

Bundesgesundheitsbl 2021 · 64:1084–1092
<https://doi.org/10.1007/s00103-021-03386-x>
 Eingegangen: 26. März 2021
 Angenommen: 28. Juni 2021
 Online publiziert: 23. Juli 2021
 © Der/die Autor(en) 2021



Carsten Oliver Schmidt¹ · Juliane Fluck^{2,3,4} · Martin Golebiewski⁵ ·
 Linus Grabenhenrich⁶ · Horst Hahn^{7,8} · Toralf Kirsten^{9,10} · Sebastian Klammt¹¹ ·
 Matthias Löbe¹² · Ulrich Sax¹³ · Sylvia Thun¹⁴ · Iris Pigeot^{15,16} · NFDI4Health Task
 Force Covid-19

¹ Institut für Community Medicine, Universitätsmedizin Greifswald, Greifswald, Deutschland

² ZB MED – Informationszentrum Lebenswissenschaften, Bonn, Deutschland

³ Institut für Geodäsie und Geoinformation, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Deutschland

⁴ Abteilung Bioinformatik, Fraunhofer Institut SCAI, Sankt Augustin, Deutschland

⁵ Heidelberger Institut für Theoretische Studien (HITS), Heidelberg, Deutschland

⁶ Robert Koch-Institut, Berlin, Deutschland

⁷ Institut für Digitale Medizin, Fraunhofer MEVIS, Bremen, Deutschland

⁸ Jacobs University, Bremen, Deutschland

⁹ Fakultät Angewandte Computer- und Biowissenschaften, Hochschule Mittweida, Mittweida, Deutschland

¹⁰ Institut für Medical Data Science, Universitätsmedizin Leipzig, Leipzig, Deutschland

¹¹ Netzwerk der Koordinierungszentren für Klinische Studien – KKS-Netzwerk e. V., Berlin, Deutschland

¹² Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig, Leipzig, Deutschland

¹³ Institut für Medizinische Informatik, Universitätsmedizin Göttingen, Göttingen, Deutschland

¹⁴ Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Berlin, Deutschland

¹⁵ Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Deutschland

¹⁶ Fachbereich Mathematik und Informatik, Universität Bremen, Bremen, Deutschland

COVID-19-Forschungsdaten leichter zugänglich machen – Aufbau einer bundesweiten Informationsinfrastruktur

Einleitung

COVID-19 stellt Individuen und Gesellschaften weltweit vor eine der größten Herausforderungen der letzten Jahrzehnte. Public-Health-Forschung, epidemiologische und klinische Studien sind daher unabdingbar, um die Ausbreitung des für die Pandemie verantwortlichen SARS-CoV-2 und dessen Varianten nachzuvollziehen, die Folgen für die Gesundheit und das soziale Leben besser zu verstehen sowie wirksame Therapie- und Impfmethoden zu identifizieren. Dadurch entsteht für Politik, Wirtschaft, Gesundheitsversorgung und Gesellschaft eine empirische Grundlage zur Eindämmung und zum Umgang mit der Pandemie. Diese

bedarf jedoch einer fortlaufenden Aktualisierung.

In sehr kurzer Zeit entstanden zahlreiche Projekte, Studien und Netzwerke zur Erforschung von SARS-CoV-2 und COVID-19. Aus der Perspektive von Forschenden sind hiermit erhebliche Herausforderungen verbunden. Es fällt zunehmend schwer, einen Überblick zu behalten. Dieser Überblick ist jedoch unabdingbar, um Forschungsaktivitäten besser zu koordinieren, ungeplante Doppelforschung zu vermeiden und Studien harmonisiert zu implementieren. Derzeit wird die Zusammenführung von Wissensquellen durch die unzureichende Berücksichtigung von Standards und einen Mangel an harmonisierten Methoden auf

allen Ebenen des Forschungsprozesses erschwert.

Aufgrund der bestehenden Pflicht zur Registrierung klinischer Studien in Registern sind deren Metadaten gut strukturiert verfügbar, z. B. in der *International Clinical Trials Registry Platform* (ICTRP) der Weltgesundheitsorganisation (WHO; [1]) oder dem Deutschen Register Klinischer Studien (DRKS; [2]). Dagegen ist die Situation für epidemiologische und Public-Health-Studien wesentlich unübersichtlicher. Zwar gibt es national und international mehrere Übersichten im Internet, z. B. zu seroepidemiologischen Studien am Robert Koch-Institut (RKI; [3]), ein COVID-19-Forschungsregister der *American Society for Microbiology* [4], die COVID-

19-Forschungsübersicht der Medizin-informatik-Initiative [5] oder des Rats für Sozial- und Wirtschaftsdaten [6], aber diese Übersichten sind in Umfang, Aktualität und Informationstiefe uneinheitlich.

Noch schwieriger wird es, wenn studienübergreifend ein detaillierter Einblick in die Protokolle, Erhebungsinstrumente, Itembanken und weitere Studiendokumente gewonnen werden soll, obwohl Technologien zur übersichtlichen Aufbereitung und Darstellung solcher Informationen grundsätzlich verfügbar sind [7–9]. Dies wäre zum Beispiel wichtig, um eigene Erhebungen mit bestehenden vergleichbar zu planen. Nur vereinzelt bieten Projekte Zugang zu relevanten Informationen. So wurde im Netzwerk Universitätsmedizin (NUM; [10]), das vor allem krankenhausbezogene Forschung koordiniert, mit dem German Corona Consensus Dataset (GECCO) ein positives Beispiel für harmonisierte Datenerhebungen auf Basis von internationalen medizinischen IT-Standards anhand eines abgestimmten Kerndatensatzes geschaffen [11]. Auch das GESIS – Leibniz-Institut für Sozialwissenschaften stellt eigene Erhebungsinstrumente und weitere Studiendokumente auf ihren Webseiten bereit [12].

Ein weiteres Problem betrifft den Zugriff auf neueste Forschungsergebnisse und Daten, die inzwischen häufig in Preprints publiziert werden, bevor sie mit oft mehrmonatigen Verzögerungen in etablierten Literaturdatenbanken wie PubMed oder Web of Science auffindbar sind. Ein zentraler Zugriff auf diese verteilt vorliegenden Preprint-Archive ist daher wichtig, um aktuelle Ergebnisse besser zu finden und die Idee des breiten Communityreviews von Preprints im Sinne einer Qualitätssicherung praktisch zu unterstützen.

