Datenbanken gespeicherten Informationen zu Personen und die in Biobanken verfügbaren Bioproben wurden in einem zentralen Datawarehouse (DWH) zusammengeführt. Forschenden innerhalb des Deutschen Zentrums für Lungenerkrankungen (DZL) sowie externen Kooperationspartnern soll darüber ein einfacher und regelkonformer Zugriff auf Bioproben und Daten ermöglicht werden. Forschende können dort nach Beantragung eines Zugangs gezielt nach Bioproben und klinischen Phänotypen zu suchen. CoMetaR enthält 3.474 verschiedene Begriffe und unterstützt die Speicherung eines einzelnen Thesaurus im Resource Description Framework (RDF)-Format, das auf dem Simple Knowledge Organization System (SKOS) und dem Dublin Core (DC) Wissensorganisationssystem basiert (Stohr, Gunther, & Majeed, 2021; Stöhr et al., 2021).

b) MDM-Portal

Das mit Mitteln der Deutschen Forschungsgemeinschaft (DFG) von den Universitäten Münster und Heidelberg aufgebaute und betriebene MDM-Portal (Medizinische Datenmodelle) ist ein Metadaten-Register zur Erstellung, Analyse, Freigabe und Wiederverwendung von medizinischen Formularen. Nach Anmeldung steht die Infrastruktur für die akademische, nicht-kommerzielle medizinische Forschung gemeinfrei zur Verfügung. Das Portal enthält Formulare in dem systemunabhängigen CDISC Operational Data Model (ODM) Format mit mehr als 500.000 Datenelementen. Es besteht die Möglichkeit, Formulare anzuschauen, zu diskutieren und in

verschiedene technische Formate zu exportieren (PDF, CSV, Excel, SQL, SPSS, R, etc.) ([Dugas et al., 2019](#); [Dugas, Meidt, et al., 2016](#); [Dugas, Neuhaus, et al., 2016](#)).

c) NFDI4Health

NFDI4Health verfolgt das Ziel, Metadaten zu Studien und ihren Erhebungen zu charakterisieren und zentral zugänglich zu machen. Personenbezogene Studiendaten verbleiben dezentral in den verschiedenen Repositorien der erhebenden Organisationen und Institutionen. Für die Auswertung der Daten wird NFDI4Health Infrastrukturen und Algorithmen nutzen und weiterentwickeln, die einer verteilten Datenspeicherung Rechnung tragen und eine verteilte Datenanalyse erlauben.

Das NFDI4Health Task Force COVID-19 Metadata Schema Version 2_0 (Metadata Schema) enthält eine Liste von Eigenschaften, die eine Ressource beschreiben. Derzeit werden zwei Haupttypen von Ressourcen unterschieden: a) Studienbeschreibungen (d.h. Metadatensätze, die eine Studie beschreiben) und b) Studiendokumente.

Das Metadatenschema besteht hauptsächlich aus Eigenschaften, die von etablierten Standards und Modellen wie DataCite Metadata Schema, Datenmodellen des ClinicalTrials.gov, German Clinical Trials Register, International Clinical Trials Registry, HL7® FHIR, MIABIS, Maelstrom Research cataloguing toolkit und DDI Controlled Vocabularies übernommen wurden.

SEEK ist eine Reihe von Tools zur Unterstützung der Verwaltung, gemeinsamen Nutzung und Erforschung von Daten und Modellen in der Systembiologie. Die SEEK-Plattform bietet Forschenden eine zugangskontrollierte, webbasierte Umgebung für die gemeinsame Nutzung und den Austausch von Daten und Modellen für die alltägliche Zusammenarbeit und die öffentliche Verbreitung. Werkzeuge zur Durchführung von Modellsimulationen, zur Darstellung von Versuchsdaten und zur Unterstützung bei der Annotation und Standardisierung von Daten ergeben zusammen eine Sammlung von Ressourcen, die sowohl die Analyse als auch die gemeinsame Nutzung unterstützen. Die zugrunde liegenden semantischen Web-Ressourcen extrahieren zusätzlich SEEK-Metadaten in RDF (Resource Description Format) und stellen sie bereit. SEEK RDF ermöglicht umfangreiche semantische Abfragen, sowohl innerhalb von SEEK als auch zwischen verwandten Ressourcen im Web der Linked Open Data ([Lobe et al., 2022](#); [Wolstencroft et al., 2015](#)).

d) PaDaWaN - Patient DataWarehouse Navigator

Padawan ist ein klinikinternes Metadaten-Repository der Universität Würzburg. Es eignet sich zur Auswertung von Daten mit Bezug zu einem Fokus, z.B. Behandelten in einer Klinik, eines

klinischen Registers oder einer Studie. Diese Daten werden mit einem ETL-Prozess in eine Datenbank geladen und können dann mit einer GUI abgefragt werden. Die Analyse textueller Daten wird unterstützt. Es handelt sich um ein Open-Source-Framework, das mit Unterstützung des Deutschen Zentrums für Herzinsuffizienz (DZHI) seit 2011 als Data Warehouse entwickelt wird ([Dietrich et al., 2018](#); [Fette et al., 2017](#)).

e) Samply.MDR

Samply.MDR wurde im Jahr 2012 als erweiterbare und integrierbare MDR-Implementierung entwickelt, um die Spezifikation von Datenelementen auf eine strukturierte, formalisierte und standardisierte Weise zu unterstützen. Die Entwicklung erfolgte im Rahmen der IT für die Clinical Communication Platform (CCP-IT) des Deutschen Krebskonsortiums (DKTK). Durch zusätzliche Anforderungen im Rahmen der Entwicklung eines Open-Source-Registersystems für Seltene Erkrankungen (OSSE) wurde der PHP-basierte Prototyp - in Übereinstimmung mit den anderen Komponenten des Samply-Toolset - in JSF re-implementiert und im November 2015 als Open-Source-Software veröffentlicht. Seitdem findet Samply.MDR in mehreren Projekten für die Bereitstellung von Metadaten für verschiedene Systeme (z.B. i2b2, li2b2) Verwendung ([Kadioglu et al., 2018](#)).

Samply ist eine Sammlung von IT-Werkzeugen und -Komponenten zur Unterstützung der Zusammenarbeit und des Datenaustauschs in Forschungsnetzwerken ([Lablans, Kadioglu, Muscholl, & Ückert, 2015](#)). Sie setzt unter anderem Teile der ISO 11179 um. Samply ist als Open Source veröffentlicht⁸, um eine hersteller- und implementierungsunabhängige Daten- und Systemintegration zu ermöglichen.

5 Metadaten in Europa

Die Joint Action (JA) Towards the European Health Data Space (TEHDAS) ist das politische Entwicklungsinstrument, das für die Entwicklung von Optionen für gemeinsame Rahmenbedingungen zur Unterstützung der grenzüberschreitenden Sekundärnutzung von Gesundheitsdaten verantwortlich ist. Ziel ist es, die Europäische Kommission (EK) bei ihrem Legislativvorschlag für den Europäischen Gesundheitsdatenraum für die Sekundärnutzung (EHDS2) zu unterstützen. Ziel des EHDS2 ist es, dass die europäischen Bürgerinnen und Bürger, Gemeinschaften und Unternehmen in Zukunft von einem sicheren und nahtlosen Zugang zu Gesundheitsdaten profitieren, unabhängig davon, wo diese gespeichert sind. Die TEHDAS JA startete im Februar 2021 und läuft bis zum 1. August 2023.

⁸ <https://bitbucket.org/medicalinformatics/>

Innerhalb der TEHDAS JA wird das Arbeitspaket 7 (WP7) "Connecting the dots" die technischen Optionen für eine effektive Sekundärnutzung von Gesundheitsdaten durch den Europäischen Gesundheitsdatenraum für die Sekundärnutzung von Gesundheitsdaten (EHDS2) detailliert beschreiben. TEHDAS definiert die Sekundärnutzung von Daten als "wenn Daten für einen anderen Zweck verwendet werden als den, für den sie ursprünglich erhoben wurden".

Die Datenfindungsphase ist der Ausgangspunkt des gesamten User-Journey-Prozesses und somit der erste Kontaktpunkt zwischen Datennutzenden und dem EHDS2. In dieser Phase suchen Datennutzende nach den Daten, die zur Beantwortung einer Forschungsfrage benötigt werden. Voraussetzung für die Durchführung dieser Suche ist, dass die bei den verschiedenen für die Datenverarbeitung Verantwortlichen oder bei den Verarbeitenden verfügbaren Daten im EHDS2 über durchsuchbare Metadatenkataloge veröffentlicht wurden. Die erforderlichen Publikationsdienste fassen die Softwareelemente zusammen, die von den Datenverarbeitern und den Knotenpunkten verwendet werden, um die Metadaten-Deskriptoren der Originaldaten zu kuratieren, sie in einem Metadaten-Katalog zu organisieren und diesen Katalog an einem bestimmten Ort zu platzieren, der später über die Suchdienste zugänglich sein wird. Die Dienste für die Veröffentlichung von Metadaten sind eng mit dem Rahmen für die Bewertung der Datenqualität (Data Quality Assessment Framework, DQAF) verknüpft.

Die Nutzung der Metadatenveröffentlichungsdienste ist also eine Vorbedingung für die Datenermittlungsphase. Ohne die Veröffentlichung der Metadaten-Deskriptoren wird die Auffindbarkeit der vorhandenen Daten nahezu unmöglich sein.

Für die Verortung ergeben sich verschiedene Optionen:

- (1) Zentralisiert: Metadaten werden in einem einzigen Metadatenkatalog gespeichert.
- (2) Verteilt: Die Metadaten der vorhandenen Daten werden in regionalen Knoten gespeichert, mit denen die Datenquellen verbunden sind und über föderierte Abfragen auffindbar sind.
- (3) Hybride Systeme: Eine übergeordnete Metadatenbeschreibung, z. B. die Anzahl der Datensätze in einer Sammlung und die Pathologien einer solchen Sammlung, wird in einem zentralen Knotenkatalog für Dienste gespeichert. Weitere detaillierte Metadaten, z. B. Alter der Betroffenen und Komorbiditäten, werden in den Knoten gespeichert, die die Datenkontrolleure oder -verarbeiter verbinden, die diese Daten speichern.

