SOZIALMEDIZNISCHER DATENVERBUND GESUNDHEITSMETROPOLE RUHR

Data4Health.Ruhr

Rahmenskizze

-

Fördermaßnahme „Innovation & Strukturwandel“ im Förderbereich „WIR! – Wandel durch Innovation in der Region“ des Bundesministeriums für Bildung und Forschung

Vorbemerkung: *Bei der vorliegenden Rahmenskizze handelt es sich nicht um einen eigenen Projektantrag, sondern um die Beschreibung des Gesamtkontextes der sukzessive zu beantragenden Einzelprojektvorhaben im Entwicklungsschwerpunkt „Sozialmedizinischer Datenverbund“ des Innovationsraums „Urban Health Data Space“ der Digital Health Factory Ruhr. Die Rahmenskizze bildet somit die inhaltliche Klammer um diese Projektfamilie und stellt dabei die Basis im Sinne des Strategieentwicklungsprozesses zu diesem Entwicklungsschwerpunkt dar.*

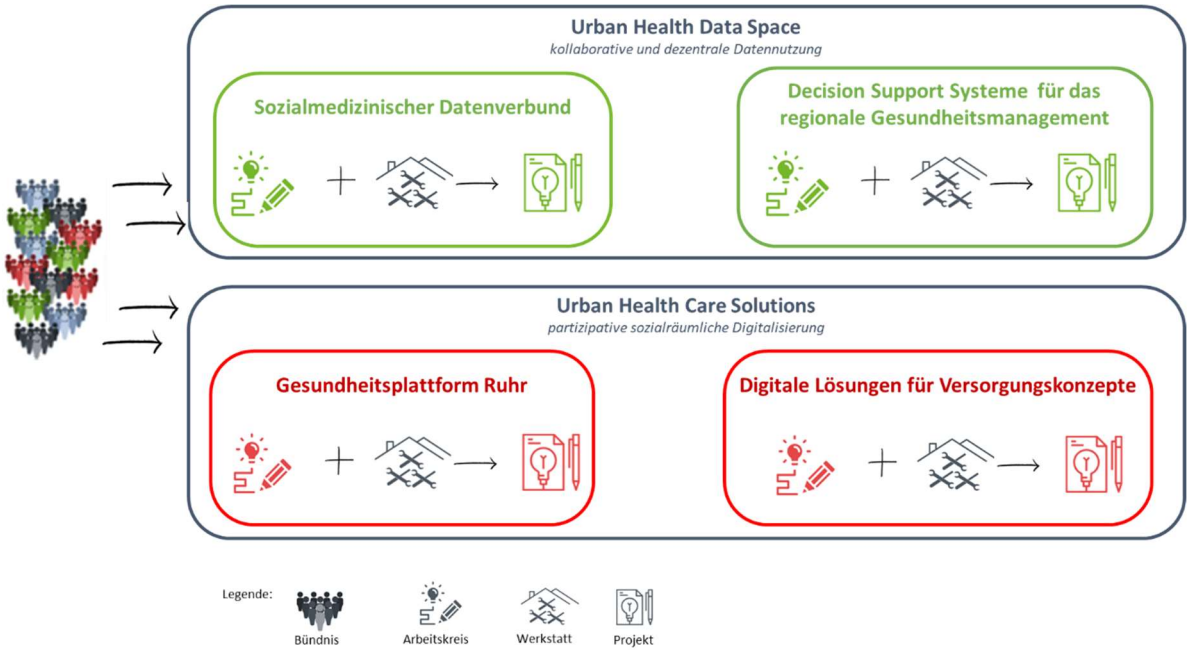

Abbildung 1: Innovationsräume und Entwicklungsschwerpunkte der *Digital Health Factory Ruhr*

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

I. Adressierte Problemlage: Mangelnde Zusammenführung vorhandener Datengrundlagen für eine passgenaue Gesundheitsförderung und -versorgung

Hinter einer in der Gesamtbetrachtung guten Versorgungssituation großstädtischer Räume verbirgt sich eine stark ausgeprägte Tendenz zur sozialräumlichen Segregation von Stadtgebieten, in denen sich gesundheitliche Auffälligkeiten, soziale Benachteiligungen und zunehmende Defizite in der Gesundheitsversorgung in brisanter Weise verschränken. Genau dort, wo die Anteile ärmerer Menschen hoch, Migrationshintergründe stark, die Ressourcen vulnerabler Gruppen schwach ausgeprägt, Erkrankungsraten hoch und Lebenserwartungen eher niedrig sind, nehmen die Probleme in der Bereitstellung und Inanspruchnahme von Versorgungsangeboten zu. Gleichzeitig gibt es übergreifende Veränderungen in den urbanen Arbeits- und Lebensweisen, deren Spektrum vom Ernährungs- und Bewegungsverhalten bis hin zur psychischen Wirkung urbaner Stressfaktoren reichen.