Darüber hinaus wurden im Laufe der Pandemie weitere Hürden offensichtlich, die eine effiziente Forschung erschweren: Obwohl teilweise dieselben Personen in verschiedene Studien eingeschlossen wurden und weitere Gesundheitsdaten dieser Personen z. B. bei den Krankenkassen gespeichert sind, fehlen ausreichende Optionen, diese Daten auf in-

Bundesgesundheitsbl 2021 · 64:1084–1092 <https://doi.org/10.1007/s00103-021-03386-x>
© Der/die Autor(en) 2021

C. O. Schmidt · J. Fluck · M. Golebiewski · L. Grabenhenrich · H. Hahn · T. Kirsten · S. Klammt · M. Löbe · U. Sax · S. Thun · I. Pigeot · NFDI4Health Task Force Covid-19

COVID-19-Forschungsdaten leichter zugänglich machen – Aufbau einer bundesweiten Informationsinfrastruktur

Zusammenfassung

Public-Health-Forschung, epidemiologische und klinische Studien sind erforderlich, um die COVID-19-Pandemie besser zu verstehen und geeignete Maßnahmen zu ergreifen. Daher wurden auch in Deutschland zahlreiche Forschungsprojekte initiiert. Zum heutigen Zeitpunkt ist es ob der Fülle an Informationen jedoch kaum noch möglich, einen Überblick über die vielfältigen Forschungsaktivitäten und deren Ergebnisse zu erhalten. Im Rahmen der Initiative „Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten“ (NFDI4Health) schafft die „Task Force COVID-19“ einen leichteren Zugang zu SARS-CoV-2- und COVID-19-bezogenen klinischen, epidemiologischen und Public-Health-Forschungsdaten. Dabei werden die sogenannten FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) berücksichtigt, die eine schnellere Kommunikation von Ergebnissen befördern sollen. Zu den wesentlichen Arbeitsinhalten der Taskforce gehören die Erstellung eines Studienportals

mit Metadaten, Erhebungsinstrumenten, Studiendokumenten, Studienergebnissen und Veröffentlichungen sowie einer Suchmaschine für Preprint-Publikationen. Weitere Inhalte sind ein Konzept zur Verknüpfung von Forschungs- und Routinedaten, Services zum verbesserten Umgang mit Bilddaten und die Anwendung standardisierter Analyseroutinen für harmonisierte Qualitätsbewertungen. Die im Aufbau befindliche Infrastruktur erleichtert die Auffindbarkeit von und den Umgang mit deutscher COVID-19-Forschung. Die im Rahmen der NFDI4Health Task Force COVID-19 begonnenen Entwicklungen sind für weitere Forschungsthemen nachnutzbar, da die adressierten Herausforderungen generisch für die Auffindbarkeit von und den Umgang mit Forschungsdaten sind.

Schlüsselwörter

COVID-19-Studienportal · FAIR-Prinzipien · Epidemiologie · Public Health · Klinische Studien

Making COVID-19 research data more accessible—building a nationwide information infrastructure

Abstract

Public health research and epidemiological and clinical studies are necessary to understand the COVID-19 pandemic and to take appropriate action. Therefore, since early 2020, numerous research projects have also been initiated in Germany. However, due to the large amount of information, it is currently difficult to get an overview of the diverse research activities and their results. Based on the “Federated research data infrastructure for personal health data” (NFDI4Health) initiative, the “COVID-19 task force” is able to create easier access to SARS-CoV-2- and COVID-19-related clinical, epidemiological, and public health research data. Therefore, the so-called FAIR data principles (findable, accessible, interoperable, reusable) are taken into account and should allow an expedited communication of results. The most essential work of the task force includes the generation of a study portal

with metadata, selected instruments, other study documents, and study results as well as a search engine for preprint publications. Additional contents include a concept for the linkage between research and routine data, a service for an enhanced practice of image data, and the application of a standardized analysis routine for harmonized quality assessment. This infrastructure, currently being established, will facilitate the findability and handling of German COVID-19 research. The developments initiated in the context of the NFDI4Health COVID-19 task force are reusable for further research topics, as the challenges addressed are generic for the findability of and the handling with research data.

Keywords

COVID-19 study portal · FAIR principles · Epidemiology · Public health · Clinical trials

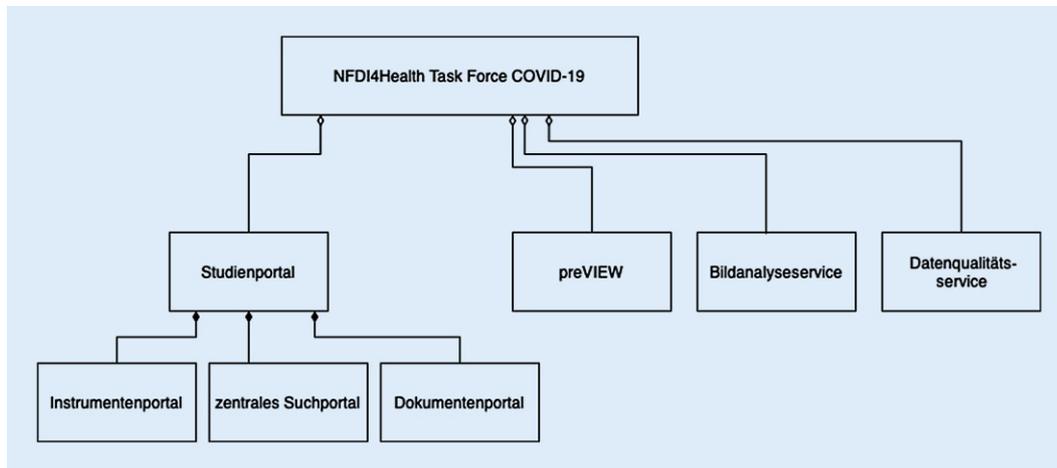


Abb. 1 ◀ Übersicht zu Services und Aktivitäten der NFDI4Health Task Force COVID-19

dividueller Ebene zu verknüpfen. Dies beschränkt die Möglichkeiten, ein ausreichend umfassendes Bild des Krankheitsgeschehens zu erhalten, um verlässlichere Aussagen zur Verlaufsprognose oder zu Impffolgen schnell zu erhalten. Dies ist der Fall, obwohl es außerhalb von reinen Forschungsprojekten positive Beispiele gibt, die durch eine entsprechende Gesetzgebung abgedeckt sind, wie etwa die Zusammenführung von Melde- und Sequenzierungsdaten am RKI, um verbesserte Aussagen zu Virusvarianten zu erhalten.

Zusammenfassend erfüllen die deutschen klinischen Studien zu COVID-19 sowie entsprechende Datenbestände in Epidemiologie und Public Health trotz positiver Beispiele die Ansprüche der sogenannten FAIR-Prinzipien [13] noch nicht im vollen Umfang. Dabei steht FAIR für die Auffindbarkeit (Findable), Zugänglichkeit (Accessible), Interoperabilität (Interoperable) und Wiederverwendbarkeit (Reusable) von Forschungsdaten LINK zu [14]. Um diesem Defizit zu begegnen, wurde als Teil des deutschen interdisziplinären Netzwerkprojekts „Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten“ (NFDI4Health; [15]) die Task Force COVID-19 etabliert [16]. Ihr Ziel ist es, eine bundesweite Informationsinfrastruktur zu entwickeln, um medizinische, epidemiologische und Public-Health-Forschung nach FAIR-Kriterien leichter zugänglich zu machen und Forschungsergebnisse besser zu kommunizieren. Dabei betrachtet die

NFDI4Health Task Force COVID-19 neben Forschung in Bezug auf PatientInnen mit COVID-19 auch die Public-Health-Folgen des Pandemieausbruchs auf die Allgemeinbevölkerung.