Tabelle 3: Beispiele für europäische und internationale Metadatenkataloge

Bezeichnung	Kennzeichen	Referenzen
BBMRI-ERIC	<ul style="list-style-type: none"> • Portal für Zugang zu Biobanken und biomolekularen Ressourcen • nationale Knotenpunkte in 21 Mitgliedstaaten • Tools: BBMRI-ERIC Directory, Locator und Finder, RD-Connect Biobank und Registry Finder 	https://www.bbMRI-eric.eu

Bezeichnung	Kennzeichen	Referenzen
	<ul style="list-style-type: none"> • Datenmodell des Verzeichnisses basiert auf MIABIS Core 2.0 	
CRMDR	<ul style="list-style-type: none"> • Stand 01/2021 bezieht der MDR Studiendaten von: <ul style="list-style-type: none"> ○ Clinical Trials.gov (CTG) (https://clinicaltrials.gov/) ○ EUCTR (https://www.clinicaltrialsregister.eu/) ○ ISRCTN (http://www.isrctn.com/) ○ ICTRP der WHO (https://www.who.int/ictcp/en/) • Suchportal über ECRIN Weiterentwicklung erfolgt über EOSC-life 	https://www.crmdr.org
EMIF	<ul style="list-style-type: none"> • Metadaten-Katalog der European Medical Information Framework (EMIF) • 18 bevölkerungsbasierte Datenpartner • 55 kohortenbasierte Datenpartner • 14 Länder • OMOP-CDM (europaweit 10 Datenbanken) • Einsatz von ATLAS (OHDSI) zur Durchführung wissenschaftlicher Analysen von standardisierten Gesundheitsdaten 	https://emif-catalogue.eu/about/
ERDRI.mdr	<ul style="list-style-type: none"> • Metadatenkatalog der ERDRI für seltene Erkrankungen • Basiert auf samplly.mdr • Set mit 16 Common Data Elements (CDE) • Entwicklung von domänenspezifischen Datenelementen (DCDE) 	https://eu-rd-platform.jrc.ec.europa.eu/mdr/
Federated EGA - European Genome-Phenome Archive	<ul style="list-style-type: none"> • Archivierung personenbezogener genetischer und phänotypischer Daten • Daten aus 4.500 Studien mit 6.800 Datensätzen und 15 PB Daten • REST-API Schnittstelle zur Durchführung von Suchen 	https://ega-archive.org/
Maelstrom Research cataloguing toolkit	<ul style="list-style-type: none"> • Kanadischer Metadatenkatalog, der in europäischen Kooperationsprojekten eingesetzt wird • Projekt von DataSHaPER • Eigens entwickelte Software Suite (ONYX, OPAL, ROCK, MICA, DataSHIELD, AGATE) • 1.612.606 gesammelte Variablen aus • 24 Netzwerken und • 372 individuellen Studien 	https://www.maelstrom-research.org/page/catalogue
MINERVA	<ul style="list-style-type: none"> • Stärkere Nutzung von Real-World-Daten in der Arzneimittelentwicklung: Metadaten für die Auffindbarkeit von Daten und die Replizierbarkeit von Studien • Metadata for data discoverability and study replicability in observational studies • 436 Variablen (241 davon vorrangig für regulatorische Zwecke) 	https://www.encepp.eu/encepp/view-Resource.htm?id=39323

a) BBMRI-ERIC

BBMRI-ERIC ist eine der größten Forschungsinfrastrukturen für die Gesundheitsforschung in Europa und bietet ein Portal für den Zugang zu Biobanken und biomolekularen Ressourcen, das von den nationalen Knotenpunkten in 21 Mitgliedstaaten und der IARC/WHO als internationaler Organisation koordiniert wird. BBMRI-ERIC zielt darauf ab, die Zugänglichkeit und Interoperabilität der bestehenden umfassenden Sammlungen von biologischen Proben aus verschiedenen (Teil-)Bevölkerungen Europas, einschließlich seltener Krankheiten, zu verbessern.

Um Biobanken und biomolekulare Ressourcen auffindbar zu machen, betreibt BBMRI-ERIC ein Portfolio von Werkzeugen: BBMRI-ERIC Directory, Locator und Finder, zusammen mit spezialisierten Tools wie RD-Connect Biobank und Registry Finder. Diese Werkzeuge ermöglichen es Forschenden, Biobanken und biomolekulare Ressourcen auf der Grundlage ihrer Metadaten zu finden, einschließlich aggregierter Informationen über Spenderinnen und Spender, verfügbarer

Daten und Proben oder virtuelle Kohorten zu bilden sowie die Verfügbarkeit mit Hilfe von datenschutzfreundlichen Abfragen von pseudonymisierten Daten bei den Quelleninstitutionen zu ermitteln. Das Datenmodell des Verzeichnisses basiert auf MIABIS Core 2.0 ([Merino-Martinez et al., 2016](#)) mit dem Konzept von Biobanken (Institutionen) und Sammlungen (von Daten und Proben).

b) CRMDR

Das MDR-System wurde ursprünglich von ECRIN (European Clinical Research Infrastructure Network) in Zusammenarbeit mit ONEDATA⁹ und INFN (Istituto Nazionale di Fisica Nucleare) in Bologna konzipiert und entwickelt. Dies geschah im Rahmen des H2020-Projekts eXtreme - DataCloud (XDC). Das MDR-System verwendet PostgreSQL als Datenspeicher. In Postgres werden die Daten in mehreren Datenbanken gespeichert, von denen jede in verschiedene Datenschemata oder Gruppen zusammengehöriger Tabellen unterteilt werden kann. Die erste Entwicklungsstufe umfasste die Integration der von ECRIN gesammelten Datenquellen, den Export als json-Datei-Metadaten in das OneData-Dateimanagementsystem sowie die Indexierung über Elastic Search und die Bereitstellung im Webportal ([Canham & Ohmann, 2016](#)). Im Rahmen des H2020-Projekts EOSC-Life unter Durchführung von ECRIN erfolgt die Ausweitung der Datenerfassung auf eine größere Anzahl und Vielfalt von Datenquellen sowie die Automatisierung der Funktionsweise des MDR. Die Daten werden nun auf von ECRIN verwalteten Servern (bereitgestellt von OVH) gesammelt. Ein von ECRIN verwaltetes Portal wurde entwickelt, um das ursprünglich von OneData bereitgestellte Portal zu ersetzen. Es ist geplant, dieses Portal mit den EOSC-Hub-Diensten der European Open Science Cloud zu integrieren. Das Metadatenschema wird kontinuierlich aktualisiert und erweitert und liegt in einer aktuellen Version 6.0 vor ([Canham, 2021](#)). Es ist geplant, das fertige Portal mit anderen Diensten im EOSC-Hub der European Open Science Cloud zu integrieren¹⁰.

c) EMIF

Der Metadaten-Katalog des European Medical Information Framework (EMIF) ermöglicht die Bona-fide-Recherche in bevölkerungsbasierten (z. B. aus elektronischen Gesundheitsakten (EHR) abgeleiteten, regionalen Datenbanken) und kohortenbasierten (vorwiegend im Bereich der Alzheimer-Krankheit) Datenquellen und dient insbesondere der Suche nach potenziellen Datenpartnern für Studien. Mit Stand Juni 2018 sind 18 bevölkerungsbasierte und 55 kohortenbasierte Datenpartner aus 14 Ländern, die im Rahmen von EMIF zusammenarbeiten, im Ka-

⁹ <https://onedata.org/>

¹⁰ [https://wiki.crmdr.org/index.php?title=Progress_\(EOSC_Life\)](https://wiki.crmdr.org/index.php?title=Progress_(EOSC_Life))

talog enthalten. Neben Metadaten-Informationen zu den Datenquellen wurden Tools zur Bewertung der Eignung der Datenquellen und der Initiierung neuer Forschungsstudien mit den Datenquellen in einer sicheren Umgebung entwickelt.

Über die EMIF-Plattform sind Daten von mehr als 62 Millionen europäischen Erwachsenen und Kindern mittels eines Zusammenschlusses von Gesundheitsdatenbanken und Kohorten aus sieben verschiedenen Ländern (DK, IT, NL, UK, ES, EE) in einen Informationsrahmen für die einheitliche Wiederverwendung integriert. Sie ist repräsentativ für die verschiedenen Arten bestehender Datenquellen wie z. B. bevölkerungsbezogene Register, krankenhausbezogene Datenbanken, Kohorten, nationale Register oder Biobanken. Der Katalog bietet vertikale Projekte durch Bildung von Communities, wie z. B. EMIF-Electronic Health Record Data (EMIF-EHR) und EMIF-Alzheimer's Disease (EMIF-AD).

EMIF wurde 2013 bis 2018 über Innovative Medicines Initiative Joint Undertaking (IMI-JU) gefördert. Das Projektkonsortium besteht aus 57 Teams aus dem akademischen Bereich, kleinen und mittleren Unternehmen sowie Mitgliedern des Europäischen Verbandes der Pharmazeutischen Industrie und ihrer Verbände (EFPIA).

Das Extraktionswerkzeug Jerboa Reloaded wurde im Rahmen des EMIF-Projekts entwickelt, um die Datenextraktion und -verarbeitung aus den EHR-Datenbanken zu unterstützen. Es wird in einem so genannten verteilten Netzwerkdesign verwendet, d. h. es führt De-Identifizierung, Verknüpfung, Analyse und Aggregation lokal an jedem Datenquellenstandort durch. Jerboa führt ein Skript aus, das alle Parameter eines bestimmten Studiendesigns enthält. Dies hat den Vorteil, dass die lokalen Analysen auf eine gemeinsame, standardisierte Weise durchgeführt werden und nicht von kleinen Unterschieden in der Implementierung durch lokale Statistikerinnen und Statistiker abhängen, und dass für jede Studie nur die für diese spezielle Studie erforderlichen Daten in einem analytischen Datensatz gemeinsam genutzt werden ([Trifan, van der Lei, Díaz, & Oliveira, 2018](#)).

d) ERDRI.mdr

European RD Registry Infrastructure (ERDRI) des Joint Research Centre (JRC) der EU dient dazu, die Integration heterogener Daten aus verschiedenen Registern für seltene Krankheiten zu erleichtern und die Grundlage für die Sicherstellung der semantischen Interoperabilität zwischen Registern für seltene Krankheiten (RD) zu bilden. ERDRI wird seit 2019 entwickelt. Kernelemente sind das Metadaten Repository ERDRI.mdr und das Registerverzeichnis ERDRI.DOR. Das Metadaten Repository ERDRI.mdr enthält Sammlungen von Metadaten, die die verwendeten Datenelemente eines Registers spezifizieren, einschließlich der Bezeichnung der verwendeten Datenelemente, ihrer Definition und Maßeinheiten. Langfristig wird die Harmonisierung der verwendeten Datenelemente und Datenstrukturen in bestehenden und zukünftigen Registern

angestrebt, hierzu wurden ein "Set gemeinsamer Datenelemente (Common Data Elements, CDE) für die Registrierung seltener Krankheiten"¹¹ sowie weitere domänenspezifische CDEs entwickelt, deren Verwendung durch Register die Standardisierung und semantische Interoperabilität unterstützt ([Abaza et al., 2022](#); [Groenen et al., 2021](#); [Schaaf et al., 2021](#)).

Das Registerverzeichnis ERDRI.dor bietet einen Überblick über die teilnehmenden Register mit ihren wichtigsten Merkmalen und Beschreibungen. Die Dateneingabe und -pflege erfolgt durch die Registerinhaber.

e) Europäisches Genom-Phänomen-Archiv (EGA)

Das im Jahr 2008 gestartete Europäische Genom-Phänomen-Archiv (EGA) dient der dauerhaften Archivierung und gemeinsamen Nutzung aller Arten von personenbezogenen genetischen und phänotypischen Daten, die aus biomedizinischen Forschungsprojekten stammen.

Im Jahr 2022 sind über 4500 Studien mit 6800 Datensätzen mit fast 15 PB sensibler menschlicher Daten in EGA archiviert. Die EGA-Website ermöglicht die Identifizierung relevanter Datensätzen durch die Suche nach öffentlich zugänglichen Metadaten auf verschiedene Weise, u.a. nach Freitext, Text, gesteuerten Listen, Zugängen und anderen Merkmalen¹². Die Suchmaschine berücksichtigt gängige Rechtschreibfehler, Groß- und Kleinschreibung und die meisten Interpunktionszeichen und schlägt auch ähnliche Begriffskombinationen mit einer höheren Anzahl von Ergebnissen vor, um die Effizienz der Suche zu erhöhen. Softwarebasierte Suchen sind über die Metadaten mit der EGA Metadata API möglich¹³.

