

NFDI

Nationale Forschungsdateninfrastruktur und Datenmanagement in NFDI4Health

Prof. Dr. Juliane Fluck

Leitung Wissensmanagement

ZB MED - Informationszentrum Lebenswissenschaften

Thementag zum Umgang mit Forschungsdaten,
Greifswald, 13. Oktober 2020

➤ Sprecherin NFDI4Health (Start Oktober 2020)

➤ Seit Oktober 2018 bei ZB MED

Leitung der Forschungsabteilung Wissensmanagement mit gleichzeitiger Professur an der Universität Bonn

➤ 18 Jahre als Wissenschaftlerin bei
Fraunhofer SCAI Abteilung Bioinformatik

- Text mining/
Informationsextraktion
- Image mining
- Data mining
- Semantische Datenintegration

Biology
Medicine
Chemistry
Pharmacology



- Teil I: Warum gewinnen Daten und Forschungsdatenmanagement an Relevanz
- Teil II: Allgemeine Informationen zu NFDI
- Teil III; NFDI4Health Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten und
NFDI4Health Task Force COVID-19

Paradigma: Datengetriebene Wissenschaft

Beispiele von ‚Large Scale‘ Experimenten:

➤ Hochdurchsatzverfahren

➤ Next Generation Sequencing

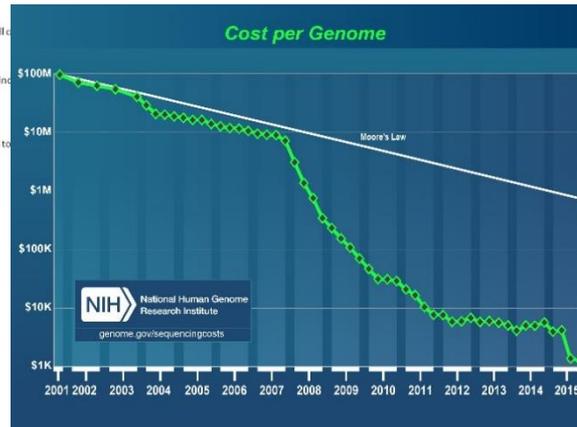
High Throughput Screening (HTS)

1. Dispense Library 2. Dispense Target/Reagent 3. Assay Readout

Examples of Screening Libraries

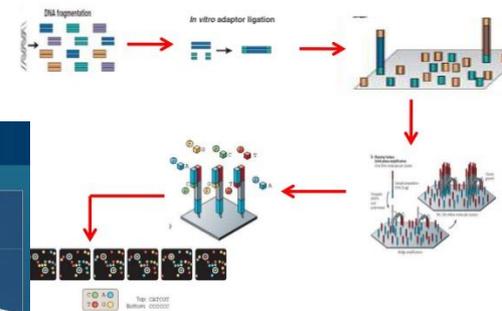
- Pharmaceutical Industry
 - Drug compounds, screened against a protein/cell
 - Natural products
- Biotechnology Industry
 - Proteins, screened against another protein for binding
 - PCR primers, used for gene expression
 - siRNA, for genome-wide gene silencing studies
- Environmental/Public Health
 - Organic compounds screened against living cells to

<https://nanohub.org/>



<https://www.genome.gov/sequencingcosts/>

TECHNOLOGY: NEXT GENERATION SEQUENCING



<https://www.slideshare.net/lampweb/>

[cal-applications-of-next-generation-sequencing](#)

Drastische Kostenreduktion: unter 1000 Euro

Massives Wachstum von Daten

- Standards für Datenpublikation nur in bestimmten Teilbereichen etabliert; FAIR Data Principles lange nicht erreicht
- Wenn Daten abgelegt werden, sind sie nur unvollständig annotiert und selten interoperabel
- Daten Kuratierung und Integration wird zum Hauptfaktor für Analyse heterogener Daten

Höherer (maschinen-lesbarer) Wissensbedarf zur Interpretation der Daten

- kaum noch ohne algorithmische Unterstützung möglich

Massives Wachstum unstrukturierter Information (z.B. wissenschaftliche Publikation)