Wesentliche Arbeitsinhalte der NFDI4Health Task Force COVID-19 umfassen:

1. die Erstellung eines Studienportals zum Auffinden deutscher COVID-19-Forschungsinitiativen mit strukturierten Gesundheitsdaten aus epidemiologischen und klinischen Studien inkl. Impfstudien, administrativen Datenbanken, der Primärversorgung und der Gesundheitsberichterstattung, das auch die semantisch aufbereitete vergleichende Darstellung von Items aus Erhebungsinstrumenten umfasst;
2. die auf Text-Mining (Extraktion aus großen Textmengen) basierende Aufbereitung und Darstellung von SARS-CoV-2- und COVID-19-bezogener Forschung unter Nutzung der Inhalte verschiedener Preprint-Server in einer semantischen Suchmaschine;
3. ein Konzept zur Verknüpfung von Forschungs- und Routinedaten;
4. Services zum verbesserten Umgang mit Bilddaten;
5. die Anwendung standardisierter Analyseroutinen für harmonisierte Qualitätsbewertungen.

Die primäre Zielgruppe der Infrastruktur sind Forschende, die Studien zu SARS-CoV-2 oder COVID-19 durchführen oder planen. Die intendierte Anwendung

der ersten beiden Arbeitsinhalte betrifft insbesondere das leichtere Auffinden deutscher COVID-19-Ressourcen zur besseren Planung eigener Studien sowie zur Bewertung von Studienergebnissen. Mit den Arbeitsinhalten 3–5 werden Hilfsmittel an die Hand gegeben, um die Qualität spezifischer Aspekte von COVID-19-bezogener Forschung wie das Verknüpfen von Daten aus verschiedenen Quellen (Record-Linkage) oder Bildanalyse zu unterstützen. Eine Übersicht zu den verschiedenen Arbeitsinhalten findet sich in [Abb. 1](#). Der Schwerpunkt dieser Publikation liegt auf der Darstellung des Studienportals und der Suchmaschine für Preprint-Publikationen (Arbeitsinhalte 1–2).

COVID-19-Studienportal

Um ein Studienportal entwickeln zu können (s. Arbeitsinhalt 1), das einerseits mit Registern für klinische Studien kompatibel ist, insbesondere dem ICTRP [1] und dem DRKS [2], und andererseits epidemiologische und Public-Health-Studien angemessen abbildet, musste ein geeignetes Datenmodell formuliert werden. Mit diesem wird eine strukturierte Erfassung von studienbezogenen Metadaten in einer Datenbank ermöglicht. Zu diesem Zweck greift es auf Attribute und Wertelisten aus einer Reihe von Vorarbeiten zurück, z. B. den *Minimum Information About Biobank Data Sharing* (MIABIS; [17, 18]), das Maelstrom-Datenmodell [19] sowie dem DataCite-Metadatenchema [20]. Ein Mapping

Studies Instruments

Refine your search. Currently 18 studies are matched

seroprevalence

Quick filters

- Available Documents
- Available Instruments
- Primary Design
 - Non-interventional 17
 - Interventional 1
- Study Status
 - Recruiting 15
 - Completed 1
 - Not yet recruiting 1
 - Suspended 1
- Study Type

29 columns hidden Sort fields

View Details	Acronyms	Title	Source Identifiers	Webpage
Open	CovidimmunEye	Immune Status: Antibodies Against Severe Acute Respiratory Syndrome Coronavirus 2 in a...	NCT04448338	
Open	SchoolCovIDD19	Prospektive Erfassung der SARS-CoV-2 Seropositivität bei Schulkindern nach Ende der un...	DRKS00022455	
Open	Co-HCW	Erhebung der Seroprävalenz und Infektionsstatus von SARS-CoV-2 bei Mitarbeitern des U...	DRKS00022432	
Open	SLKovid	SLKovid: Seroprävalenz von SARS-CoV-2 bei Mitarbeiterinnen und Mitarbeitern eines Klini...	DRKS00022226	
Open	AK-TIR-COVID-19	Seroprävalenz von Antikörpern gegen COVID-19 bei Klinikpersonal im Vergleich zu einem ...	DRKS00021813	
Open	KiTaCovIDD19	SARS-CoV-2 Surveillance im Vorschulalter (COVID-19)	DRKS00022729	
Open	PrEP-Co	SARS-CoV-2 Serologie bei HIV PrEP Gebräuchern während der COVID-19 Pandemie: Die ...	DRKS00022403	
Open	Wü-KiTa-CoV	Würzburger Kinderbetreuungsstudie in der COVID-19 Pandemie: Akzeptanz unterschiedli...	DRKS00023721	Visit
Open	KoCo19-CED-1	Versorgungssituation, Gesundheitsprobleme und psychosoziale Belastungen während der ...	DRKS00022155	
Open	AmbCovIDD19	Ambcovidd19 Erfassung von Krankheitsverläufen, Infektionsketten und Seroprävalenz bei Covid- 19 Infe...	DRKS00022549	
Open	SeMaCo	Serologische Untersuchungen bei Blutspendern des Großraums Magdeburg auf Antikörpe...	DRKS00023263	
Open	FamilyCovIDD19	Erfassung der SARS-CoV-2-Seroprävalenz in Haushaltsgemeinschaften und Transmission ...	DRKS00022564	
Open	Co-NAN	CoVID-19-Ausbruch in Neustadt am Rennsteig: Erhebung von Seroprävalenz, Infektionsst...	DRKS00022416	
Open	COKITS	Screening for SARS-CoV-2-Infections in Teachers, Childcare Educators, Pupils and Presch...	NCT04619407	
Open	CovSeroUKL	Seroprävalenz von SARS CoV-2/Covid-19 in Blutspendern und Patienten am Universitätskl...	DRKS00023364	
Open		Prevalence and flight-associated risk factors of COVID-19 infection experienced by refuge...	ISRCTN17401860	Visit
Open	COVID-19 BaWü	Prävalenz von COVID-19 bei Kindern in Baden-Württemberg	DRKS00021521	
Open	SARS-CoV-2 seroprevalenc...	Serielle Querschnittsstudie zur Erhebung der Seroprävalenz von SARS-CoV-2 Infektionen ...	DRKS00021709	

Abb. 2 ▲ Suchportal mit einem Ergebnisausschnitt, der Seroprävalenzstudien umfasst

gegen die Standards HL7 FHIR [21] und das Clinical Data Interchange Standards Consortium – Operational Data Model (CDISC ODM) wurden umgesetzt, um eine angemessene Interoperabilität zu gewährleisten. HL7 FHIR sowie CDISC ODM sind Standards zum Austausch strukturierter klinischer Daten zwischen Geräten und Organisationen. Da keiner der Standards eine vollständige Abdeckung der Problemdomäne bot, wurde das Datenmodell zunächst nicht konsistent innerhalb eines dieser Standards formuliert. Details des Datenmodells sind in einer eigenen Publikation verfügbar [22]. Durch die Kompatibilität mit dem Datenformat von DataCite können Studiendokumente und Instrumente einzeln publiziert und ein DOI (Document Object Identifier) vergeben werden [23]. So wird auch „graue Literatur“ referenzier- und zitierbar. Darüber hinaus kann eine forschungsfreundliche Lizenz für deren Nachnutzung im Creative Commons Framework [24] vergeben werden.