(Konzept zur Digital Health Factory Ruhr – Digitale Lösungen für die Gesundheitsversorgung in Ballungsräumen)

Fehlende Passgenauigkeit in der sozialraumbezogenen Gesundheitsförderung und -versorgung

Stadtbezirke, die in Sozialmonitorings einen niedrigen Sozialstatus aufweisen, fallen gleichzeitig durch eine höhere Krankheitslast bei verschiedenen Erkrankungen der Bevölkerung auf. Die Gesundheitsberichterstattung des Landes Nordrhein-Westfalen zeigt dies sehr deutlich für das Ruhrgebiet (Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen 2020), wobei innerhalb der Kommunen erhebliche Unterschiede auf Ebene der Stadtteile bestehen. Darüber hinaus lassen sich bestimmte – meist als vulnerabel bezeichnete – Gruppen quer durch alle Bevölkerungsschichten und unabhängig von räumlichen Abgrenzungen finden. Um diesen Ausprägungen der gesundheitsbezogenen sozialen Ungleichheit zu begegnen, existieren bereits diverse Ansätze, die auf die Unterstützung sowie eine verbesserte Einbindung der einzelnen Gruppen abzielen. Ziel ist hierbei die Schaffung bedarfsgerechter Angebote in Gesundheitsförderung, Prävention und Versorgung.

Hinweise darauf, dass gesundheitsbezogene Angebote in „benachteiligten“ Sozialräumen nicht nur in ihrem Umfang hinter anderen Stadtgebieten zurückbleiben, sondern auch in ihrer qualitativen Ausrichtung den besonderen Bedarfen der heterogenen Gruppen nicht ausreichend gerecht werden, werfen die Frage nach geeigneten Instrumenten für flächendeckend bedarfsorientierte und hochwertige gesundheitsbezogene Angebote in Ballungsräumen auf. Diese beinhalten zum einen präventive Ansätze, die auf die Herstellung einer einheitlich hohen Gesundheitskompetenz oder die Schaffung verbesserter Rahmenbedingungen für Gesundheit (Umweltfaktoren, Bewegungs- u. Erholungsangebote) abzielen. Zum anderen müssen die Instrumente die unterschiedlichen Bedarfe der Menschen für eine sachdienliche Inanspruchnahme von Gesundheitsleistungen adressieren – auch um Fehlversorgungstendenzen entgegenzuwirken.

GEFÖRDERT VOM
wir! Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Zwar kann im Allgemeinen sogar ein erhöhter Bedarf an Gesundheitsleistungen für bestimmte Sozialräume oder Stadtteile festgestellt werden, doch sicher ist bislang nur, dass sich Lösungsansätze nicht in rein quantitativen Maßgaben – wie z.B. den bestehenden Bedarfsindizes, welche u.a. bei der Identifikation von Sonderbedarfen in der Bedarfsplanung herangezogen werden – erschöpfen dürfen. Welche Instrumente für einen bestimmten Sozialraum geeignet sind, um Hemmfaktoren für eine einheitlich hohe Gesundheitsqualität zu reduzieren, kann aber aufgrund einer fragmentierten Datenlage auch durch die Akteure vor Ort derzeit bestenfalls antizipiert werden.