- Information und Wissen wird zumeist durch Text transportiert

Forschungsdatenmanagement bei ZB MED: Services und Forschung

Kompetenz-, Beratungs- und Vernetzungszentrum in den Lebenswissenschaften

| Beraten | Publizieren | Vernetzen | Forschung |
|--|--|---|--|
| <ul style="list-style-type: none"> ▶ FAQs ▶ Persönliche Beratung ▶ Workshops mit unterschiedlichen Zielgruppen ▶ Vorträge ▶ Tutorials | <ul style="list-style-type: none"> ▶ DOI-Service ▶ Lebenswissenschaftliche Forschungsdaten-repositorien (auf Basis re3data.org) ▶ Publikationsinfrastruktur (Fachrepositorium Lebenswissenschaften) | <ul style="list-style-type: none"> ▶ DINI/nestor AG Forschungsdaten ▶ Vorträge auf lebenswissenschaftlichen und bibliothekarischen Tagungen ▶ AK Forschungsdaten der Leibniz-Gemeinschaft ▶ Research Data Alliance (international und national) | <ul style="list-style-type: none"> ▶ Qualitätssicherung von Forschungsdaten ▶ Pilotprojekte zu systematischem Forschungsdatenmanagement ▶ Beteiligung an Konsortien zum Aufbau von Forschungsdateninfrastruktur (z.B. MI-I, NFDI4Life) ▶ |

Services und Forschung entlang des Lebenszyklus von Forschungsdaten

NFDI4Health
Suchportal für personenbezogene Gesundheitsdaten



Forschungsdaten suchen und nutzen

Forschungsdaten archivieren

Digitales Langzeitarchiv

Forschungsdatenmanagement planen

RDMO4Life – dynamische Forschungsdatenmanagement-Planung in den Lebenswissenschaften



Forschungsdaten publizieren

NFDI4Health – Publikation von Studien

Wegweiser Elektronische Laborbücher

Forschungsdaten dokumentieren

Forschungsdaten referenzierbar machen

DOI-Service

Deutsche Datenstrategie

- ▶ BMBF Aktionsplan: ‚Natürlich. Digital. Nachhaltig‘

https://www.bmbf.de/upload_filestore/pub/Natuerlich_Digital_Nachhaltig.pdf

- ▶ ‚Datenstrategie der Bundesregierung‘

<https://www.bundestkanzlerin.de/bkin-de/mediathek/live-aus-dem-kanzleramt/>

[expertenanhoerung-zur-datenstrategie-im-kanzleramt-1715230!mediathek?query](https://www.bundestkanzlerin.de/bkin-de/mediathek/live-aus-dem-kanzleramt/expertenanhoerung-zur-datenstrategie-im-kanzleramt-1715230!mediathek?query)

- ▶ **Nationale Forschungsdateninfrastruktur**

basiert auf Empfehlungen des Rats für Informationsinfrastrukturen (RII)

- ▶ **Neues RII Papier: Herausforderung Datenqualität**

- ▶ Wissenschaftsrat plant "Papier zum Wandel durch datenintensive Forschung"

Vorlauf NFDI

- ❑ **2016** empfiehlt der Rat für Informationsinfrastrukturen (RfII) Aufbau einer Nationalen Forschungsdateninfrastruktur „Dieses soll als bundesweites, verteiltes und wachsendes Netzwerk arbeitsteilig angelegt sein“*
- ❑ **2018** Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungs-dateninfrastruktur (NFDI) vom 26.11. 2018
 - Mit der NFDI sollen die heute oft dezentral, projektförmig und temporär gelagerten **Datenbestände von Wissenschaft und Forschung für das deutsche Wissenschaftssystem systematisch erschlossen werden.**“
 - „Die NFDI soll Standards im Datenmanagement setzen und als **digitaler, regional verteilter und vernetzter Wissensspeicher** Forschungsdaten nachhaltig sichern und nutzbar machen.“
 - „Nach den sogenannten **FAIR-Prinzipien*** sollen Forschungsdaten auffindbar, zugänglich, interoperabel und nachnutzbar sein.“

*RfII (2016): Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, 160 S.