Um die relationale Datenbank zu befüllen, die das Datenmodell implementiert, wurde im ersten Schritt ein

automatisches Verfahren entwickelt, um Einträge zu klinischen Studien aus den Registern DRKS und WHO ICTRP regelmäßig auszulesen. Im zweiten Schritt wurden epidemiologische und Public-Health-Studien anhand geeigneter Suchmaschinen wie Google und PubMed identifiziert, die jeweiligen Studienleitungen individuell kontaktiert und um ihre Teilnahme an dem Studienportal gebeten. Studienmerkmale wurden in diesem Prozess manuell erfasst.

Das COVID-19-Studienportal [25] mit seinen 3 Teilkomponenten zentrales Suchportal, Dokumentenportal und Instrumentenportal bedient verschiedene Anwendungsfälle, die nachfolgend beschrieben sind.

Zentrales Suchportal

Das zentrale Suchportal (Abb. 2) ist eine als browserbasierte Single Page Application entwickelte Anwendung, welche NutzerInnen einen einfachen Zugang zu Studieninformationen bietet. Zusätzlich werden NutzerInnen über diesen Dienst zu den anderen Webportalen der NFDI4Health Task Force COVID-

19 weitergeleitet. Stand Mai 2021 waren 691 Studien eingeschlossen.

Zur Integration der Studieninformationen aus verschiedenen Datenquellen wurde das oben erwähnte Datenmodell [23] verwendet. Dieses erlaubt neben der Beschreibung von Studien auch die Darstellung von anderen Ressourcen mit entsprechenden Metadaten. Dazu zählen Erhebungsinstrumente (z.B. Frage- und Erfassungsbögen) sowie andere Studiendokumente (z.B. Data Dictionaries, Studienprotokolle, Patienteninformationen und Einwilligungserklärungen). Beziehungen zwischen den Ressourcen (Studien, Erhebungsinstrumenten, Dokumenten, Forschenden usw.) können hierarchisch abgebildet werden.

Dokumentenportal

Im Rahmen der europäischen Forschungsdateninitiative FAIRDOM wurde die Plattform „SEEK“ auf Basis der Software Ruby on Rails entwickelt [9, 26]. SEEK wurde im Rahmen der NFDI4Health Task Force COVID-19 an die Anforderungen des Studienportals angepasst. Die Plattform ermöglicht

The screenshot shows the SEEK platform interface. On the left, a tree view lists various studies, with 'CORONA-MONITORING lokal' selected. The main area displays the study's details, including its description, SEEK ID, investigation, projects, and person responsible. Below this, an 'Extended Metadata' section provides specific information about the study's design, eligibility, and target sample size.

Selected: CORONA-MONITORING lokal (Study)
 Description: Die Studie CORONA-MONITORING lokal ist eine bevölkerungsbezogene seroepidemiologische Beobachtungsstudie, bei der an...
 SEEK ID: <https://seek.studyhub.nfd4health.de/studies/1118>

CORONA-MONITORING lokal
 Die Studie CORONA-MONITORING lokal ist eine bevölkerungsbezogene seroepidemiologische Beobachtungsstudie, bei der an vier besonders von der COVID-19-Epidemie betroffenen Orten in Deutschland wiederholte Querschnittsuntersuchungen mit jeweils unabhängigen Stichproben durchgeführt werden sollen.

SEEK ID: <https://seek.studyhub.nfd4health.de/studies/1118>
 Investigation: Studies in the German nfd4health Study Hub Covid-19
 Projects: COVID-19 related studies and tools in Germany
 Person responsible: Not available

Extended Metadata (NFD4Health Study metadata)
 Parent id: 0
 Resource id: 10136
 Resource type: Study
 Resource web page: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/cmi-studie/Factsheet_Kupferzell.html
 Resource web studyhub: <http://covid19.studyhub.nfd4health.de/resource/10136>
 Resource web mica: <http://studycatalogbrowser.ship.med.uni-greifswald.de/study/corona-monitoring-lokal>
 Resource use rights: DE
 Resource source: manually collected
 Acronym: CoMoLo
 Study primary design: Non-interventional
 Study type: Cross-sectional
 Study analysis unit: Individual
 Study status: Completed
 Study eligibility: Individuals will be included if they a) are 18 years or older, b) are registered residents of one of the four study municipalities, c) can provide written consent to participate in the study, or, where necessary, written consent is provided by a legal representative, d) are able to take part in the interviews (where necessary with the help of relatives) and the examinations at the study centres or during home visits.
 Study country: Germany
 Study region: Kupferzell (Baden-Württemberg), Bad Feilnbach (Bayern), Straubing (Bayern), Berlin-Mitte (Berlin)
 Study population: General population
 Study target sample size: 8000

Related items
 People (1) Projects (1) Investigations (1) Resources (2)

[CORONA-MONITORING lokal \(Detailed Questionnaire\)](#)

a Der kurze schriftliche Fragebogen im Prüfungszentrum und eine optionale webbasierte oder telefonische Nachuntersuchung sammeln weitere Informationen zu versch. Symptomen und Krankengeschichte.

b Study target sample size: 8000

Abb. 3 ▲ Ausschnitt aus den erfassten Studien mit einem Beispiel für die Darstellung einer Studie in der Plattform SEEK mit ihren Ressourcen und Dokumenten (a) sowie der Beschreibung ihrer Merkmale (b)

dadurch die Speicherung und Strukturierung vielfältiger Ressourcen und Dokumente von Studien, wie z. B. Erhebungsinstrumente, Standard Operating Procedures (SOPs), Dokumentvorlagen und Studienmanuale, und erlaubt damit einen gebündelten Zugriff auf diese, inklusive Versionsverwaltung, und Möglichkeiten zur Datenpublikation, um Inhalte zitierbar zu machen (Abb. 3). Der Zugang zu den gesammelten Metadaten und Ressourcen der Studien, sofern entsprechende Nutzungsrechte bestehen, ist über eine mit dem zentralen Suchportal verlinkte Benutzeroberfläche (Web-Frontend) sowie über die Nutzung einer programmatischen Schnittstelle (API) möglich, welche zur Suche der Studien auch an das Studienportal angeschlossen ist.

Instrumentenportal

Das Instrumentenportal erleichtert den Zugang zu Inhalten von Erhebungsinstrumenten, z. B. von Survey-Fragebögen und deren Items, d. h. Variablen und zugehörige Wertelisten. Um dies zu ermöglichen, wurden semantische Suchoptionen in den frei verfügbaren Softwareanwendungen OPAL und MICA genutzt [7], die in Java, JavaScript und PHP pro-

grammiert sind. Zu diesem Zweck wurden ausgewählte Erhebungsinstrumente für die OPAL-Datenbank aufbereitet und eine semantische Annotation unter Zuhilfenahme der Maelstrom-Taxonomie durchgeführt [19]. Die Taxonomie umfasst 18 Domänen (z. B. soziodemografische und ökonomische Charakteristika, Erkrankungen, nichtpharmakologische Interventionen), die wiederum in 135 Subdomänen (z. B. ICD-Bereiche in der Domäne Erkrankungen) unterteilt sind. Dies ermöglicht eine Suche und Darstellung von Items nach inhaltlichen Kriterien (Abb. 4). Anfang Mai 2021 waren 23 Instrumente mit 3506 Items abgebildet und semantisch annotiert.