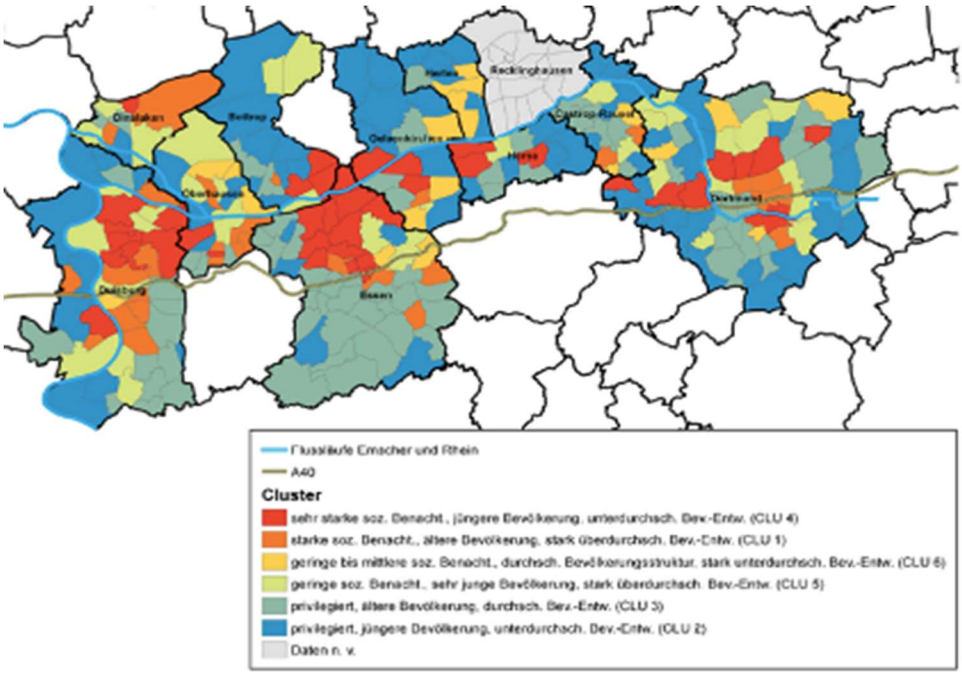

Abbildung 2: Clusterzugehörigkeit der Stadtteile hinsichtlich der Merkmale zur sozialen Benachteiligung, zur Bevölkerungsstruktur und zur Bevölkerungsentwicklung mit dem Ziel der Ausweisung spezifischer Sozialräume (Farwick et al. 2012)

Hürden in der Evidenzgenerierung durch fragmentierte Datenhaltung

Hinsichtlich der Passgenauigkeit der verschiedenen möglichen Instrumente stellen sich somit einige Herausforderungen, für deren Begegnung eine vertiefte Datenlage notwendig wäre, die sich derzeit weder aus den Sozialmonitorings der Kommunen noch aus den verschiedenen weiteren Datenquellen versorgungsrelevanter Stakeholder ableiten allein lässt:

- So existieren auf Seiten der **Kommunen** zwar vereinzelt Daten über die gesundheitlichen Stati, die sich stadtteilbezogen mit weiteren in der *Sozialberichterstattung* statistisch erfassten Aspekten wie z.B. dem Sozialleistungsbezug, dem Migrationsanteil oder der Alleinerziehendenquote in Beziehung setzen lassen, doch stellen die kommunalen Berichterstattungen in der Regel trotz ihrer umfangreichen Indikatorenlisten eine wenig feingranulare Übersicht dar. So nimmt die *Gesundheitsberichterstattung* zumeist nur einen punktuellen Ausschnitt des Lebensverlaufs (wie z.B. die Schuleingangsuntersuchungen,

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Renteneintritt, Sterbefälle), oder nur ausgewählte Gesundheitsparameter – häufig aggregiert – in den Blick (Landeszentrum Gesundheit NRW 2022).

- Daten aus der **GKV-Abrechnung**, wie sie z.B. bei *Krankenkassen,* den *Kassenärztlichen Vereinigungen* oder *Abrechnungsdienstleistern für Pflege, Heil- und Hilfsmittel* vorliegen, können nützliche Informationen über die Inanspruchnahme von Gesundheitsleistungen liefern. Für sich alleine stehend können diese ihrerseits aber nicht den tatsächlichen Bedarf an Gesundheitsleistungen, sondern lediglich die Inanspruchnahme des Angebots unter den aktuellen Rahmenbedingungen abbilden. Zudem kommt es im Ruhrgebiet aufgrund der engen räumlichen Ballung zu Verflechtungen der Versorgung entlang von KV- und Krankenkassengrenzen.
- Über die **statistischen Landesämter** lassen sich sowohl Daten zu Themen wie Gesundheitsausgaben, Pflege oder gesundheitsrelevanten Verhalten abrufen als auch zu Bevölkerungserhebungen in Bezug auf Geburtenraten, Migration oder der Lebenserwartung. Jedoch stellen die Daten einen landesweiten Überblick dar und geben keine Aufschlüsse über regionale Besonderheiten oder Ballungsräume.
- Auch über **Longitudinalstudien und öffentlich zugängliche Datenbanken** wie dem *Sozio-ökonomischen Panel (SOEP)*, der *Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)*, der *Nationalen Kohorte (NAKO)* oder medizinischen Datenbanken wie dem *Landeskrebsregister NRW* lassen sich repräsentative Mikrodaten durch Wiederholungsbefragungen in Deutschland zu unterschiedlichen Schwerpunktthemen nutzbar machen und mit Gesundheitsfragestellungen verknüpfen.
- **Weitere Daten zu gesundheitsrelevanten Determinanten** liegen bei verschiedenen Institutionen vor. So verfügen die lokalen *Umweltämter und das Landesamt für Umwelt und Verbraucherschutz (LANUV)* über Daten zu Lärm- und Luftbelastung sowie Hitze, das *Landesinstitut für Arbeitsschutz und Arbeitsgestaltung* führt regelmäßig Beschäftigtenbefragungen zu Arbeit, Sicherheit und Gesundheit durch und der *Regionalverband Ruhr* verfügt mit seinem *ruhrFIS* über Daten unter anderem zur Daseinsvorsorge.