Rollen in den Konsortien

▪ **Antragstellende Einrichtung**

- koordiniert die Antragstellung/Konsortium
- verwaltet zentral die Mittel des Konsortiums
- stellt den Sprecher des Konsortiums

▪ **Mitantragstellende Einrichtungen**

- stellen Co-Sprecher,
- die für Teilbereiche verantwortlich sind
- erhalten ihre Mittel über die antragstellende Einrichtung

▪ **Beteiligte**

- leisten einen nachhaltigen Beitrag zum Arbeitsprogramm
- das Konsortium kann Mittel für die Beteiligten beantragen

Bewilligte NFDI Konsortien

9 Konsortien wurden in der ersten Runde bewilligt und sind seit dem 01. Oktober am Start:

Medizin

- **NFDI4Health** Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten
- **GHGA** Deutsches Humangenom-Phenomarchiv

Biologie

- **NFDI4BioDiversity**: Biodiversität, Ökologie und Umweltdaten
- **DataPLANT**: Daten in Pflanzen-Grundlagenforschung

Bewilligte NFDI Konsortien

Natur- und Ingenieurwissenschaften

- NFDI4Ing Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften
- NFDI4cat: NFDI für Wissenschaften mit Bezug zur Katalyse
- NFDI4Chem: Fachkonsortium Chemie in der NFDI

Geistes- und Sozialwissenschaften

- KonsortSWD Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschafts-wissenschaften
- NFDI4Culture – Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern

Weitere Informationen:

<https://www.dfg.de/foerderung/programme/nfdi/>

NFDI Verein als übergeordnete Struktur

- **Direktorat in Karlsruhe, Direktor: Prof. York Sure Vetter**
- **Im Verein kann jeder Mitglied werden unabhängig von einer NFDI-Förderung**
- **Übergeordnete Arbeitsgruppen zur Etablierung einer gemeinsamen NFDI als Endziel**
- **Beispiele möglicher Themen**
 - Kollaborative Governance und allgemeiner Rahmen
 - Einbeziehung der Community (Nutzer)
 - Technische Infrastruktur und Konzepte
 - Legale und ethische Aspekte
 -

Weitere Informationen: <https://www.nfdi.de/>



Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten

Konsortium (I) - Leitung

Zentrale Dienste bei ZB MED – Informationszentrum Lebenswissenschaften

- Deutsche Zentralbibliothek für Medizin
 - Nationaler Auftrag für Informationsversorgung und Infrastrukturentwicklung
 - Ideal geeignet, um zentrale Services von NFDI4Health nachhaltig anzubieten
- ZB MED als Dachorganisation für
 - Bibliotheken / Communities zur Entwicklung von Standards und Infrastrukturen
 - Lebenswissenschaftliche / biomedizinisch-informatische Communities

Domänen-Expertise und Co-Lead

Iris Pigeot, Leiterin des

Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS

Konsortium (II) – Beteiligung vor Ort

Universitätsmedizin Greifswald
Institut für Community Medicine

➤ **Prof. Dr. Carsten Oliver Schmidt**

Leiter Funktionsbereich **Qualität in der Gesundheitsforschung**

➤ **Prof. Dr. Dagmar Walthemath**

Abt. Versorgungsepidemiologie und Community Health /
Forschungsbereich **Medizininformatik**

Konsortium (III)

- Forschungsinstitute involviert in Datenerhebung und -analyse personenbezogener Gesundheitsdaten



- Infrastruktureinrichtungen / Experten in Entwicklung von Standards und Methoden



- Institutionen mit Fokus auf Einbezug der Nutzer-Community



- 47 namhafte Institutionen als Beteiligte

Mission von NFDI4Health

Unsere Mission ist es,

- den **Mehrwert der Forschung** in Epidemiologie, Public Health und in auf klinischen Studien gestützter Medizin zu erhöhen,
- indem qualitätsgesicherte Daten international gemäß den **FAIR-Prinzipien** zugänglich gemacht werden.

Aktuelle **COVID-19-Pandemie** zeigt diesen Bedarf!

Untergruppe von NFDI4Health hat zum 1. Juli DFG-Förderung erhalten:

NFDI4Health Task Force COVID-19

Laufzeit: 07.2020 – 06.2023; Fördervolumen insgesamt € 1.065.084

Koordination: Iris Pigiot, BIPS

Aktuelle Herausforderungen

- **Auffindbarkeit** der Daten oft beeinträchtigt
 - Trotz existierender Studienregister-Portale
- **Metadatenbeschreibung** häufig fehlend
- **Interoperabilität** zwischen verschiedenen Datenquellen in der Regel nicht gegeben
 - Jede Institution wendet eigene Standards an
- Möglichkeiten für **geregelten Datenzugang** sehr **eingeschränkt**
- Kaum **Data Mining-** und **Maschine Learning-Ansätze** möglich