Der öffentliche Zugang besteht über den Link [27]. Abgebildet sind sowohl die Fragen und Items als auch deren Antwortkategorien. Eine grafische Übersicht gibt für jedes Instrument Aufschluss über die Anzahl der Items sortiert nach inhaltlichem Bereich, z. B. Lebensstile, Diagnosen und Symptome.

Semantische Suchmaschine für Preprints

Um einen zentralen Zugriff auf Preprints zu COVID-19 zu schaffen, wurden in einem ersten Schritt Metadaten von

den Preprint-Servern medRxiv, bioRxiv, ChemRxiv, ResearchSquare, arXiv und Preprints.org abgefragt und in ein gemeinsames Datenschema konvertiert. Im nächsten Schritt wurden diese Metadaten, z. B. Titel und Abstracts, automatisch mittels Text-Mining [28, 29] unter Zuhilfenahme geeigneter Terminologie indiziert, um die Forschenden durch die semantische Suchfunktionalität bei der Extraktion relevanter COVID-19-Informationen zu unterstützen, beispielsweise basierend auf Konzepten, Autoren, Publikationsdatum oder Quellen. Begleitend wurde eine Terminologie erstellt, um virale SARS-CoV-2-Proteine mithilfe eines wörterbuchbasierten Algorithmus zu erkennen [30]. NutzerInnen stehen die semantischen Suchfunktionalitäten über ein webbasiertes Nutzerinterface und eine Programmierschnittstelle zur Verfügung.

Die semantische Suchmaschine für COVID-19-Preprints, kurz genannt „COVID-19 preVIEW“, umfasst Stand Mai 2021 mehr als 27.000 Preprints aus 6 verschiedenen Servern und ist öffentlich zugänglich unter dem Link [30]. Die Weboberfläche zeigt einen Überblick über die neuesten Publikationen mit Metadaten, wie beispielsweise Titel, DOI und Abstract (Abb. 5). Des Wei-

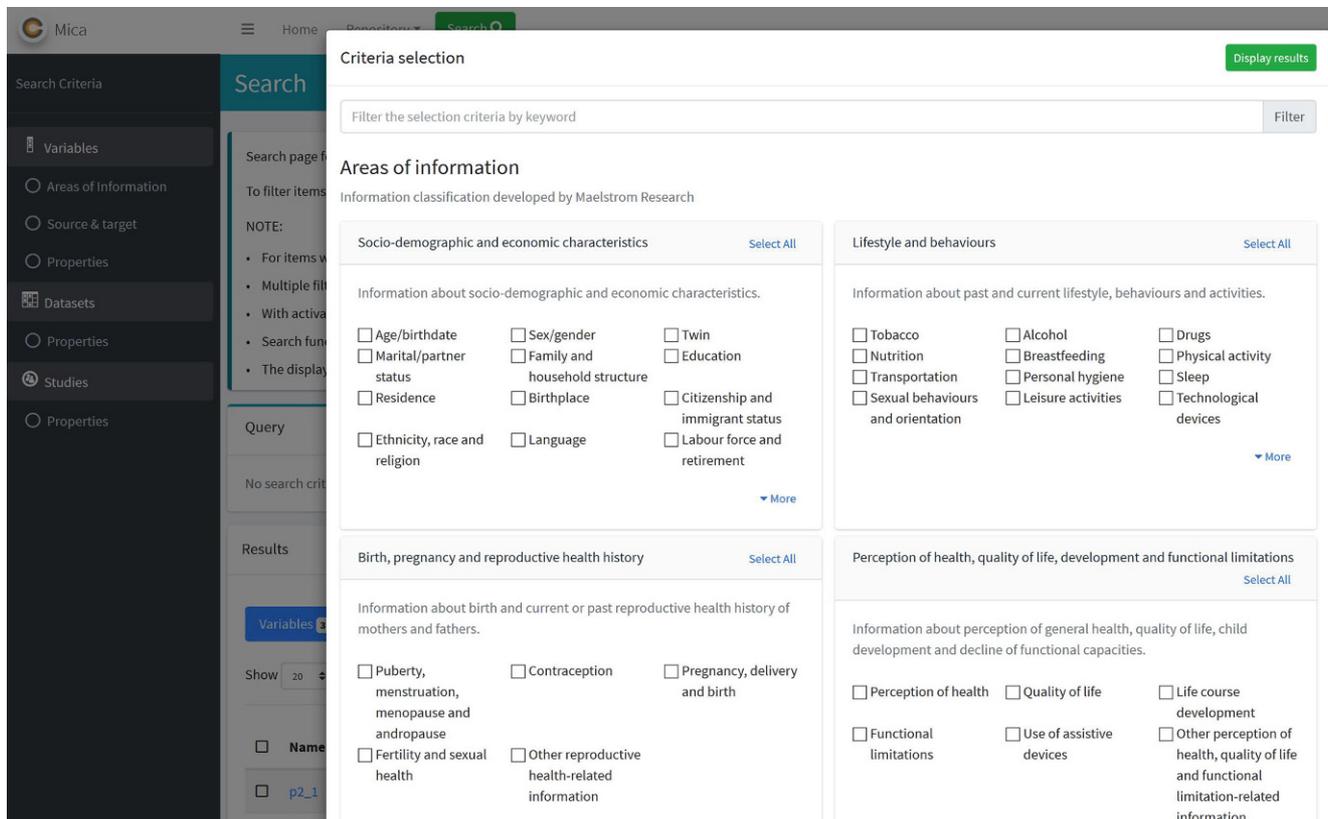


Abb. 4 ▲ Beispiel der semantischen Suche für das Instrumentenportal

teren werden die häufigsten Konzepte für jede Entitätsklasse angezeigt und können ebenfalls zur Suche hinzugefügt werden. Eine Tabelle zeigt die Häufigkeiten der Terme. Weitere Funktionalitäten erleichtern die Informationsextraktion, wie z. B. der Export des Subkorpus im Endnote- oder BibTex-Format.

Weitere Arbeitsinhalte der NFDI4Health Task Force COVID-19

Mit dem bisherigen Überblick sind wesentliche, aber nicht alle Arbeitsinhalte der NFDI4Health Task Force COVID-19 dargestellt. So wird als wichtiger Bestandteil der FAIRness von Forschungsdaten die Notwendigkeit einer stärker harmonisierten Herangehensweise an die Datenqualität berücksichtigt. Auf Basis eines bestehenden Datenqualitätskonzeptes [31] und darauf beruhender generischer Analyseroutinen in den Programmiersprachen R [32, 33] und Stata sind Anwendungsbeispiele auch auf COVID-19-Forschungsdaten bezogen, um har-

monisierte Datenqualitätsanalysen zu illustrieren. Ein Zugang besteht über die Website der Universitätsmedizin Greifswald [34].