Diese Liste stellt lediglich beispielhaft das Potenzial von bereits in verschiedenen primären Kontexten erhobenen Datenbeständen bei unterschiedlichen Akteuren dar und könnte beliebig erweitert werden. Doch eine separierte Betrachtung dieser Datenbestände bietet keine ausreichende Wissensbasis für die Ausgestaltung bedarfsgerechter Angebote. Eine Kombination der aus den einzelnen Datenbeständen zu gewinnenden Informationen birgt dagegen ein nicht zu unterschätzendes Potenzial.

Bedarf eines generischen Lösungsansatzes für die Datenanalyse

In der Kombination verschiedener der oben angesprochenen Datenquellen kann somit ein Schlüssel für die Generierung von Evidenz für die passgenaue Ausgestaltung sozialräumlich orientierter gesundheitsbezogener Lösungen liegen. Doch genau hierin befindet sich gleichzeitig auch eine

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

große Hürde, da die Verfügbarmachung der Informationen aus Sekundärdaten aufwendig und datenschutzrechtlich komplex zu betrachten ist.

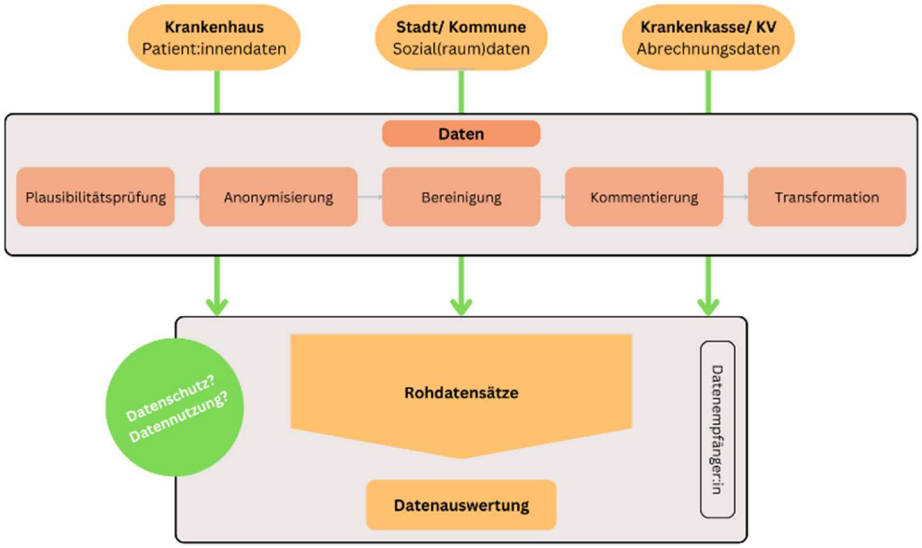

Abbildung 3: klassischer Weg der Datenlieferung in der Versorgungsforschung; exemplarisch (eigene Darstellung)

Die Aufbereitung von Datensätzen unterliegt bestimmten Regeln, zu denen u.a. die Kommentierung, die Anonymisierung, die Bereinigung sowie die Transformation der Daten zählen (Döring et al. 2016). Bei der (Weiter-)Verarbeitung von aus in anderen Kontexten erhobenen Daten – bei denen es sich also nicht um Primärdaten handelt – müssen diese Schritte häufig schon auf Seiten der Datenlieferanten erfolgen. Genau dieser oftmals arbeitsintensive Umstand stellt nicht selten schon allein aufgrund mangelnder personeller Ressourcen ein Hindernis für die tatsächliche Durchführung solcher Vorhaben dar. Hinzu kommt, dass die notwendigen Abstimmungsprozesse für jedes Vorhaben und jede Fragestellung immer wieder neu erarbeitet werden, was nicht nur dazu führt, dass es zu prozessualen Redundanzen kommt, die sich in ihren Aufwänden und Wartezeiten bis zur ersten tatsächlichen Datenlieferung potenzieren, sondern auch die Beantwortung von (Folge-)Fragestellungen nahezu unmöglich macht.