Strategie

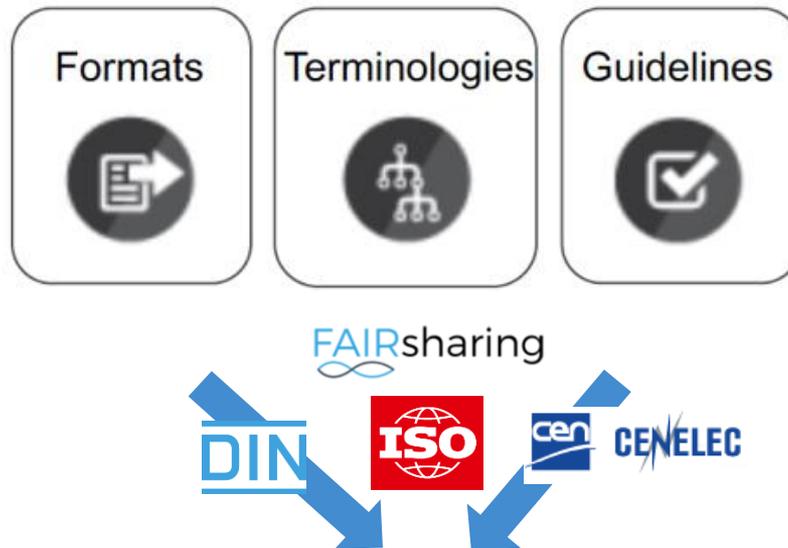
- **Verbesserung der Auffindbarkeit von Gesundheitsdaten**
 - Unterstützung von Datenpublikationen, Central Search Hub
 - Standardisierung von Metadaten, Verbesserung der Interoperabilität
- **Implementierung eines übergeordneten Datenzugangs- und Datennutzungsprozesses (Central Access Point)**
 - Gewährleistung einer Nutzung ausschließlich im Einklang mit Einverständniserklärungen und Datenschutzrichtlinien
- **Weiterentwicklung von Diensten**, die kontrollierten Zugriff auf verteilt vorliegende Daten mittels Analysetools erlauben
- **Entwicklung von Diensten für dynamisches und sicheres Linkage** von Primär-, Sekundär- und Registerdaten

Unser Portfolio: Konzepte/Lösungen, um Datenschutz zu wahren und Datenzugang zu ermöglichen

- Generische Konzepte und Dienste für Datenaustausch und Datenschutz
- Legales Grundgerüst für Einwilligungsstandards und Datenzugriffsverfahren
- Entwicklung von Konzepten und Methoden zur Bewertung von **Datenschutzrisiken**
- Pilotierung der Implementation von Datenstandards und von Zugangswegen für **Record/Daten Linkage von (Sekundär-)daten**
- **Synthetische Gesundheitsdaten** (Konzept der differentiellen Privatheit)
- **Verteilte Datenanalyse**