EGA betreibt ein verteiltes Datenzugriffsmodell, bei dem Anfragen an die für die Datenverarbeitung Verantwortlichen und nicht an die EGA gestellt werden. Der Daten Controller besteht aus einer oder mehreren Personen in einem DAC, der Zugriffsanträge entgegen nimmt und sie auf der Grundlage der beabsichtigten Datennutzung prüft. Die Bedingungen werden in einer Datennutzungsvereinbarung (Data Access Agreement, DAA) festgelegt. Diese umfasst Datenverwaltungs- und Sicherheitsrichtlinien, Bedingungen für Veröffentlichungen oder Sperrvermerke sowie Einschränkungen der Datennutzung oder -weitergabe.

Das EGA enthält ausschließlich Daten, die von Einzelpersonen gesammelt wurden, deren Einverständniserklärungen die Freigabe der Daten nur für spezifische Forschungszwecke oder für Bona-fide-Forschende erlauben. Die EGA bietet die notwendige Sicherheit, um den Zugang zu kontrollieren und die Vertraulichkeit der Betroffenen zu wahren, während sie gleichzeitig denjenigen Forschenden und in der Klinik Tätigen Zugang gewährt, die zur Einsichtnahme in die

¹¹ Set of common data elements. European Commission. 2019.

https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en

¹² <https://ega-archive.org/howtosearch>

¹³ <https://ega-archive.org/metadata/how-to-use-the-api>

Daten berechtigt sind. In allen Fällen werden die Entscheidungen über den Datenzugang von der zuständigen datenzugangsberechtigten Stelle (DAO) und nicht von der EGA getroffen. Bei der DAO handelt es sich in der Regel um dieselbe Organisation, die das ursprüngliche Studienprotokoll genehmigt und überwacht hat, oder um eine von dieser genehmigenden Organisation benannten Stelle.

Der EGA-Verbund ist die wichtigste globale Ressource für die Entdeckung und den Zugang zu sensiblen humanen Omics-Daten und damit verbundenen Daten, die für die Sekundärnutzung freigegeben sind, über ein Netz nationaler Humandatenpeicher zur Beschleunigung der Krankheitsforschung und zur Verbesserung der menschlichen Gesundheit.

f) Maelstrom-Katalog

In Zusammenarbeit mit internationalen Forschungsnetzwerken hat Maelstrom Research einen Standardansatz zur Dokumentation und Verbreitung epidemiologischer Studienmetadaten entwickelt. Das standardisierte Studienbeschreibungsmodell und die entwickelte Ontologie zur Variablenklassifizierung werden zur Erstellung umfassender Kataloge mit Studienmetadaten in zahlreichen Kooperationsprojekten verwendet. Einige dieser netzspezifischen Kataloge sind in den Maelstrom-Katalog integriert, der webbasiert einsehbar und durchsuchbar ist. Er enthält umfassende Informationen über epidemiologische Forschungsnetze und Studien sowie über die von ihnen gesammelten Daten. Darüber hinaus bietet der Katalog auch Informationen über Harmonisierungsinitiativen und harmonisierte Daten, die von diesen Forschungsnetzwerken erzeugt wurden.

Verwendet wird die eigens entwickelte OBiBa-Software-Suite¹⁴, die aus einer Reihe von eigenständigen Anwendungen besteht (ONYX, OPAL, ROCK, MICA, DataSHIELD, AGATE) ([Doiron, Marcon, Fortier, Burton, & Ferretti, 2017](#)). Sie unterstützen verschiedene Datenmanagement-Aktivitäten einer Studie und können integriert werden, um ein umfassendes Informationsmanagement- und Analysesystem für einzelne Studien zu schaffen. Die Komponenten ermöglichen eine Datenharmonisierung und -föderation für Studiennetzwerke und werden auch in nationalen und europäischen Projekten eingesetzt ([Bergeron, Doiron, Marcon, Ferretti, & Fortier, 2018](#)).

Der Katalog verfügt über 1.612.606 gesammelte Variablen aus 24 Netzwerken und 372 individuellen Studien. In acht Harmonisierungsinitiativen wurden mit 34 Harmonisierungsprotokollen 9.355 Variablen harmonisiert¹⁵.

¹⁴ <https://www.obiba.org/>

¹⁵ <https://www.maelstrom-research.org/page/catalogue>

g) MINERVA

Im November 2020 wurde das Projekt "Strengthening Use of Real-World Data in Medicines Development: Metadata for Data Discoverability and Study Replicability" (MINERVA; EU PAS Register number EUPAS39322) initiiert. Der Schwerpunkt der Studie lag auf der Definition des Metadatensatzes, einschließlich der Einbindung von Interessengruppen, sowie auf der Entwicklung eines Leitfadens für bewährte Verfahren, der die Metadaten und Empfehlungen auf der Grundlage des Pilotprojekts beschreibt. Ziel des Projekts war auch die Pilotierung einer Metadatenliste¹⁶ und eines FAIR-Daten-Proof-of-Concept (POC)-Katalogs zur Unterstützung von Forschenden, Datenzugangspartnern und Nutzenden im Prozess von der Identifizierung und Machbarkeit bis zum Studiendesign und zur Durchführung von Studien sowie zur Dokumentation der Stärken und Grenzen der Studienergebnisse.

Hintergrund des Projektes ist die Empfehlung der Big Data Task Force der Leitungen der Arzneimittelagenturen und der Europäischen Arzneimittelagentur (HMA-EMA) "die Auffindbarkeit von Daten durch die Identifizierung von Metadaten zu fördern" und zu ermöglichen.

Ziel dieser Empfehlung ist es über die Identifizierung der wichtigsten Metadaten für regulatorische Entscheidungen über die Wahl der Datenquelle eine Stärkung der aktuellen Ressourcen-datenbank des Europäischen Netzes der Zentren für Pharmakoepidemiologie und Pharmakovigilanz (ENCePP) zu bewirken und die Anwendung der FAIR-Grundsätze zu fördern.¹⁷

Die Metadatenliste umfasste 436 Variablen, von denen 241, die als vorrangig für regulatorische Zwecke eingestuft wurden, im Rahmen des Pilotprojekts erfasst wurden. Der Proof-of-Concept-Katalog wurde in die folgenden Bereiche unterteilt: Institutionen, Datenquellen, Datenbanken, gemeinsame Datenmodelle, Netzwerke und Studien. Die 15 Datenquellen umfassten eine unterschiedliche Anzahl von Datenbanken, die von 1 bis 16 reichte. Die Vollständigkeit der qualitativen Metadaten variierte von Datenquelle zu Datenquelle. Die Empfehlungen wurden in einem Leitfaden zusammengestellt, der im EU-PAS-Register verfügbar ist.¹⁸

6 Probleme / Herausforderungen

Daten in medizinisch-wissenschaftlichen Registern werden in der Regel hochstrukturiert zu einem bestimmten Zweck erhoben ([Niemeyer et al., 2021](#)). Für einen europäischen Gesundheitsdatenraum ebenso wie für ein nationales Gesundheitsdatenökosystem stellen Register daher eine wichtige Quelle dar. Dieses Potential wird bislang nicht ausgeschöpft, nur etwa 5% der Register (eigene Auswertung Registergutachten, Stand 30.10.2021) geben an, dass ihre Daten

¹⁶ <https://www.encepp.eu/encepp/openAttachment/documents.otherDocument-2/45372>

¹⁷ https://www.hma.eu/fileadmin/dateien/HMA_joint/00-About_HMA/03-Working_Groups/Big_Data/HMA-EMA_Joint_Big_Data_Taskforce_Phase_II_report_Evolving_Data_Driven_Regulation.pdf

¹⁸ <https://www.rtihs.org/publications/minerva-metadata-data-discoverability-and-study-replicability-observational-studies>

für externe Forschungsanfragen genutzt werden. Eine der Ursachen liegt möglicherweise in der in vielen Fällen noch nicht umfassenden und konsequenten Umsetzung der FAIR-Kriterien ([Boeckhout, Zielhuis, & Bredenoord, 2018](#); [Holub et al., 2018](#); [Stausberg & Harkener, 2019](#); [Wilkinson et al., 2016](#)) in Registern.

Die Verwendung von einheitlichen CDEs aus standardisierten und öffentlich zugänglichen Metadatenkatalogen können Register dabei unterstützen, ihr Register auf eine Nachnutzung ihrer Daten durch externe Forschungsvorhaben vorzubereiten und z.B. die Durchführung einer Machbarkeitsanfrage erheblich zu vereinfachen, den Aufwand der Datenzusammenführung mit anderen Datenquellen zu erleichtern, wenn nicht sogar zu ermöglichen.

Nach Einschätzung der befragten Expertinnen und Experten (vgl. Tabelle 1) liegen die Ursachen u.a. in fehlender Kenntnis über bereits vorhandene und verfügbare gut beschriebene Metadaten und Datenelemente (CDEs) in Metadaten-Repositories. Doch auch die Verwendung von standardisierten CDEs stelle für die Register im Aufbau nicht notwendigerweise eine Entlastung dar, die erforderlichen Schritte zur Nutzung werden als aufwändig angesehen und ggf. vorhandene Werkzeuge der Repositories zur Erleichterung nicht genutzt, da sie einen erhöhten Einarbeitungsaufwand erfordern. Die fehlende Kenntnis beziehe sich z.T. auch auf die Datentypen und die in den Repositories verwendeten Spezifikationen der Datentypen. Darüber hinaus fehlten vielen Metadaten-Repositories Werkzeuge und technische Schnittstellen, die eine unmittelbare Anbindung an ein Register ermöglichen würden oder eine umfassende Unterstützung für semantische Annotation von Metadaten anböten ([Sasse et al., 2022](#)).

Ähnliches gelte für die Hinterlegung von Datenelementen in Repositories, hier würde in der Regel der ISO-Norm folgend eine Validierung des Elementes erwartet, diesen Aufwand nähmen nur wenige Register auf sich.

7 Verknüpfbarkeit von Metadaten

7.1 Interoperabilität von Metadaten / Standardisierung

Die meisten Formulare und Itemkataloge aus Forschungsprojekten und Gesundheitsstudien entsprechen nicht den FAIR-Prinzipien, sie sind weder leicht auffindbar noch wiederverwendbar. Dies liegt sicherlich zum einen an vorliegenden Beschränkungen zur Veröffentlichung, zum anderen aber daran, dass Daten nicht unter Interoperabilitätsgesichtspunkten veröffentlicht werden, d.h. ohne beschreibende Metadaten ([Dugas et al., 2015](#)). Die Verwendung unterschiedlicher Datenformate und der zugehörigen Metadatenformate im Gesundheitswesen führt zu einer Heterogenität der Metadatenelemente. Dies führt dazu, dass mitunter eine voll-

ständige Übereinstimmung der Datenfelder nicht erreicht werden kann. Sei es, dass die Datenformate unterschiedliche Anwendungsbereiche unterstützen oder dass sie unterschiedliche Freiheitsgrade bei der Entwicklung von Erweiterungen zulassen. ([Bonisch, Kesztyus, & Kesztyus, 2022](#)).