Darüber hinaus wurden in der NFDI4Health Task Force COVID-19 Tools zur Beurteilung der Qualität bei der Bildgebung weiterentwickelt. Dies wurde am Anwendungsbeispiel der Lunge umgesetzt (Abb. 6). Dafür wird die bereits bestehende Plattform Grand Challenge [35] verwendet. Grand Challenge bietet eine Umgebung, um Bilddaten hochzuladen, zu verwalten und Methoden auf die Daten anzuwenden. Bis zu 600 Schnittbilder einer Computertomographie führt die vom Fraunhofer MEVIS, Institut für Digitale Medizin, entwickelte Software zu dreidimensionalen Darstellungen zusammen. Algorithmen für eine automatische Bildanalyse benötigen eine bestimmte Datenqualität, sodass es wichtig ist, auch die Beurteilung der Qualität der zu untersuchenden Daten zu (teil-)automatisieren. Die ausgewählten Daten können zunächst in verschiedenen Ansichten interaktiv betrachtet

werden. Relevante Strukturen, wie von COVID-19 betroffene Regionen in der Lunge, werden farblich hervorgehoben. Die Wahrscheinlichkeit einer COVID-19-Erkrankung sowie der Schweregrad des Lungenbefalls werden automatisch berechnet und angezeigt. Die Ergebnisse können als Bericht exportiert und runtergeladen werden. Dabei werden für die Datenschnittstellen standardisierte DICOM-Formate (*Digital Imaging and Communications in Medicine*; [36]) eingesetzt.

Aufbauend auf einschlägigen Publikationen [37–40] wird zudem ein Konzept zur Verknüpfung verschiedener Datenquellen wie primären Forschungsdaten, Krankenhausdaten, Abrechnungsdaten der ambulanten Versorgung, Sequenzierdaten und Bilddaten entwickelt. Dieses Konzept wird durch ein Datenschutzkonzept und eine Vorlage zur Einholung einer informierten Einwilligung ergänzt. Als Herausforderung erweist sich in diesem Teilprojekt der Umgang mit rechtlichen und logistischen Voraussetzungen. Zur Verknüpfung von

Corpus overview for the search: seroprevalence AND Germany AND cohort

seroprevalence AND Germany AND cohort

advanced search

displaying 1 - 5 records in total 10 records per page

Expand/collapse all abstracts

Seroprevalence of SARS CoV-2 antibodies in healthcare workers and administration employees: a prospective surveillance study at a 1,400-bed university hospital in Germany

Authors: Christina Bahrs; Aurelia Kimmig; Sebastian Weis; Juliane Ankert; Stefan Hagel; Jens Maschmann; Andreas Stallmach; Andrea Steiner; Michael Bauer; Wilhelm Behringer; Michael Baier; Miriam Kesselmeier; Cora Richert; Florian Zepf; Martin Walter; Andre Scherag; Michael Kiehntopf; Bettina Loeffler; Mathias W Pletz

doi:10.1101/2020.09.29.20203737 Date: 2020-09-29 Source: medRxiv

Abstract

Testing the backbone of the healthcare system: a prospective serological-epidemiological cohort study of healthcare workers in rural Germany.

Authors: Stephanie Hoffmann; Jacob Spallek; Heinz-Detlef Gremmels; Juliane Schiebel; Frank Hulert

doi:10.21203/rs.3.rs-84703v1 Date: 2020-09-28 Source: ResearchSquare

Abstract

Seroprevalence and correlates of SARS-CoV-2 neutralizing antibodies: Results from a population-based study in Bonn, Germany

Authors: N. Ahmad Aziz; Victor M Corman; Antje K.C. Echterhoff; Anja Richter; Antonio Schmandke; Marie Luisa Schmidt; Thomas H. Schmidt; Folgerdiena M De Vries; Christian Drosten; Monique M.B. Breteler

doi:10.1101/2020.08.24.20181206 Date: 2020-08-29 Source: medRxiv

Sources

- medRxiv
- ResearchSquare
- arXiv
- Preprints.org
- ChemRxiv
- bioRxiv

Annotations

- All
- None
- MeSH Disease
- HGNC Genes
- SARS-CoV-2 Proteins

Show

This service is developed in the project **nfdi4health task force covid-19** which is a part of nfdi4health.

Abb. 5 ▲ Preprint-Suchportal mit einem Ergebnisausschnitt, der Seroprävalenzstudien umfasst



Abb. 6 ◀ 3-D-Visualisierung der Lungenanatomie und von COVID-19 betroffenen Regionen. (Fraunhofer MEVIS, Foto: mit freundlicher Genehmigung © Jens Lehmkuhler/U Bremen Research Alliance)

Routine- und Forschungsdaten sind datenschutzkonforme Einwilligungen und Pseudonymisierungsansätze notwendig. In den meisten Datenquellen werden jedoch nur zweckgebundene Auswertungen durch Einverständnisse und Einwilligungen der ProbandInnen bzw. PatientInnen abgesichert, d. h., eine Verknüpfung verschiedener Datenquellen ist selten vorgesehen. Diese Hürde wurde bereits von der Medizininformatik-Initiative [41] erkannt und ein sogenannter Broad Consent definiert, der eine dif-

ferenzierte Zustimmung zu einzelnen Verwendungsoptionen abbildet und der durch Datenschutzverantwortliche der Länder akzeptiert wurde. Als zweite, logistische Hürde erwies sich das Fehlen eines übergeordneten bzw. harmonisierten Treuhandkonzepts. Personenidentifizierende und klinische Daten werden durch Dateneigner in lokalen Treuhandstellen verwaltet und unterliegen individuellen Pseudonymisierungsverfahren. Für eine effektivere, übergreifende Nutzung der Daten ist ein übergreifendes Treuhand-

konzept erforderlich, das die einzelnen Treuhandstellen miteinander koppelt. Dadurch könnten spezifische Pseudonyme studien- bzw. organisationsübergreifend unter Nutzung spezieller Linkage-Verfahren abgeglichen werden.

Diskussion

Die NFDI4Health Task Force COVID-19 entwickelt Konzepte, Methoden und Werkzeuge, die eine bessere Übersicht über medizinische, epidemiologische und Public-Health-Forschung zu SARS-CoV-2 und COVID-19 erlauben. Komplementiert wird dies durch Empfehlungen zum Verknüpfen verschiedener Datenquellen sowie zur harmonisierten Analyse der Datenqualität von Bild- und anderen Forschungsdaten.

Das Studienportal schafft auf Basis neuer Standards zur gemeinsamen Beschreibung von Studien, Erhebungsinstrumenten und weiteren Dokumenten eine Grundlage zur Integration von Informationen aus bisher getrennten Domänen, wie beispielsweise klinischen und epidemiologischen Studien. Durch die studienübergreifend einheitliche se-

mentische Annotation von Items aus COVID-19-Erhebungsinstrumenten besteht im Gegensatz zu anderen Portalen eine leichte Such- und Vergleichbarkeit nach inhaltlichen Kriterien. Zudem wurde mit COVID-19 preVIEW eine semantische Suchmaschine für COVID-19-Preprints entwickelt, die u. a. einen Überblick über die neuesten Publikationen, inklusive Metadaten, ermöglicht.