Unser Portfolio: Standardisierung



Standards

- Datenpublikationsrichtlinie
- (Meta-)datenstandards
- Datenqualitätsstandards

Harmonisierung der existierenden Standards & Adaption & Erweiterung

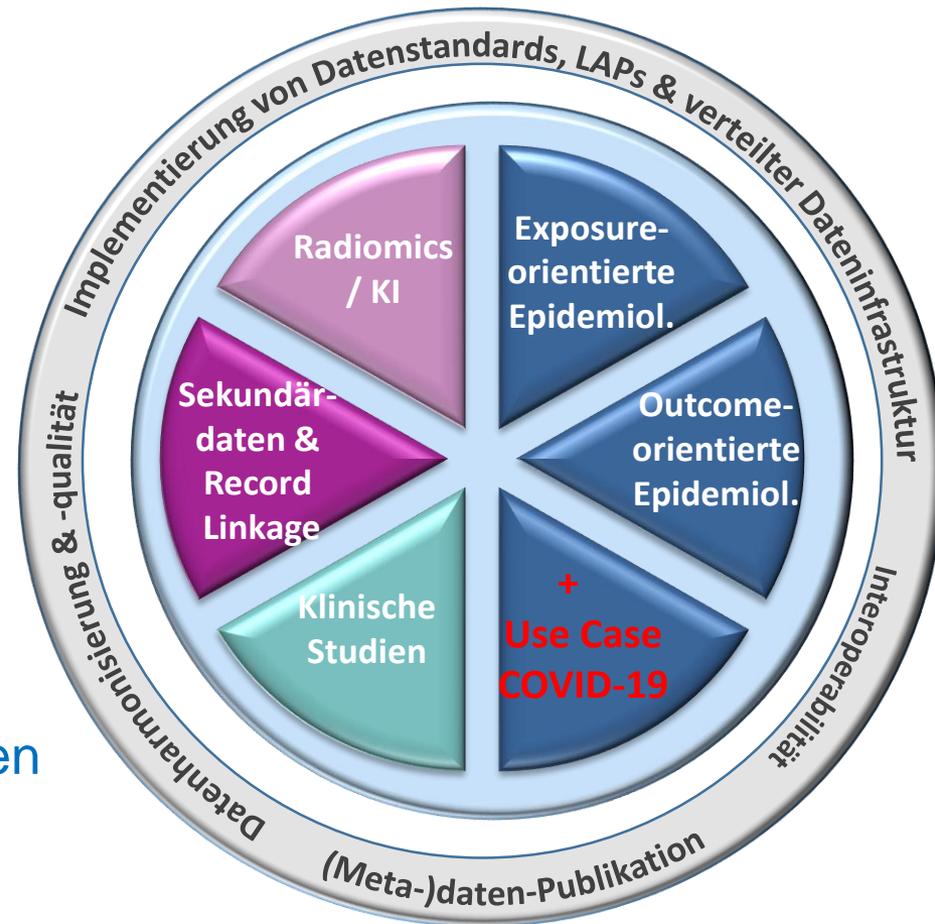
Organisationen, Initiativen: MII, GMDS, TMF, GoFAIR, GA4GH, CDISC, HL7, ISO, CEN, DIN, ...

Formate und Terminologien: ODM, BRIDG, FHIR, ICD, LOINC, ...

Identifier: WHO-UTN, DOI, ...

wurden ausgewählt, um

- verschiedene Nutzer-Communities einzubinden
- die relevanten Aspekte zu adressieren und Blaupausen zu liefern
- Services und Standards in wichtigen Datenquellen zu implementieren
- Ansätze zur Sicherstellung der Vertraulichkeit und zur Einhaltung Datenschutz- und ethischen Richtlinien aufzuzeigen



Exkurs in die aktuelle Realität der Covid-19 Pandemie

1. Einzigartige Zusammenarbeit von Wissenschaftlern zur Erforschung von Covid-19
2. Einzigartige schnelle Unterstützung durch verschiedenste Funding-Initiativen
3. Ganz wichtig für Deutschland: brillante (Vor)Arbeit der Medizininformatik-Initiative
4. Hoher Informationsbedarf, der nicht durch die üblichen Publikationskanäle gedeckt werden kann

-> Relevanz von

- Pressemitteilungen
- Twitter
- Preprints

-> ZB MED Covid-19 Hub

ZB MED Covid-19 Informationen (1)



→ ZB MED COVID-19 Hub

To support research we offer various tools and data sets around SARS-CoV-2 and COVID-19.

[→ learn more](#)



→ Publications and other sources

ZB MED reviews freely accessible scientific publications and valid internet sources lists them here in a systematic way.

[→ learn more](#)



→ LIVIVO COVID-19 collection

About 50,000 entries from various sources summarized in the ZB MED search portal.

[→ learn more](#)



→ Share your data on SARS-CoV-2/COVID-19

ZB MED offers publishing infrastructure and advice.

[→ learn more](#)

ZB MED Covid-19 Informationen (2)

HOME

OUR SERVICES ▾

EXTERNAL RESOURCES ▾

CONTACT

LEGAL NOTICE

PREPRINT VIEWER
LITERATURE OVERVIEW
LITERATURE SEARCH PORTAL
BIOINFORMATICS TOOLS
PUBLISHING DATA

TEXT MINING AND NLP
BIOINFORMATICS DATA AND TOOLS
MEDICAL DATA AND TOOLS
INITIATIVES AND FUNDING OPPORTUNITIES
HACKATHONS AND OTHER EVENTS



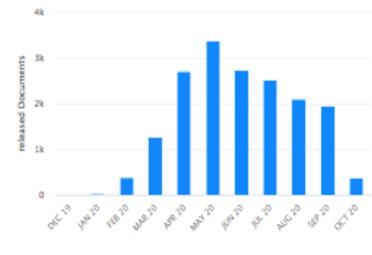
COVID-19 Pandemic: ZB MED establishes special services to support research