Die Standardisierung von Metadaten und die Verwendung von CDEs aus Metadaten-Repositories sind wichtige Schritte zur Herstellung von Interoperabilität von Registern. Eine aktuelle Arbeit des Duke Clinical Research Institute (DCRI) untersuchte an 38 nationalen klinischen Registern die Anwendung und Nutzung von Datenstandards, um Kandidaten für die vergleichende Bewertung und die Entwicklung von (standardisierten) Metadaten zu identifizieren. Nur sehr wenige klinische Konzepte (wie z.B. Geschlecht oder Tabakkonsum) wurden mit genau der gleichen Darstellung der Datenelemente von einem Register zum nächsten erfasst; es gab keine Konzepte, die in allen 38 Registern auf genau die gleiche Weise (d. h. übereinstimmend) erfasst wurden ([Tcheng et al., 2020](#)). Obgleich die Anzahl der befragten Register sicherlich nicht repräsentativ für die US-amerikanische Registerlandschaft sein kann und auch nicht übertragbar ist auf die deutsche Registerlandschaft, lässt sich doch feststellen, dass hier beim Aufbau der Register wenig Augenmerk auf die Herstellung von Interoperabilität (semantisch wie auch syntaktisch) gelegt wurde. Abgesehen von den dadurch entstehenden Herausforderungen für eine mögliche Datenübernahme aus klinischen Systemen, stellt dies auch für die Nachnutzung und Analyse zur Beantwortung registerübergreifender Forschungsfragen eine nicht zu unterschätzende Hürde dar. Zu dieser Schlussfolgerung kommt auch eine Untersuchung der US-amerikanischen Registerlandschaft aus dem Jahr 2017, die eine Ursache in der mangelnden Strukturierung und Standardisierung von klinischen Dokumentationssystemen sieht, aufgrund dessen die meisten Register in gemischten Datenerfassungsumgebungen arbeiteten und auf die manuelle Dateneingabe durch Abstraktion von Krankenakten angewiesen seien. Automatisch aus elektronischen Quellen extrahierte Daten, ließen sich in Registern in der Regel nur über maßgeschneiderte technische Schnittstellen erfassen, die die Quelldaten in kompatible Formate umwandeln und an das Register übertragen ([Blumenthal, 2017](#)). Die Bereitstellung von „maßgeschneiderten“ Schnittstellen, seien sie zum Zwecke des Daten-Im- oder Exports gedacht, sind für die wenigsten Register eine gangbare Lösung.

Ausgelöst durch den Bedarf der forschenden Industrie, Teilnehmende und klinische Daten für Studien zu gewinnen, wurde die europäische Plattform Electronic Healthcare Record for Clinical Research (EHR4CR)¹⁹ aufgebaut, die die Effizienz der Durchführung klinischer Studien verbessern soll. Als ein Haupthindernis für die Wiederverwendung von routinemäßig erfassten Daten

¹⁹ <https://www.imi.europa.eu/projects-results/project-factsheets/ehr4cr>

für die klinische Forschung wurde die Heterogenität der Informationssysteme im Gesundheitswesen identifiziert ([Daniel et al., 2014](#)). Für den Aufbau eines gemeinsamen "semantischen MDR", das einen Rahmen für die bidirektionale bzw. MDR-übergreifende Zuordnung bietet und föderierte Abfragen ermöglicht, wurden Mappings für 178 EHR4CR- und 199 SALUS-Metadanelemente in der semantischen MDR persistiert, mit dem Ziel, eine EU-weite rechenfähige semantische Integration von Daten aus der klinischen Versorgung, der klinischen Forschung und der Patientensicherheit zu entwickeln. Trotz der Bemühungen konstatiert die Arbeitsgruppe, dass die formale Darstellung multimodaler und mehrstufiger Daten zur Unterstützung der Dateninteroperabilität zwischen klinischen Forschungs- und Versorgungsbereichen immer noch eine Herausforderung darstellt. Das Terminologie Mapping in den Krankenhäusern sei der größte Engpass bei der Datenstandardisierung. Darüber hinaus erfordere die Herstellung von semantischer Interoperabilität innerhalb eines breiten internationalen Forschungsnetzes einen strengen Governance-Prozess, um die Qualität des Datenstandardisierungsprozesses zu gewährleisten ([Daniel et al., 2017](#)).

Für die Anwendung und Nutzung von standardisierten annotierten Metadaten im Zusammenhang mit der Interoperabilität von epidemiologischen Registern (Population-Based Patient Registries (PBPR)) wurde in einem Beitrag von Nicholson und Perego anhand eines praktischen Beispiels gezeigt, wie eine Lösung durch die Umsetzung eines bestehenden Rahmens auf der Grundlage des Konzepts föderierter, semantischer Metadatenregister erreicht werden kann. Ein wichtiges Merkmal, das die Verwendung dieses Frameworks motiviert, ist, dass es schrittweise und unabhängig innerhalb jeder Domäne von Patienten-Registern implementiert werden kann. Durch die Anwendung von Grundsätzen für verknüpfte offene Daten erweitert der Rahmen die ISO/IEC 11179-Norm, um sowohl syntaktische als auch semantische Interoperabilität von Datenelementen mit den Mitteln zur Spezifikation automatischer Extraktionskripte für den Abruf von Daten aus verschiedenen Inhaltsmodellen von Registern zu ermöglichen ([Nicholson & Perego, 2020](#)).

Der steigende Bedarf an gut beschriebenen und strukturierten Daten zeigt sich auch an der steigenden Anzahl von Projekten und Initiativen zur Unterstützung der gemeinsamen Nutzung von Daten und der Vielzahl der resultierenden verschiedenen Metadaten-Repositories. Diese sind so heterogen wie die Daten, die sie beschreiben, einige davon sind allgemeiner Art, andere spezifisch für Krankheitsbereiche. Der Komplexitätsgrad wird in dem Maße zunehmen, in dem die Notwendigkeit der gemeinsamen Nutzung von Daten akzeptiert wird, daran ändern auch einzelnen Initiativen zur Zusammenarbeit nichts. Daten und Dokumente werden daher auf eine Vielzahl von Repositorien und Speicherorten verteilt sein, von denen viele spezialisiert sind und bestimmten Forschungsgemeinschaften, geografischen Gebieten oder Institutionen dienen,

auch wenn einige größere, allgemeinere Repositorien entstehen. Die Dateien selbst liegen wahrscheinlich in einer Vielzahl von unterschiedlichen Formaten und Dateitypen vor. Das Zusammentragen der relevanten Daten, Papiere und Dokumente im Zusammenhang mit einer bestimmten Studie oder Überprüfung und sogar die Feststellung, was verfügbar ist und unter welchen Bedingungen, kann schwierig, zeitaufwändig und teuer werden.

Einen Ansatz ein gewisses Maß an Interoperabilität herzustellen könnte ein gemeinsames und flexibel erweiterbares Metadatenschema darstellen, wie es von Canham und Ohmann ([2016](#)) entworfen wurde. Das Schema setzt auf dem DataCite-Schema²⁰ auf und erweitert bzw. vereinfacht es. Die Metadatendefinition sollen so einfach wie möglich gehalten werden und sich auf die Beschreibung des Datenobjekts konzentrieren. Weitergehende Informationen zu den Objekten wie Studienattribute oder -deskriptoren können aufgenommen werden, ihre Form ist jedoch nicht vorgeschrieben. Auf diesem Schema können Softwarelösungen aufsetzen und dadurch Aufwände für die Erstellung und das Sammeln der Metadaten gesenkt und die Identifizierung relevanter Datenobjekte unterstützt werden. In Verbindung mit einem allgemeinen "Metadaten-Repository" könnte es sowohl einen lokalen als auch einen zentralen, öffentlichen Metadatensatz erzeugen, so dass die Informationen nicht mehr über eine API-Abfrage oder einen Datei-Import gesammelt werden müssten ([Canham & Ohmann, 2016](#)). Das beschriebene Metadatenschema wurde für die „protokollgesteuerte Forschung“ entwickelt und primär ist hier klinische Forschung gemeint. Auch Register sind im optimalen Fall (register)protokollgesteuert, eine Anpassung bzw. Anwendung für Register könnte erfolgreich sein. Limitierend ist hier jedoch, dass es im Gegensatz zur klinischen Forschung keine verbindlichen Vorgaben für die Protokolle gibt und der Effekt hinter den Erwartungen zurückbleiben könnte.

Eine Standardisierung von Metadaten ist auch über eine sog. Bottom-up-Standardisierung möglich, hier wird der Weg der Standardisierung über die Wiederverwendung begangen. Bei der Erstellung von Datensätzen und Formularen werden den Forschenden Metadaten aus einem Verzeichnis nach Häufigkeit ihrer Verwendung angeboten, dadurch entsteht eine Art communitygetriebene Standardisierung durch Einsatz der immer selben Items. Wenn eine Datenelementdefinition in vielen realen Umgebungen verwendet wird, deutet dies darauf hin, dass diese Definition bereits getestet wurde und gut akzeptiert ist. Darüber hinaus gibt es bereits viele Datensätze, die potenziell mit Daten aus einem neu konzipierten System, das eine solche Datenelementdefinition übernimmt, verglichen werden können. Häufige Metadatendefinitionen werden als qualitativ hochwertig mit erhöhtem Wiederverwendungspotenzial bewertet und höher eingestuft, äquivalente Definitionen werden automatisch erkannt und zusammengefasst, gespeichert wird nur eine einzige Kopie mit der Anzahl ihrer Vorkommen.

²⁰ <https://datacite.org/>

Diese „Pragmatische MDR“ hat in zwei Anwendungsfällen gezeigt, dass die Wiederverwendung von Datenelementen durch diesen Ansatz erleichtert wird und eine auf der Häufigkeit basierende Bewertung von Datenelementen als Indikator für deren Qualität gelten kann. Für komplexe und damit auch seltenere klinische Konzepte lässt sich in diesem ersten Piloten nicht dieselbe Schlussfolgerung ziehen. Die Verfasserinnen sehen daher auch keinen Ersatz der ressourcenintensiven Top-Down-Ansätze, sondern eine wertvolle Ergänzung und Orientierung ([Hegselmann et al., 2021](#)).

Nach Einschätzung der Expertisegruppe ist eine Einigung auf Standards in der Community national wie international ein wichtiger und langer Prozess. Die Auffindbarkeit von Registern und der enthaltenen deskriptiven Metadaten stellt einen ersten Schritt dar. Die Veröffentlichung von Variablenkatalogen aller auffindbarer Register könnte eine nächste Stufe sein. Entscheidend für die Verknüpfbarkeit von Metadaten von Registern untereinander oder mit anderen Datenquellen wird jedoch die Etablierung von Standardschnittstellen für Metadaten für den Im- und Export in einem strukturierten Format sein.