Die in dieser Arbeit präsentierten Tools und Services befinden sich noch in der Weiterentwicklung und werden durch Feedback der NutzerInnen fortlaufend optimiert. Dies betrifft sowohl die Standards und deren technische Umsetzung als auch die Integration von neuen Inhalten. Zur schnelleren Umsetzung wurde das Datenmodell zur Beschreibung von Studien und Ressourcen zunächst zwar unter Berücksichtigung bestehender Standards, aber doch als eigenständiges Konzept aufgesetzt. Um die Interoperabilität zu anderen Systemen zu gewährleisten und damit auch die Befüllung der Plattform mit Studien und deren Metadaten über entsprechende Schnittstellen weiter automatisieren zu können, ist zukünftig eine Anbindung des Datenmodells an etablierte Standards wichtig, insbesondere an den Interoperabilitätsstandard HL7 FHIR [21]. Auch die Implementation von weltweit gängigen domänenspezifischen Ontologien (Begrifflichkeiten) aus dem medizinischen Bereich, allen voran SNOMED CT [31, 42], wäre wichtig, da sich hierdurch mächtigere semantische Such- und Klassifikationsmöglichkeiten ergäben. Die verwendeten Werkzeuge wie MICA oder SEEK können komplexere Terminologien abbilden; entsprechende Funktionalitäten befinden sich derzeit in Vorbereitung. Weiterhin ist geplant, die vorhandenen Funktionalitäten von SEEK hinsichtlich Strukturierung und Klassifizierung der Studien und Inhalte umfassender zu nutzen.

Technisch wurden im Rahmen der NFDI4Health Task Force COVID-19 mehrere Web-Frontends als Bestandteile eines umfassenden Studienportals verwendet. Dieses Vorgehen bedarf einer weiteren Evaluation, da die Frontends zwar unterschiedliche Schwerpunkte haben, aber dennoch teils überlappen-

de Funktionalitäten besitzen. Deren Zusammenspiel muss im weiteren Projektverlauf unter Berücksichtigung von NutzerInnen-Feedbacks optimiert werden. Eine weitere Herausforderung ist das manuelle Hinzufügen und Kuratieren neuer Metadaten zur Beschreibung von Studien und weiteren Ressourcen. Dies kann nur im Zusammenspiel mit den Verantwortlichen unter der Bedingung knapper zeitlicher Ressourcen umgesetzt werden, was die Akzeptanz reduziert. Daher werden Optionen erprobt, den Prozess der Informationssammlung zu verschlanken, etwa durch Programmierung geeigneter Tools zur Onlineeingabe. Auch die Nutzung etablierter Lizenzmodelle wie Creative Commons zur Klärung der rechtlichen Rahmenbedingungen für eine Nachnutzung von Studiendokumenten in einem zentralen Portal ist nicht trivial. Solche Lizenzmodelle bedürfen noch einer breiteren Verankerung bei den Forschenden.

Die im Rahmen der NFDI4Health Task Force COVID-19 begonnenen Entwicklungen werden im Rahmen der NFDI4Health aufgegriffen und verstetigt. Dies ist sinnvoll, weil die im Zusammenhang mit der aktuellen Pandemie identifizierten Defizite nicht spezifisch für COVID-19-Forschung sind, sondern Forschungsaktivitäten in Public Health, Epidemiologie und klinischen Studien allgemein betreffen. Daher werden Erkenntnisse aus der NFDI4Health Task Force COVID-19 richtungsweisend für weitere Entwicklungen der NFDI4Health sein. Zur Gewährleistung einer hohen Nachhaltigkeit wird auch mit anderen Initiativen auf nationaler und internationaler Ebene eng kooperiert.

Fazit

Die im Aufbau befindliche Infrastruktur der NFDI4Health Task Force COVID-19 ermöglicht die bessere Auffindbarkeit von Forschung und deren Ergebnissen zu SARS-CoV-2 und COVID-19 aus Public Health, Epidemiologie und klinischen Studien mit Fokus auf Deutschland. Die im Rahmen der NFDI4Health Task Force COVID-19 umgesetzten Entwicklungen sind über COVID-19 hinaus relevant, da die adressierten Defizite und Heraus-

forderungen generisch für die Auffindbarkeit von und den Umgang mit Forschungsdaten sind.

Korrespondenzadresse

Prof. Dr. Carsten Oliver Schmidt

Institut für Community Medicine, Universitätsmedizin Greifswald
Walther-Rathenau-Str. 48, 17475 Greifswald, Deutschland
carsten.schmidt@uni-greifswald.de

Mitglieder der NFDI4Health Task Force COVID-19.

Wolfgang Ahrens (Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Deutschland; Universität Bremen, Fachbereich Mathematik und Informatik, Bremen, Deutschland); Johannes Darms (ZB MED Informationszentrum Lebenswissenschaften, Bonn, Deutschland); Jörg Henke (Universitätsmedizin Greifswald, Institut für Community Medicine, Greifswald, Deutschland); Xiaoming Hu (Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Deutschland); Sophie Klopfenstein (Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Deutschland); Lisa Langnickel (ZB MED Informationszentrum Lebenswissenschaften, Bonn, Deutschland); Bianca Lassen-Schmidt (Fraunhofer MEVIS, Institut für Digitale Medizin, Bremen, Deutschland); Hermann Pohlabein (Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Deutschland); Michael Lieser (Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Deutschland); Anatol-Fiete Näher (Robert Koch-Institut, Berlin, Deutschland); Markus Scholz (Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig, Leipzig, Deutschland); Carina Vorisek (Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Deutschland); Dagmar Waltemath (Universitätsmedizin Greifswald, Institut für Community Medicine, Greifswald, Deutschland); Hannes Wünsche (Robert Koch-Institut, Berlin, Deutschland)

Danksagung. Wir bedanken uns bei der Maelstrom-Forschungsgruppe unter Leitung von Dr. Isabel Fortier und bei Sofiya Koleva von der McGill-Universität in Montreal, Canada, für die semantische Annotation von COVID-19-Erhebungsinstrumenten.

Förderung. Die dieser Publikation zugrunde liegenden Arbeiten wurden im Rahmen der NFDI4Health Task Force COVID-19 (www.nfdi4health.de/task-force-covid-19-2) umgesetzt, mit Förderung durch die Deutsche Forschungsgemeinschaft (DFG, Projektnummer 451265285, PI 345/17-1; SCHM 2744/9-1).