Gradual expansion of services to support research on SARS-CoV-2 in the field of virology and related areas



preVIEW: COVID-19

Corpus overview



advanced search

« 1 2 3 4 5 ... 1749 1750 » displaying 1 - 10 records in total 17499 10 records per page

Expand/collapse all abstracts

Kidney injury molecule-1 is a potential receptor for SARS-CoV-2

Authors: Chen Yang; Yu Zhang; Hong Chen; Yuchen Chen; Dong Yang; Ziwei Shen; Xiaomu Wang; Xinran Liu; Mingxui Xiong; Kun Huang

doi:10.1101/2020.10.09.334052 Date: 2020-10-10 Source: bioRxiv

Abstract

COVID-19 patients present high incidence of kidney abnormalities, which are associated with poor prognosis and high mortality. Identification of SARS-CoV-2 in kidney of COVID-19 patients suggests renal tropism and direct infection. Presently, it is generally recognized that SARS-CoV-2 initiates invasion through binding of receptor-binding domain (RBD) of spike protein to host cell-membrane receptor ACE2, however, whether there is additional target of SARS-CoV-2 in kidney remains unclear. Kidney injury molecule-1 (KIM1) is a transmembrane protein that drastically up-regulated after renal injury. Here, binding between SARS-CoV-2-RBD and the extracellular Ig V domain of KIM1 was identified by molecular simulations and co-immunoprecipitation, which was comparable in affinity to that of ACE2 to SARS-CoV-2. Moreover, KIM1 facilitated cell entry of SARS-CoV-2-RBD, which was potentially blocked by a rationally designed KIM1-derived polypeptide. Together, the findings suggest KIM1 may mediate and exacerbate SARS-CoV-2 infection in a vicious cycle, and KIM1 could be further explored as a therapeutic target.

Soluble Spike DNA vaccine provides long-term protective immunity against SAR-CoV-2 in mice and nonhuman primates

Authors: Yong Bok Seo; You Suk Suh; Ji In Ryu; Hwanhee Jang; Hanseul Oh; Bon-Sang Koo; Sang-Hwan Seo; Jung Joo Hong; Manki Song; Sung-Joo Kim; Young Chul Sung

doi:10.1101/2020.10.09.334136 Date: 2020-10-10 Source: bioRxiv

Abstract

SARS-CoV-2 Elicits Robust Adaptive Immune Responses Regardless of Disease Severity

Authors: Stine Sofie Frank Nielsen; Line K Vibholm; Ida Monrad; Rikke Olesen; Giacomo S Frattari; Marie H Pahus; Jesper F Hejlen; Jesper D Gunst; Christian Erikstrup; Andreas Hollefer; Rune Hartmann; Lars Østergaard; Ole S Søgaard; Mariane H Schleimann; Martin Tolstrup

doi:10.1101/2020.10.08.331645 Date: 2020-10-09 Source: bioRxiv

Abstract

Remote home monitoring (virtual wards) during the COVID-19 pandemic: a living systematic review

Authors: Cecilia Vindrola-Padros; Kelly Singh; Manbinder S Sidhu; Theo Georgioulou; Chris Sherlaw-Johnson; Sonila M Tomini; Matthew Inada-Kim; Karen Kirkham; Allison Streetly; Naomi J Fulop; Rune Hartmann; Lars Østergaard; Ole S Søgaard; Mariane H Schleimann; Martin Tolstrup

doi:10.1101/2020.10.07.20208587 Date: 2020-10-09 Source: medRxiv

Overview

- MeSH Disease
 - Coronavirus Infections (3747)
 - Infections (3470)
 - Severe Acute Respiratory Syndrome (3021)
 - Death (1421)
 - Pneumonia (1052)

Human Phenotype

- Pneumonia (1344)
- Fever (803)
- Cough (659)
- Hypertension (471)
- Anxiety (462)

Transmission

- age categories (3697)
- Transmission (3285)
- gender (1718)
- fomite (1353)
- contact tracing (1217)

Seroprevalence

- performance (1882)
- serum sample (1649)
- antibody (1250)
- seroprevalence (1232)
- serological test (652)

The ZB MED preprint Viewer preVIEW includes all COVID-19 related preprints from medRxiv and bioRxiv, from ChemRxiv, from ResearchSquare, from arXiv and from Preprints.org and is updated on a daily basis (7am CET/CEST). The web page can also be accessed via API.