7.2 Zusammenführung von Metadaten

Eines der Ziele der Medizininformatik-Initiative ist die Bereitstellung von Daten aus der Routineversorgung gemäß den FAIR-Prinzipien. Ein wichtiger Bestandteil ist die Einrichtung von "Datenintegrationszentren" (DIZ oder DIC) an deutschen Universitätskliniken und Partnerinstitutionen zur Unterstützung spezieller klinischer Anwendungsfälle. Diese Zentren sollen zeigen, wie Daten, Informationen und Wissen aus dem Gesundheitswesen, der klinischen und biomedizinischen Forschung über die Grenzen von Standorten hinweg verknüpft werden können. Die technischen Konzepte der Konsortien sind bewusst unterschiedlich hinsichtlich Anwendungsfall, Umfang und technische Umsetzung. Für den Austausch zwischen den Konsortien sind die Zusammenführung und Harmonisierung der Metainformationen über die auszutauschenden Daten von außerordentlicher Bedeutung. Es sind daher Mapping-Modelle notwendig, um einen interkonsortialen Austausch von Metadatenelementen zu ermöglichen. Ein Beispiel bietet das Mapping-Modell zwischen dem CDISC ODM- und dem Smply-Metadatenansatz, der von einer großen Anzahl von Institutionen verwendet wird ([Kock-Schoppenhauer et al., 2018](#)).

Ein im Jahr 2021 gestartetes Projekt der Nachwuchsgruppe befasst sich mit dem Thema „FAIRe und Reliable Analysestrukturen in Medizinischen Datenintegrationszentren“ (FAIRMedDIC)²¹. Das Medical Data Integration Center (MeDIC) der Universitätsmedizin Göttingen (UMG) führt

²¹ <https://www.medizininformatik-initiative.de/de/konsortien/nachwuchsgruppen-fuer-die-medizininformatik>

medizinische Informationen und die dazugehörigen Metadaten aus Krankenhausinformationssystemen und Datenbanken der klinischen Forschung (u.a. Daten aus Studien und Registern, wie z.B. Fallberichte, Patientenberichte oder Befunde) in einem Data Warehouse zusammen. Diese liegen in unterschiedlichen Formaten vor. Beim Harvesting der Metadaten im Rahmen des Extract-Transform-Load-Prozesses werden die Metadaten über ein MeDIC-spezifisches Datenprotokoll extrahiert und geladen, um Duplikate zu vermeiden. Im Rahmen dieses Forschungsprojekts wurden die Spezifikationen der Datenformate CDISC, OMOP, openEHR und FHIR untersucht und alle zugehörigen Metadatenelemente extrahiert und gegenübergestellt. Die untersuchten Formate nutzen unterschiedliche Metadatenbeschreibungen und eine einfache Übertragung zwischen den Formaten führt unweigerlich zu Informationsverlusten. CDISC beispielsweise hat Strukturen hinzugefügt, die auf dem ODM-Format aufbauen und in Verbindung damit die Grundlage für die Studiendokumentation in der klinischen Versorgung bilden. OpenEHR wiederum ist für die Speicherung medizinischer Daten in einer EHR konzipiert, während FHIR für den Austausch von Daten zwischen verschiedenen Institutionen gedacht ist. OMOP hingegen bietet ein gemeinsames Datenformat, um Daten aus verschiedenen Datenbanken zu vereinheitlichen. Die Verfassenden konnten zeigen, dass keines der Datenformate alle Metadaten enthält, die für den erfolgreichen Betrieb des MeDIC zum Zwecke eines zuverlässigen Datenmanagements erforderlich sind. Es wurde daher ein spezifisches Konvergenzformat vorgeschlagen, das die beschriebenen Herausforderungen umgeht ([Bonisch et al., 2022](#)).

7.3 Kuratierung und Pflege von Metadaten

Ein wichtiger Aspekt der Verknüpfbarkeit von Metadaten ist die notwendige und ressourcenintensive Kuratierung und Annotierung von Metadaten, ohne die eine Interoperabilität nicht nachhaltig sichergestellt werden kann. Diese stellt nach wie vor eine Herausforderung dar, da sie bislang in der Regel ressourcenintensiv und manuell im Zusammenspiel mit den Verantwortlichen durchgeführt werden muss. Zur Steigerung der Akzeptanz dieses Prozesses in der Wissenschafts-Community werden daher Optionen erprobt, den Prozess der Informationssammlung zu verschlanken, etwa durch Programmierung geeigneter Tools zur Onlineeingabe ([Schmidt et al., 2021](#)).

8 Harmonisierungsinitiativen

Die retrospektive Harmonisierung von Metadaten ist ein aktuelles Thema in verschiedenen Forschungsprojekten.

Das EU-Innovative Medicine Projekt EH DEN²² organisiert offene Ausschreibungen für Datenpartner, um finanzielle Unterstützung für die Standardisierung ihrer Daten gemäß dem OMOP Common Data Model zu erhalten. Im Rahmen von EH DEN werden kleine und mittlere Unternehmen (KMU) für die Unterstützung der Datentransformation ausgebildet und finanziert.

Das weiter oben bereits vorgestellte Projekt Maelstrom nutzt Ansätze zur semantischen Annotation von Daten, um Forschungsdaten FAIR zu machen, so wurden Leitlinien zur Harmonisierung epidemiologischer Forschungsdaten veröffentlicht und ein Modell für Studienmetadaten sowie die Maelstrom-Klassifikation, die aus 18 Bereichen und 135 Unterbereichen besteht, erstellt. Die Datenannotationen werden von Sachverständigen vorgenommen und mit einem hausinternen automatischen Klassifikator auf der Grundlage einer maschinellen Lernmethode validiert ([Bergeron et al., 2018](#)). Das Ergebnis ist, dass die harmonisierten Daten über Open-Source-Softwareanwendungen wie OPAL und MICA ([Doiron et al., 2017](#)) durchsucht werden können.

Die Wiederverwendung biomedizinischer Daten für die Forschung ist eine anspruchsvolle Aufgabe, insbesondere wenn Daten aus verschiedenen Quellen zusammengeführt und harmonisiert werden sollen. Diese Herausforderungen ergeben sich zum einen aus syntaktischen und semantischen Unterschieden in den Quelldaten und zum anderen aus potenziellen Datenqualitätsproblemen.

Im Projekt ADOPT BBM RI-ERIC wurde am Beispiel Colon Cancer Data Collection (CCDC) ein mittels Algorithmen gesteuerter Extraktions-Transformations-Lade-Prozess (ETL)-Ansatz für die Integration heterogener Daten entwickelt. Die Identifizierung der korrekten Zuordnungen wird dabei durch einen lexikalischen Abgleichprozess in MDRMatcher²³ übernommen und durch einen Bag-of-Words-Algorithmus konnte in fast 80 % der Fälle eine vollautomatische, korrekte Zuordnung erzielt werden. Andererseits erwies sich der generische Ansatz der Datentypkonvertierung in ETLHelper als geeignet, um die nachfolgenden Datentransformationen durchzuführen und Fehler in den Biobankdaten zu erkennen ([Mate et al., 2019](#)).

Im Rahmen des MDR- Symposium 2020 wurden die unterschiedlichen Ansätze zur Harmonisierung unter dem Motto „Let’s make MDR simple!“ diskutiert²⁴. Vorgestellt wurde u.a. einheitliche Abfrageschnittstelle für Metadaten-Repositories, die den ISO 11179-Standard für Metadaten-Repositories und die Abfragesprache GraphQL kombiniert. Es wurde eine Referenzimplementierung auf der Grundlage des bestehenden Sampl y.MDR durchgeführt. Die Schnittstelle

²² <https://www.ehden.eu/open-calls/harmonisation-fund/>

²³ Der Quellcode der Tools, einschließlich der vollständigen Dokumentation und eines kleinen Demo-Datensatzes mit künstlichen Daten, ist unter der GPL3-Lizenz auf GitHub verfügbar: <https://github.com/seb-mate/ADOPT-BBMRI-ERIC-ETL-Tools>

²⁴ https://www.mdr-symposium.de/wp-content/uploads/2020/11/Hannes_Ulrich.pdf

erleichtert den Zugang zu Metadaten, ermöglicht eine bessere Interaktion mit Metadaten und ist eine Grundlage für die Verbindung bestehender Repositories ([Ulrich et al., 2019](#)).

Im Oktober und November 2021 veranstaltete die Europäische Kommission (EC) drei Workshops, um Handlungsprioritäten für FAIR-Daten für die COVID-19-Reaktion zu ermitteln und Empfehlungen für von der Europäischen Union (EU) finanzierten Projekte und Infrastrukturen für die Wiederverwendung von Gesundheitsdaten und die EC selbst abzuleiten. Der Bericht listet u.a. die Aktivitäten und Investitionen der EC in die „Fairification“ der Forschung und formuliert als Ergebnis die folgenden Schlüsselempfehlungen für die EC, die von der EU geförderten Forschungskonsortien und -infrastrukturen sowie die Open-Science-Gemeinschaft und die Politik:

- (1) Verbesserung und Koordinierung der Umsetzung von FAIR
- (2) Förderung der Qualität von Daten und Metadaten auf Teilnehmendenebene
- (3) Unterstützung von Verbindungen zwischen Datentypen auf der Ebene der Teilnehmenden
- (4) Entwicklung einer transparenten, ethischen und reaktionsfähigen Governance für die Wiederverwendung von Daten

Interoperable Daten auf Teilnehmendenebene und beschreibende Metadaten, die unter Verwendung von in der Gemeinschaft entwickelten Standards erfasst werden, sind die Voraussetzungen, um Daten wiederzuverwenden. Der Ansatz „FAIR by design“ ist kosteneffizienter und führt zu qualitativ hochwertigeren Daten als Daten, die nachträglich FAIRifiziert werden. „FAIR by design“ erfordert keine nachträgliche Transformation in einen interoperablen Standard. Erkannt wird jedoch auch, dass der Aufbau von Kapazitäten und die Bewältigung der Herausforderungen für die Austauschstandards Zeit und Ressourcen in Anspruch nehmen werden ([Commission, Innovation, & Maxwell, 2022](#)).

Einen ersten Ansatz hierfür – auch und gerade für neue Register – könnte die Verwendung von standardisierten EDC-Systemen sein, die sowohl interoperable Formate (syntaktisch) nutzen als auch die Wiederverwendung von Metadatenelementen ermöglichen, was zu einem vereinfachten Metadatenentwicklungsprozess und Herstellung von Datenkompatibilität in der Entwurfsphase führt ([Greulich, Heggemann, & Dugas, 2021](#)).

Die beiden europäischen Projekte EHES Pilot Project (2009-2012)²⁵ und BRIDGE Health Project (2015-2017)²⁶ dienten der Verbindung nationaler Kohorten und v.a. der europaweiten Auswertbarkeit nationaler Gesundheitsuntersuchungserhebungen. Die Europäische Gesundheitsuntersuchungserhebung (EHES) ist eine Zusammenarbeit zwischen den Organisierenden der

²⁵ <https://ehes.info/index.htm>

²⁶ <https://www.bridge-health.eu>

nationalen Gesundheitsuntersuchungserhebungen (HES) in Europa. EHES unterstützt den Aufbau von Kapazitäten in den EU-Mitgliedstaaten und zielt darauf ab, eine hohe Qualität und Vergleichbarkeit der Erhebungen zu gewährleisten. Die EHES umfasst Fragebögen sowie körperliche Messungen, z. B. des Blutdrucks, und die Entnahme von biologischen Proben, z. B. von Blut oder Urin.