Funding. Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. C.O. Schmidt, J. Fluck, M. Golebiewski, L. Grabenhenrich, H. Hahn, T. Kirsten, S. Klammt, M. Löbe, U. Sax, S. Thun und I. Pigeot geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- World Health Organization (2021) ICTRP search portal. <https://www.who.int/clinical-trials-registry-platform/the-ictrp-search-portal>. Zugegriffen: 24. März 2021
- Deutsches Register Klinischer Studien (2020) <https://www.drks.de/>. Zugegriffen: 24. März 2021
- Robert Koch Institute (2021) Seroprevalence studies in selected population groups. https://www.rki.de/EN/Content/Infections/epidemiology/outbreaks/COVID-19/AK-Studien-english/Sero_Groups.html. Zugegriffen: 24. März 2021
- American Society for Microbiology (2021) COVID-19 research registry. <https://asm.org/COVID/COVID-19-Research-Registry/Epidemiology>. Zugegriffen: 24. März 2021
- Medizininformatik-Initiative (2021) Covid-19-Forschungsübersicht. <https://www.medizininformatik-initiative.de/en/node/410>. Zugegriffen: 24. März 2021
- Rat für Sozial- und Wirtschaftsdaten (2021) Studien zur Corona-Pandemie. <https://www.konsortswd.de/ratswd/themen/corona/studien/>. Zugegriffen: 24. März 2021
- 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. *Int J Epidemiol* 46:1372–1378
- Dugas M, Neuhaus P, Meidt A et al (2016) Portal of medical data models: Information infrastructure for medical research and healthcare (Database (Oxford) 2016:bav121)
- Wolstencroft K, Owen S, Krebs O et al (2015) SEEK: A systems biology data and model management platform. *BMC Syst Biol* 9:33
- Netzwerk Universitätsmedizin (2021) <https://www.netzwerk-universitaetsmedizin.de/>. Zugegriffen: 24. März 2021
- Sass J, Bartschke A, Lehne M et al (2020) The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Med Inform Decis Mak* 20:341
- GESIS – Leibniz-Institut für Sozialwissenschaften (2021) Erhebungsinstrumente. <https://www.gesis.org/angebot/studien-planen-und-daten-erheben/erhebungsinstrumente>. Zugegriffen: 24. März 2021
- NFDI4Health (2021) NFDI4Health – Eine Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten. <https://www.nfdi4health.de/>. Zugegriffen: 24. März 2021
- Wilkinson MD, Dumontier M, Aalbersberg J et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018
- NFDI4Health (2021) Task force COVID-19. <https://www.nfdi4health.de/de/task-force-covid-19/>. Zugegriffen: 24. März 2021
- Merino-Martinez R, Norlin L, van Enckevort D et al (2016) Toward global biobank integration by implementation of the Minimum Information About Biobank data Sharing (MIABIS 2.0 Core). *Biopreserv Biobank* 14:298–306
- Eklund N, Andrianarisoa NH, Van Enckevort E et al (2020) Extending the minimum information about biobank data sharing terminology to describe samples, sample donors, and events. *Biopreserv Biobank* 18:155–164
- Bergeron J, Doiron D, Marcon Y, Ferretti V, Fortier I (2018) Fostering population-based cohort data discovery: the maelstrom research cataloguing toolkit. *PLoS ONE* 13(7):e200926
- DataCite Schema (2019) Datacite metadata schema 4.3. <https://doi.org/10.14454/7xq3-zf69>. Zugegriffen: 24. März 2021
- HL7 FHIR (2019) Documentation index. <http://hl7.org/fhir/documentation.html>. Zugegriffen: 24. März 2021
- NFDI4Health Task Force COVID-19, Schmidt CO, Darms J, Shutsko A et al (2021) Facilitating study and item level browsing for clinical and epidemiological COVID-19 studies. *Stud Health Technol Inform* 281:794–798. <https://doi.org/10.3233/SHTI210284>
- FAIRDOMHub (2021) NFDI4Health Task Force COVID-19 metadata schema mapping. https://fairdomhub.org/data_files/3973. Zugegriffen: 24. März 2021
- Creative Commons (2021) Homepage. <https://creativecommons.org/>. Zugegriffen: 24. März 2021
- NFDI4Health (2021) German central health study hub Covid-19. <https://covid19.studyhub.nfdi4health.de>. Zugegriffen: 24. März 2021
- Wolstencroft K, Krebs O, Snoep JL et al (2017) FAIRDOMHub: A repository and collaboration environment for sharing systems biology research. *Nucleic Acids Res* 45:D404–D407
- Mica (2020) NFDI4Health – COVID-19 questionnaires. <https://mica.covid19.studyhub.nfdi4health.de/>. Zugegriffen: 11. Mai 2021
- Leaman R, Lu Z (2016) TaggerOne: joint named entity recognition and normalization with semi-Markov models. *Bioinformatics* 32(18):2839–2846
- Wei C-H, Kao H-Y, Lu Z (2015) GNormPlus: an integrative approach for tagging genes, gene families, and protein domains. *Biomed Res Int* 2015:918710
- Funk C, Baumgartner W, Garcia B et al (2014) Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinform* 15(1):59
- Deutsche Zentralbibliothek für Medizin (ZB MED) – Informationszentrum Lebenswissenschaften (2021) preVIEW: COVID-19. <https://preview.zbmed.de>. Zugegriffen: 11. Mai 2021
- Schmidt CO, Struckmann S, Enzenbach C et al (2021) Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. *BMC Med Res Methodol* 21:63
- Richter A, Schmidt CO, Struckmann S (2021) dataquieR—R package version 1.04. <https://cran.r-project.org/web/packages/dataquieR/index.html>. Zugegriffen: 24. März 2021
- Universitätsmedizin Greifswald (2021) Standards and tools for data quality assessment in epidemiological studies. <https://dfg-qa.ship-med.uni-greifswald.de/>. Zugegriffen: 11. Mai 2021
- Grand Challenge (2021) Grand Challenge—A platform for end-to-end development of machine learning solutions in biomedical imaging. <https://grand-challenge.org/>. Zugegriffen: 11. Mai 2021
- DICOM (2021) Digital imaging and communications in medicine. <https://www.dicomstandard.org/>. Zugegriffen: 11. Mai 2021
- March S, Antoni M, Kieschke J et al (2018) Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. *Gesundheitswesen* 80(03):e20–e31
- Christen P, Ranbaduge T, Schnell R (2020) Linking sensitive data. Methods and techniques for practical privacy-preserving information sharing. Springer, Cham
- Langner I, Riedel O, Czwikla J et al (2020) Linkage of routine data to other data sources in Germany: A practical example illustrating challenges and solutions. *Gesundheitswesen* 82(S02):S117–S121
- March S, Andrich S, Drepper J et al (2019) Gute Praxis Datenlinkage (GPD). *Gesundheitswesen* 81(8/9):636–650
- Semler SC, Wissing F, Heyder R (2018) German Medical Informatics Initiative. *Methods Inf Med* 57(S01):e50–e56
- Millar J (2016) The need for a global language—SNOMED CT introduction. *Stud Health Technol Inform* 225:683–685
- Medizininformatik-Initiative (2020) Delivering SNOMED CT. Den Einstieg in die Nutzung von SNOMED CT erleichtern. <https://www.medizininformatik-initiative.de/de/den-einstieg-die-nutzung-von-snomed-ct-erleichtern>. Zugegriffen: 24. März 2021