Sources

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



Annotations

- All
- None
- MeSH Disease
- Human Phenotype
- Transmission
- Seroprevalence

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



nfdi4health is one of the funded consortia of the National Research Data Infrastructure programme of the DFG.



NFDI4Health Task Force COVID-19

- Aufbau einer bundesweiten **Informations- und Beratungsinfrastruktur** für COVID-19-Forschung
- Fokus über klinische Patientendaten hinaus auf **Folgen des Pandemieausbruchs** für öffentliche Gesundheit, wie z.B.
 - Morbidität
 - Mortalität
 - Inanspruchnahme der Gesundheitsversorgung
 - Lebensqualität
 - Auswirkungen sozialer Isolation

Ziele der Task Force COVID-19 (I)

- Verbesserung der **FAIRness** von COVID-19-Forschungsdatensammlungen
 - Inventar deutscher Studien zu COVID-19 mit strukturierten Gesundheitsdaten → zentralisierter Datenkatalog mit Suchfunktion
 - Verbesserter Zugang
 - Maßnahmen für verbesserten Datenaustausch
 - Einheitliche Bewertungen der Datenqualität

-> erste Version des Suchportals mit einem Überblick deutscher Studien soll Ende des Monats freigeschaltet werden

Ziele der Task Force COVID-19 (II)

- Förderung der zeitnahen Durchführung von **harmonisierten Forschungsarbeiten** im Zusammenhang mit COVID-19
 - Leitfäden, Schulungsmaterialien und Standards z.B. zum Datenmanagement, zu Erhebungsinstrumenten, zur Nutzung von Routinedaten
 - Qualitätsgeprüfte Analyseroutinen und Optionen zur interaktiven webbasierten Datenexploration und -visualisierung
 - Services zur (teil-)automatisierten Auswertung medizinischer Bilddaten
- > erste eLearning-Kursen zur Qualifizierung von Prüfern und Mitgliedern von Studienteams während Kontaktbeschränkungen etabliert (verantwortlich KKSNI)**

Ziele der Task Force COVID-19 (III)

- Methoden zum **Record Linkage** verschiedener COVID-19-Datensätze
 - Konzept zur Verknüpfung relevanter Datenquellen
 - Krankenhausdaten
 - Ambulante Verschreibungsdaten
 - Sequenzierdaten
 - Bilddaten
 - Soziodemographische und psychosoziale Daten
 - Inklusive Datenschutzkonzept und Vorlage für informierte Einwilligung