Das Projekt BRIDGE Health zielte darauf ab, den Übergang zu einem nachhaltigen und integrierten EU-Gesundheitsinformationssystem sowohl für die öffentliche Gesundheit als auch für Forschungszwecke vorzubereiten. Das Projekt wurde im Mai 2015 gestartet und hatte eine Laufzeit von 30 Monaten. Es umfasste 31 Mitwirkende in 16 Ländern. Das Projekt vereinte EU-Projekte in den Bereichen Bevölkerungs- und Gesundheitssystemmonitoring, Entwicklung von Indikatoren, Erhebungen zu Gesundheitsuntersuchungen, Umwelt und Gesundheit, Register für Verletzungen und Krankheiten der Bevölkerung, klinische und administrative Systeme zur Erhebung von Gesundheitsdaten sowie Methoden zur Überwachung und Bewertung der Gesundheitssysteme.

9 Erste Impulse für Handlungsempfehlungen

Die klinische Forschung ist zunehmend abhängig von konsistent definierten Merkmalen, etwa bei der Wiederverwendung klinischer Daten aus der Versorgung in Studien, der Zusammenführung von Studiendaten in multizentrischen Studien oder gar der Durchführung von Metastudien. International gibt es erhebliche Anstrengungen, so genannte Metadaten-Repositories (MDR) bzw. offizielle Register zu etablieren, um relevante Datenelemente wie einen Blutdruck einheitlich zu benennen, zu definieren und zu annotieren (i.Allg. mit Codes aus standardisierten Vokabularien). Darauf basierend lassen sich Software-Dienste zur semantischen Datenintegration bereitstellen.

Die meisten der hier aufgeführten und vorgestellten Publikationen beziehen sich auf klinische Studien oder die Nutzung von Daten aus der stationären Routineversorgung. Mögliche Ursachen hierfür liegen sicherlich im akuten und stetig wachsenden Bedarf Daten aus diesen beiden Quellen zu Forschungszwecken nutzbar zu machen und der Vielzahl an nationalen und europäischen geförderten Initiativen mit genau diesem Auftrag.

Viele der gewonnenen Erkenntnisse lassen sich auf Register übertragen, andere müssen weiterentwickelt und angepasst werden, um für Register anwendbar zu sein. Auf dieser Ebene der Förderung der Weiterentwicklung und Standardisierung von Registern als Methode besteht ein wachsender Bedarf an übergeordneten Initiativen und Projekten. Insbesondere die Herstellung der Anbindbarkeit von Register an den EHDS setzt Vorarbeiten auf nationaler Ebene voraus.

Vor diesem Hintergrund ergeben sich erste Ideen für Handlungsschritte:

(1) Erhöhen der Sichtbarkeit von verfügbaren Metadaten

Diese Aufgabe könnte von einer Zentralstelle für medizinische Register (ZMR) übernommen werden. Ein erster Schritt könnte die Anfrage an die im Registerverzeichnis gelisteten Register sein, ihre Metadaten bzw. Datenelemente in strukturierter Form zur Verfügung zu stellen, hier sind verschiedene Optionen denkbar. Dazu gehören die Bereitstellung von Informationen, wo die Metadaten bereits verfügbar gemacht wurden, ggf. die direkte Bereitstellung über die ZMR und die Angabe der verwendeten Standards.

In weiteren Entwicklungsschritten könnte diese Bereitstellung und Veröffentlichung in geeigneter Form in den Bewertungskatalog einfließen und im Reifegradmodell berücksichtigt werden. Die bislang in ersten pilotären Durchgängen gemachten Erfahrungen mit Bewertungskatalog und Reifegradmodell lassen die Vermutung zu, dass diese Anforderungen von einigen Registern bereits erfüllt werden und dies wiederum auf andere Register motivierend wirken könnte.

(2) Verpflichtung zur Publikation von Metadaten der Register als Vorbedingung der Förderung von Registern

In einem weiteren Schritt könnte die Verknüpfung der Erfüllung von Interoperabilitäts- und Transparenzkriterien mit einer finanziellen Förderung beim Aufbau bzw. der Weiterentwicklung von Registern einen positiven Effekt auf die Herstellung der Sichtbarkeit und die Auseinandersetzung der Register mit ihrer Standardisierung haben.

(3) Nationale Strategieentwicklung, Terminologieserver für Forschung und Versorgung

Die aktuellen Überlegungen zu einem nationalen Terminologieserver sollten auch die Anbindung von Registern einbeziehen und auch Metadaten und Metadatenservices bereitstellen. Diese könnten auch von der ZMR für den Einsatz in Registern empfohlen werden.

10 Anhang

10.1 Literaturverzeichnis

- Abaza, H., Kadioglu, D., Martin, S., Papadopoulou, A., dos Santos Vieira, B., Schaefer, F., & Storf, H. (2022). Domain-Specific Common Data Elements for Rare Disease Registration: Conceptual Approach of a European Joint Initiative Toward Semantic Interoperability in Rare Disease Research. *JMIR Med Inform*, *10*(5), e32158. doi:10.2196/32158
- Bergeron, J., Doiron, D., Marcon, Y., Ferretti, V., & Fortier, I. (2018). Fostering population-based cohort data discovery: The Maelstrom Research cataloging toolkit. *PLoS One*, *13*(7), e0200926. doi:10.1371/journal.pone.0200926
- Blumenthal, S. (2017). The Use of Clinical Registries in the United States: A Landscape Survey. *EGEMS (Washington, DC)*, *5*(1), 26. doi:10.5334/egems.248. (Accession No. 29930965)
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, *26*(7), 931-936. doi:10.1038/s41431-018-0160-0
- Bonisch, C., Kesztyus, D., & Kesztyus, T. (2022). Harvesting metadata in clinical care: a crosswalk between FHIR, OMOP, CDISC and openEHR metadata. *Sci Data*, *9*(1), 659. doi:10.1038/s41597-022-01792-7
- Canham, S. (2021). *ECRIN Metadata Schema for Clinical Research Data Objects*.
- Canham, S., & Ohmann, C. (2016). A metadata schema for data objects in clinical research. *17*(1), 557. Retrieved from <https://doi.org/10.1186/s13063-016-1686-5>
- Commission, E., Innovation, D.-G. f. R., & Maxwell, L. (2022). *Maximising investments in health research: FAIR data for a coordinated COVID-19 response: workshop report*. Retrieved from
- Daniel, C., Ouagne, D., Sadou, E., Paris, N., Hussain, S., Jaulent, M. C., & Kalra, D. (2017). Cross border semantic interoperability for learning health systems: The EHR4CR semantic resources and services. *Learning Health Systems*, *1*(1). doi:10.1002/lrh2.10014
- Daniel, C., Sinaci, A., Ouagne, D., Sadou, E., Declerck, G., Kalra, D., . . . Laleci Erturkmen, G. B. (2014). Standard-based EHR-enabled applications for clinical research and patient safety: CDISC - IHE QRPH - EHR4CR & SALUS collaboration. *AMIA Jt Summits Transl Sci Proc*, *2014*, 19-25. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25954572>
- De Moor, G., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., . . . Coorevits, P. (2015). Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of Biomedical Informatics*, *53*, 162-173. doi:10.1016/j.jbi.2014.10.006
- Dietrich, G., Krebs, J., Fette, G., Ertl, M., Kaspar, M., Stork, S., & Puppe, F. (2018). Ad Hoc Information Extraction for Clinical Data Warehouses. *Methods Inf Med*, *57*(1), e22-e29. doi:10.3414/ME17-02-0010
- Doiron, D., Marcon, Y., Fortier, I., Burton, P., & Ferretti, V. (2017). Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonization and dissemination. *International Journal of Epidemiology*, *46*(5), 1372-1378. doi:10.1093/ije/dyx180
- Dugas, M., Hegselmann, S., Riepenhausen, S., Neuhaus, P., Greulich, L., Meidt, A., & Varghese, J. (2019). Compatible Data Models at Design Stage of Medical Information Systems: Leveraging Related Data Elements from the MDM Portal. *Stud Health Technol Inform*, *264*, 113-117. doi:10.3233/SHTI190194
- Dugas, M., Jöckel, K., Gefeller, O., Knaup-Gregori, P., Friede, T., Ammenwerth, E., . . . Prokosch, H. U. (2014). Freier Zugang zu Dokumentationsformularen und Merkmalskatalogen im Gesundheitswesen. Memorandum „Open Metadata“. *GMS Med Inform Biom Epidemiol*, *10*(1). doi:10.3205/mibe00015

- Dugas, M., Jöckel, K. H., Friede, T., Gefeller, O., Kieser, M., Marscholke, M., . . . Prokosch, H. U. (2015). Memorandum "Open Metadata". Open Access to Documentation Forms and Item Catalogs in Healthcare. *Methods Inf Med*, 54(4), 376-378. doi:10.3414/ME15-05-0007
- Dugas, M., Meidt, A., Neuhaus, P., Storck, M., & Varghese, J. (2016). ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository. *BMC Med Res Methodol*, 16, 65. doi:10.1186/s12874-016-0164-9
- Dugas, M., Neuhaus, P., Meidt, A., Doods, J., Storck, M., Bruland, P., & Varghese, J. (2016). Portal of medical data models: information infrastructure for medical research and healthcare. *Database (Oxford)*, 2016. doi:10.1093/database/bav121
- Fette, G., Kaspar, M., Dietrich, G., Ertl, M., Krebs, J., Stoerk, S., & Puppe, F. (2017). A Customizable Importer for the Clinical Data Warehouses PaDaWaN and I2B2. *Stud Health Technol Inform*, 243, 90-94. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28883177>
- Fluck, J., Lindstädt, B., Ahrens, W., Beyan, O., Buchner, B., Darms, J., . . . Pigeot, I. (2021). NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten. *Bausteine Forschungsdatenmanagement*(2), 72-85. doi:10.17192/bfdm.2021.2.8331
- Greulich, L., Hegselmann, S., & Dugas, M. (2021). An Open-Source, Standard-Compliant, and Mobile Electronic Data Capture System for Medical Research (OpenEDC): Design and Evaluation Study. *JMIR Med Inform*, 9(11), e29176. doi:10.2196/29176
- Groenen, K. H. J., Jacobsen, A., Kersloot, M. G., dos Santos Vieira, B., van Enckevort, E., Kaliyaperumal, R., . . . Kool, L. S. (2021). The de novo FAIRification process of a registry for vascular anomalies. 16(1), 376. Retrieved from <https://doi.org/10.1186/s13023-021-02004-y>
- Hegselmann, S., Storck, M., Gessner, S., Neuhaus, P., Varghese, J., Bruland, P., . . . Dugas, M. (2021). Pragmatic MDR: a metadata repository with bottom-up standardization of medical metadata through reuse. *BMC Med Inform Decis Mak*, 21(1), 160. doi:10.1186/s12911-021-01524-8
- Holub, P., Kohlmayer, F., Prasser, F., Mayrhofer, M. T., Schlünder, I., Martin, G. M., . . . Litton, J. E. (2018). Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. *Biopreservation and Biobanking*, 16(2), 97-105. doi:10.1089/bio.2017.0110
- 11179-3:2013(E) Information technology – Metadata registries (MDR) – Part 3: registry metamodel and basic attributes, Third edition, (2013 2013-02-15).
- Kadioglu, D., Breil, B., Knell, C., Lablans, M., Mate, S., Schlue, D., . . . Prokosch, H. U. (2018). Sampl.y.MDR - A Metadata Repository and Its Application in Various Research Networks. *Stud Health Technol Inform*, 253, 50-54. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30147039>
- Kersloot, M. G., Jacobsen, A., Groenen, K., & Rakov, D. (2018). Real-time FAIRification of rare disease patient registry data.
- Kock-Schoppenhauer, A. K., Ulrich, H., Wagen-Zink, S., Duhm-Harbeck, P., Ingenerf, J., Neuhaus, P., . . . Bruland, P. (2018). Compatibility Between Metadata Standards: Import Pipeline of CDISC ODM to the Sampl.y.MDR. *Stud Health Technol Inform*, 247, 221-225. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/29677955>
- Lablans, M., Kadioglu, D., Muscholl, M., & Ückert, F. (2015). Exploiting Distributed, Heterogeneous and Sensitive Data Stocks while Maintaining the Owner's Data Sovereignty. *Methods Inf Med*, 54(4), 346-352. doi:10.3414/me14-01-0137
- Lindstädt, B., & Shutsko, A. (2022). Publication Policy of the National Research Data Infrastructure for Personal Health Data (NFDI4Health) and the NFDI4Health Task Force COVID-19. In: Zenodo.
- Lobe, M., Beyan, O., Staubert, S., Meineke, F., Ammon, D., Winter, A., . . . Kirsten, T. (2019). Design of Metadata Services for Clinical Data Interoperability in Germany. *Stud Health Technol Inform*, 264, 1528-1529. doi:10.3233/SHTI190518
- Lobe, M., Ulrich, H., Beger, C., Bender, T., Bauer, C., Sax, U., . . . Winter, A. (2022). Improving Findability of Digital Assets in Research Data Repositories Using the W3C DCAT Vocabulary. *Stud Health Technol Inform*, 290, 61-65. doi:10.3233/SHTI220032

- Lopprich, M., Jones, J., Meinecke, M. C., Goldschmidt, H., & Knaup, P. (2014). A reference data model of a metadata registry preserving semantics and representations of data elements. *Stud Health Technol Inform*, 205, 368-372. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25160208>
<https://ebooks.iospress.nl/pdf/doi/10.3233/978-1-61499-432-9-368>
- Mate, S., Kampf, M., Rödle, W., Kraus, S., Proynova, R., Silander, K., . . . Prokosch, H. U. (2019). Pan-European Data Harmonization for Biobanks in ADOPT BBMRI-ERIC. *Appl Clin Inform*, 10(4), 679-692. doi:10.1055/s-0039-1695793
- Merino-Martinez, R., Norlin, L., van Enckevort, D., Anton, G., Schuffenhauer, S., Silander, K., . . . Litton, J.-E. (2016). Toward Global Biobank Integration by Implementation of the Minimum Information About Biobank Data Sharing (MIABIS 2.0 Core). *Biopreservation and Biobanking*, 14(4), 298-306. doi:10.1089/bio.2015.0070
- Nicholson, N., & Perego, A. (2020). Interoperability of population-based patient registries. *Journal of Biomedical Informatics: X*, 6-7(May), 100074. doi:10.1016/j.yjbinx.2020.100074
- Niemeyer, A., Semler, S. C., Veit, C., Hoffmann, W., van den Berg, N., Röhrig, R., . . . Beckedorf, I. (2021). *Gutachten zur Weiterentwicklung medizinischer Register zur Verbesserung der Dateneinspeisung und -anschlussfähigkeit*. Retrieved from Hamburg, Berlin: <https://www.bundesgesundheitsministerium.de/service/publikationen/details/gutachten-zur-weiterentwicklung-medizinischer-register-zur-verbesserung-der-dateneinspeisung-und-anchlussfaehigkeit-1.html>
- Rienhoff, O., & Semler, S. C. (Eds.). (2015). *Terminologien und Ordnungssysteme in der Medizin - Standortbestimmung und Handlungsbedarf in den deutschsprachigen Ländern* (Vol. Band 13). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Sasse, J., Darms, J., & Fluck, J. (2022). Semantic Metadata Annotation Services in the Biomedical Domain: A Literature Review. *Applied Sciences*, 12(2), 796. Retrieved from <https://www.mdpi.com/2076-3417/12/2/796>
- Schaaf, J., Chalmers, J., Omran, H., Pennekamp, P., Sitbon, O., Wagner, T. O. F., . . . Storf, H. (2021). The Registry Data Warehouse in the European Reference Network for Rare Respiratory Diseases - Background, Conception and Implementation. *Stud Health Technol Inform*, 278, 41-48. doi:10.3233/SHTI210049
- Schmidt, C. O., Fluck, J., Golebiewski, M., Grabenhenrich, L., Hahn, H., Kirsten, T., . . . Covid, N. F. H. T. F. (2021). [Making COVID-19 research data more accessible-building a nationwide information infrastructure]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 64(9), 1084-1092. doi:10.1007/s00103-021-03386-x
- Scott, J. (2011). The Metadata Mania.
- Sinaci, A. A., & Laleci Erturkmen, G. B. (2013). A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. *J Biomed Inform*, 46(5), 784-794. doi:10.1016/j.jbi.2013.05.009
- Stausberg, J., & Harkener, S. (2019). Bridging Documentation and Metadata Standards: Experiences from a Funding Initiative for Registries. *Stud Health Technol Inform*, 264, 1046-1050. doi:10.3233/SHTI190384
- Stausberg, J., & Harkener, S. (2021). Metadata of Registries: Results from an Initiative in Health Services Research. *Stud Health Technol Inform*, 281, 18-22. doi:10.3233/SHTI210112
- Stausberg, J., Harkener, S., Burgmer, M., Engel, C., Finger, R., Heinz, C., . . . Wegner, J. (2022). Metadata Definition in Registries: What Is a Data Element? *Stud Health Technol Inform*, 294, 174-178. doi:10.3233/SHTI220432
- Stohr, M. R., Gunther, A., & Majeed, R. W. (2021). ISO 21526 Conform Metadata Editor for FAIR Unicode SKOS Thesauri. *Stud Health Technol Inform*, 278, 94-100. doi:10.3233/SHTI210056
- Stöhr, M. R., Günther, A., & Majeed, R. W. (2021). The Collaborative Metadata Repository (CoMetaR) Web App: Quantitative and Qualitative Usability Evaluation. *JMIR Med Inform*, 9(11), e30308. doi:10.2196/30308

- Tcheng, J. E., Drozda, J. P., Gabriel, D., Heath, A., Wilgus, R. W., Williams, M., . . . Windle, J. R. (2020). Achieving Data Liquidity: Lessons Learned from Analysis of 38 Clinical Registries (The Duke-Pew Data Interoperability Project. *AMIA Annu Symp Proc(1942-597X (Electronic))*, 2019: 2864–2873.
- Trifan, A., van der Lei, J., Díaz, C., & Oliveira, J. L. (2018). A Methodology for Fine-Grained Access Control in Exposing Biomedical Data. *Stud Health Technol Inform*, 247, 561-565.
- Ulrich, H., Kern, J., Tas, D., Kock-Schoppenhauer, A. K., Uckert, F., Ingenerf, J., & Lablans, M. (2019). QL(4)MDR: a GraphQL query language for ISO 11179-based metadata repositories. *BMC Med Inform Decis Mak*, 19(1), 45. doi:10.1186/s12911-019-0794-z
- Ulrich, H., Kock-Schoppenhauer, A.-K., Deppenwiese, N., Gött, R., Kern, J., Lablans, M., . . . Ingenerf, J. (2022). Understanding the Nature of Metadata: Systematic Review. *J Med Internet Res*, 24(1), e25440. doi:10.2196/25440
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi:10.1038/sdata.2016.18
- Wolstencroft, K., Owen, S., Krebs, O., Nguyen, Q., Stanford, N. J., Golebiewski, M., . . . Goble, C. (2015). SEEK: a systems biology data and model management platform. *BMC Syst Biol*, 9, 33. doi:10.1186/s12918-015-0174-y

10.2 Tabellenverzeichnis

Tabelle 1: Teilnehmende Einrichtungen Expertiseworkshop 18.11.2022.....	4
Tabelle 1: Metadatenkataloge und -repositories in Deutschland	8
Tabelle 2: Beispiele für europäische und internationale Metadatenkataloge.....	12

10.3 Abkürzungsverzeichnis und Glossar

Abkürzung / Begriff	Beschreibung
AD	Alzheimer Disease
ADOPT	implementAtion anD OPERation of the gateway for healTh into BBMRI-ERIC
AG	Arbeitsgruppe
AGATE	Softwaretool der OBiBa Suite
AMIA	American Medical Informatics Association (http://amia.org)
API	Application Programming Interface: Softwareschnittstelle zu Programmen oder Betriebssystemen
ATLAS	Softwaretool zur Durchführung wissenschaftlicher Analysen von standardisierten Gesundheitsdaten
BBMRI	[European] Biobanking and Biomolecular Resources Research Infrastructure, ein ESFRI-Projekt (http://www.bbmri.de/)
BBMRI-ERIC	Europäische Biobanken-Infrastruktur (www.bbmri-eric.eu)
BMBF	Bundesministerium für Bildung und Forschung (www.bmbf.de)
BMC	BioMed Central; Verlag (www.biomedcentral.com)
BMG	Bundesministerium für Gesundheit (www.bundesgesundheitsministerium.de)
BRIDGE	“Bridging Information and Data Generation for Evidence-based Health policy and research” (BRIDGE Health)
CCDC	Colon Cancer Data Collection
CCP	Clinical Communication Platform des DKTK
CDE	Common Data Element
CDISC	Clinical Data Interchange Standards Consortium (www.cdisc.org)
CDM	Common Data Model(s)
CEN	Comité Européen de Normalisation, Europäisches Komitee für Normung (www.cenorm.be)
CERN	Conseil Européen pour la Recherche Nucléaire (www.cern.ch)
CoMetaR	
CoV	Coronavirus
COVID	Coronavirus Disease 2019; meldepflichtige Infektionskrankheit in Folge einer Infektion mit dem Coronavirus SARS-CoV-2
CRMDR	Clinical Research Metadata Repository
CSV	Comma Separated Values
DAA	Data Access Agreement
DAC	Data Access Comitee
DAO	Data Authority Organisation
DataCite	Internationale Non-Profit-Organisation zur zitierfähigen Auszeichnung von Forschungsdaten mit Hilfe des DOI-Systems (www.datacite.org)
DataSHIELD	Open-Source-Software zur sicheren, föderativen Auswertung biomedizinischer Daten auf der Basis von R (www.datashield.org)
DC	Dublin Core (siehe DCMI)
DCAT	Data Catalog Vocabulary (Version 3) von Dataset Exchange Working Group (DXWG)
DCDE	Domainspecific Common Data Elements
DCMI	Dublin Core Metadata Initiative (http://dublincore.org)
DCRI	Duke Clinical Research Institute
DDI	Data Documentation Initiative (www.ddialliance.org)
DESAIC	Diplomate of the European Society of Anaesthesiology and Intensive Care
DFG	Deutsche Forschungsgemeinschaft (www.dfg.de)
DIC	Data Integration Center, s. DIZ
DIZ	Datenintegrationszentrum im Förderkonzept Medizininformatik des BMBF
DK	Dänemark
DKTK	Deutsches Konsortium für Translationale Krebsforschung (https://dktk.dkfz.de)
DOI	Digital Object Identifier System der IDF (www.doi.org)