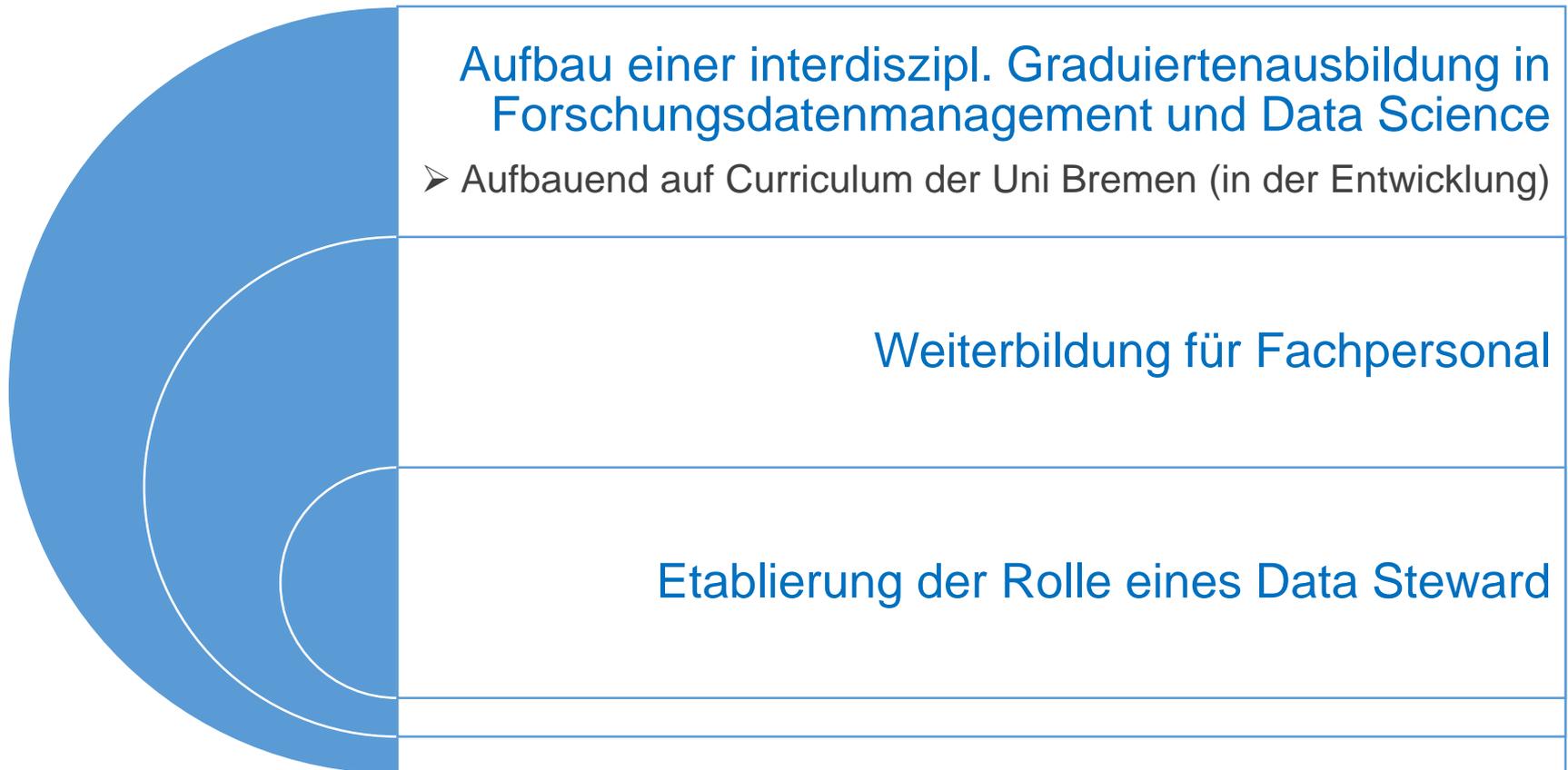
Zurück zu NFDI4Health: Vielfältige Elemente der Nutzereinbindung



2. Communityworkshop am 6. November 2020

Weitere Informationen: <https://www.nfdi4health.de/community-workshop>

Unser Portfolio: Training und Weiterbildung



- Dissemination des Fortbildungsmaterials und der Curricula als offene Ressource
- Evaluation und Feedback als Teil der allgemeinen Strategie zur Nutzereinbindung

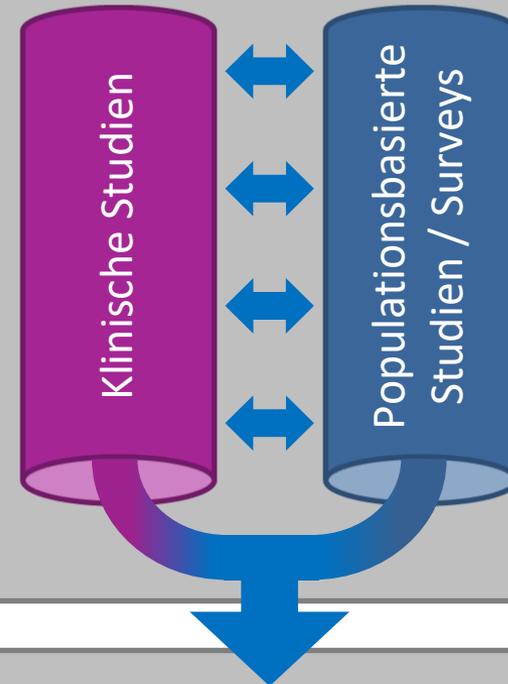
Verschmelzung von epidemiologischer, Public Health- und klinischer Forschung

Gemeinsame Such- und Zugriffsservices

Gemeinsame Standards für (Meta-)Daten

Gemeinsame Datenstrukturen

Gemeinsame Einverständniserklärungen



Öffnung der Säulen und Verstärkung der wissenschaftlichen Nachnutzung von personenbezogenen Gesundheitsdaten

Vielen Dank für die Aufmerksamkeit!

Kontakt

Juliane Fluck

ZB MED – Informationszentrum
Lebenswissenschaften

fluck@zbmed.de

www.nfdi4health.de