Abkürzung / Begriff	Beschreibung
DOR	Directory of Registries (ERDRI)
DQAF	Data Quality Assessment Framework
DWH	Data Ware House
DZHI	Deutsches Zentrum für Herzinsuffizienz (www.dzhi.de)
DZL	Deutsches Zentrum für Lungenforschung (www.dzg-lungenforschung.de)
EC	European Commission
ECG	Electrocardiogram
eCRF	electronic Case Report Form
ECRIN	European Clinical Research Infrastructures Network, seit 2007 als ESFRI-Projekt gefördert (www.ecrin.org)
EDC	Electronic Data Capturing
EE	Estland
EFPIA	European Federation of Pharmaceutical Industries and Associations (www.efpia.org)
EG	ECG Tests; Finding Domain in CDISC-SDTM
EG	Europäische Gemeinschaft
EGA	European Genome-phenome Archive (https://ega-archive.org)
EHDEN	European Health Data & Evidence Network (https://www.ehden.eu/#)
EHDS	European Health Data Space, Regulierungsentwurf der Europäischen Kommission zu einem europäischen Gesundheitsdatenraum (https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en)
EHDS1	MyHealth@EU: grenzüberschreitender Austausch von Patientendaten
EHDS2	Zugang zu Daten für Forschung und Politikgestaltung (Sekundärnutzung von Gesundheitsdaten)
EHES	European Health Examination Survey
EHR	Electronic Health Record
EHR4CR	Electronic Health Records for Clinical Research, im Rahmen der IMI gefördertes EU-Projekt (www.ehr4cr.eu)
EK	Ethikkommission(en)
EMA	European Medicines Agency (www.ema.europa.eu)
EMIF	European Medical Information Framework
EN	Europäische Norm des CEN
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EOSC	European Open Science Cloud; Initiative der Europäischen Kommission (http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud)
ERDRI	European Rare Disease Registry Infrastructure
ERIC	European Research Infrastructure Consortium; europäisches Rechtsinstrument für Forschungsinfrastrukturen (http://ec.europa.eu/research/infrastructures/index.cfm?pg=eric)
ES	Spanien
ESFRI	European Strategy Forum on Research Infrastructures (www.esfri.eu)
ETL	Extract, Transform, Load: Kurzform für den Prozess, Daten aus mehreren, heterogenen Datenquellen selektiv zu lesen, zu transformieren und in einer einheitlichen Zielstruktur abzuspeichern
ETLHelper	ETL Unterstützungstool
EUCTR	EU Clinical Trials Register (www.clinicaltrialsregister.eu)
FA	Fachärztin bzw. Facharzt
FAIR	Findable, Accessible, Interoperable, Reusable
FHIR	Fast Healthcare Interoperability Resources; HL7-Standard (http://hl7.org/fhir)
FIAHSI	IAHSI Fellows
GBN	German Biobank Node; deutscher nationaler Hub im BBMRI ERIC (www.bbmri.de)
gematik	Gesellschaft für Telematikanwendungen der Gesundheitskarte mbH (www.gematik.de)
GitHub	Webbasierter Online-Dienst für die versionierte Bereitstellung von Softwarecode (http://github.com)
GMS	German Medical Science: Webportal der German Medical Science gGmbH (www.egms.de)

Abkürzung / Begriff	Beschreibung
GNU	Rekursives Akronym für "GNU's Not Unix": 1984 gegründetes Projekt zur Erstellung eines freien (Open Source) Unix Betriebssystems (www.gnu.org)
GPL	GNU General Public License: Erste und bis heute bedeutsamste Open Source Softwarelizenz (www.gnu.org/licenses/gpl.html)
GUI	Graphical User Interface: Bedienoberfläche eines Computer-Programms
HES	Health Examination Survey
HL7	Health Level Seven; Internationale SDO für den Bereich der Interoperabilität von IT-Systemen im Gesundheitswesen (www.hl7.org)
HMA	Heads of Medicines Agencies (HMA) is a network of the heads of the National Competent Authorities (NCA)
HTTP	Hyper Text Transfer Protocol
i2b2	Informatics for Integrating Biology and the Bedside (www.i2b2.org)
IAHSI	International Academy Health Science Informatics
IARC	International Agency for Research on Cancer der WHO (www.iarc.fr)
ICAT	ICAT open-source Metadatenstandard und Tools
ICTRP	International Clinical Trials Registry Platform der WHO (http://www.who.int/ictip)
IDF	International DOI Foundation (www.doi.org)
IEC	International Electrotechnical Commission (www.iec.ch)
IHE	Integrating the Healthcare Enterprise (www.ihe.net)
IMI	Innovative Medicines Initiative (www.imi-europe.org)
INFN	Istituto Nazionale di Fisica Nucleare
ISO	International Organization for Standardization (www.iso.org)
ISRCTN	International trial registry
JA	Joint Action
JMIR	Journal of Medical Internet Research (www.jmir.org)
JRC	EC Joint Research Centre (https://ec.europa.eu/jrc)
JSF	JavaServer Faces
JU	Joint Undertaking
KMU	Klein- und mittelständische Unternehmen
MDM	Medical Data Models (https://medical-data-models.org)
MDR	Metadata Repository
MDRMatcher	Metadata Repository Matching Tool
MIABIS	Minimum Information About Biobank data Sharing (Standard)
MICA	toolbox for masked independent component analysis - OBiBa
MII	Medizininformatik-Initiative des BMBF (www.medizininformatik-initiative.de)
MINERVA	Metadatenkatalog: Metadata for data discoverability and study replicability in observational studies
MIRACUM	Medizininformatik in Forschung und Versorgung in der Universitätsmedizin Medical Informatics in Research and Care in University Medicine (www.miracum.org)
NFDI	Nationale Forschungsdateninfrastruktur; Idee und Empfehlung aus dem Positionspapier „Leistung aus Vielfalt“ des RfII zum Forschungsdatenmanagement von 2016 und von der DFG 2019 ausgeschriebene Projektförderung
NFDI4Health	Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten; im Rahmen der NFDI geförderter Infrastrukturaufbau (www.nfdi4health.de)
NL	Niederlande
OBiBa	Open Source Software for Epidemiology
ODM	Operational Data Model (CDISC-Standard)
ODMedit	Metadatenwerkzeug
OHDSI	Observational Health Data Sciences and Informatics; Initiative zur Nutzung von Gesundheitsdaten für large scale analytics (www.ohdsi.org)
OMOP	Observational Medical Outcomes Partnership: Historische Initiative aus den USA zur Entwicklung von Methoden und Verfahren zur Nutzung klinischer Beobachtungsdaten zur

Abkürzung / Begriff	Beschreibung
	Überwachung der Sicherheit medizinischer Produkte, heute überwiegend aufgegangen in der OHDSI (http://omop.org)
ONEDATA	https://onedata.org/#/home
ONYX	Metadatenwerkzeug OBiBa
OPAL	Metadatenwerkzeug OBiBa
OpenEDC	Open source Electronic Data Capture
OpenEHR	Internationale Organisation zur Entwicklung interoperabler Electronic Healthcare Records (www.openehr.org)
openEHR	Internationale Organisation zur Entwicklung interoperabler Electronic Healthcare Records (www.openehr.org)
OSSE	Open-Source-Registersystem für Seltene Erkrankungen in der EU; vom BMG gefördertes Projekt im Rahmen des Aktionsplans für Menschen mit seltenen Erkrankungen
OVH	Cloudanbieter https://www.ovhcloud.com/de/about-us/
PaDaWaN	Patient DataWarehouse Navigator
PB	PetaByte
PBPR	Population-Based Patient Registries
PDF	Portable Document Format von Adobe (www.adobe.com)
PHP	Rekursives Akronym für PHP Hypertext Preprocessor: Programmiersprache zur Erstellung von Webseiten
PID	(pseudonymer) Patientenidentifikator
PID	Persistent Identifier(s)
PLOS	Public Library of Science; nicht-kommerzielles Open-Access-Projekt für wissenschaftliche Publikationen (www.plos.org)
PLoS	siehe PLOS
POC	Proof-of-Concept
PostgreSQL	Relationales Open Source Datenbank Management System (www.postgresql.org)
PubMed	Online Datenbank der U.S. National Library of Medicine (http://pubmed.gov)
QRPH	IHE Quality, Research and Public Health Domain
RD	Rare Disease
RDA	Research Data Alliance (http://europe.rd-alliance.org)
RDF	Resource Description Framework: Formale Sprache zur Bereitstellung von Metadaten im WWW (www.w3.org/RDF)
REST	Representational State Transfer; Standard zur einfachen Datenübertragung in Webanwendungen auf Basis des HTTP-Protokolls
RfII	Rat für Informationsinfrastrukturen (www.rfii.de)
ROCK	Metadatenwerkzeug OBiBa
SALUS	EU-Projekt https://sec-salus.eu/
SARS	Severe Acute Respiratory Syndrome; virale Infektionskrankheit
SDO	Standards Development Organization
SDTM	Study Data Tabulation Model (CDISC-Standard)
SEEK	Daten- und Modell-Management-Plattform zur Unterstützung offener und reproduzierbarer Forschungsergebnisse in der Systembiologie (www.seek4science.org)
SKOS	Simple Knowledge Organization System; W3C-Standard (www.w3.org/2004/02/skos)
SPSS	Datenanalysesoftware der Firma SPSS inc. (www.spss.com)
SQL	Structured Query Language (Standard-Sprache für Datenbank-Zugriff)
Synonym	Anlage mehrerer Datensätze zu einer Person mit jeweils unterschiedlichem PID
TEHDAS	Towards European Health Data Space
TMF	TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (www.tmf-ev.de)
UCS	Universal Character Set; ISO-Standard für die einheitliche numerische Codierung von Zeichen bzw. Textelementen aller bekannten Schriftkulturen und Zeichensysteme; entspricht dem Unicode-Zeichensatz
UK	United Kingdom

Abkürzung / Begriff	Beschreibung
UKW	Universitätsklinikum Würzburg
UMG	Universitätsmedizin Göttingen (www.med.uni-goettingen.de)
Unicode	Internationaler Standard für die einheitliche numerische Codierung von Zeichen bzw. Textelementen aller bekannten Schriftkulturen und Zeichensysteme; unter der Bezeichnung UCS auch als ISO-Standard veröffentlicht.
W3C	World Wide Web Consortium (www.w3.org)
WHO	World Health Organization (www.who.org)
WP	Working Party der Europäischen Arbeitsgruppe zum Datenschutz gemäß Artikel 29 der Richtlinie 95/46/EG
WWW	World Wide Web
XDC	Format XFA Device Control (XDC)
XML	extensible Markup Language
ZB MED	Deutsche Zentralbibliothek für Medizin – Informationszentrum Lebenswissenschaften (www.zbmed.de)
Zenodo	Open Science Plattform für Veröffentlichungen, Forschungsergebnisse und Forschungsdaten; betrieben vom CERN und unterstützt von der EU (https://zenodo.org)
ZMR	Zentralstelle für Medizinische Register