Um die Datenanalyse auch bei sensiblen, z.T. auch personenbezogenen Daten maximal datenschutzkonform zu gestalten, den wissenschaftlichen Anforderungen weiterhin Rechnung zu tragen, gleichzeitig aber auch den beschriebenen Problemlagen in der Umsetzung von Evaluationen zu begegnen, müssen entsprechende Lösungsansätze also folgender Maxime unterliegen:

→ Alle relevanten Daten müssen zeitnah verfügbar sein.
→ Sowohl der administrative als auch der organisatorische Aufwand für die Datenverfügbarmachung müssen gering gehalten werden.
→ Die Umsetzung von Folgefragestellungen oder auch Längsschnittstudien muss niederschwellig durchführbar sein.

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Das zu implementierende System setzt also optimalerweise auf einmal erlernten Analyselogiken auf und kann sie sowohl auf bereits bekannte Datensatzstrukturen wiederholt anwenden als auch auf neue Datensatzstrukturen übertragen und weiterentwickeln. Denkbar wären hierfür Verfahren, die sich algorithmenbasierte Auswertungsmechanismen zunutze machen und so die Grundlagen für eine niederschwellige Evidenzgenerierung für eine verbesserte Versorgungs- und Teilhabeplanung und Behandlungskoordination sowie die Verzahnung von Gesundheitsmanagement und Gesundheitspolitik ermöglichen (vgl. Konzept S. 26).

II. Zielsetzung des Vorhabens: Aufbau eines Sozialmedizinischen Datenverbunds in der Gesundheitsmetropole Ruhr

> *»Industriedaten sind Gold wert, wenn es darum geht, neue Produkte und Dienstleistungen zu entwickeln. Die Realität sieht aber noch so aus, dass 80 % der Industriedaten gesammelt, aber nie genutzt werden. Das ist reine Verschwendung. Eine echte Datenwirtschaft wäre hingegen ein kraftvoller Motor für Innovation und neue Arbeitsplätze. […] Dafür brauchen wir gemeinsame Datenräume - zum Beispiel im Energie- oder Gesundheitswesen. […] Deshalb werden wir im Rahmen von NextGenerationEU eine europäische Cloud aufbauen – auf der Grundlage von GAIA-X.«*
>
> *(Ursula von der Leyen, Rede zur Lage der Union vom 16. September 2020)*

Grundanlage des Sozialmedizinischen Datenverbunds Gesundheitsmetropole Ruhr

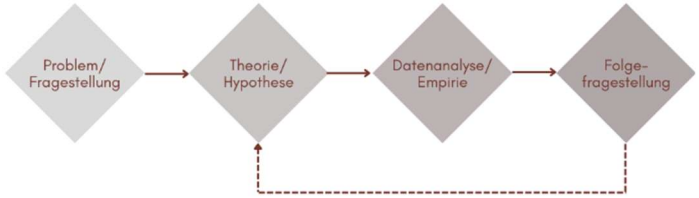

Abbildung 4: Co-Design in der Forschungsfragestellungentwicklung (eigene Darstellung nach (Büschges 1998))

Das Vorhaben versucht, mittels eines algorithmenbasierten Ansatzes, in dem dezentral und kollaborativ gearbeitet wird, standardisierte Lösungen für eine niederschwellige Sekundärdatenanalyse innerhalb des Verbundes zu etablieren. Einmal erprobt können so im Co-Design viele weitere Fragestellungen in einer komplexen sozialräumlich orientierten Versorgung kurzfristig bearbeitet und deren Ergebnisse für Entscheidungsträger nutzbar gemacht werden. Relevante Akteursgruppen aus Gesundheitsversorgung, Digital Health Industrie, Gesundheits- und Sozialplanung, Sozialversicherung und Leistungserbringung, Wissenschaft und Praxis finden sich hierfür unter dem Dach *Data4Health.Ruhr* zusammen, um gemeinsam

- relevante Fragestellungen für die Entwicklung von bedarfsgerechten Lösungsansätzen für eine sozialräumlich passgenau ausgerichtete Gesundheitsversorgung zu entwickeln,
- digitale Verfahren, Instrumente und Lösungen im Sinne von Dienstleistungen und Produkten zu konzipieren,

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

- priorisierte Fragestellungen mittels der Informationen in den verschiedensten, bereits vorhandenen Datenbanken zu beantworten und
- die daraus gewonnenen Erkenntnisse in die Praxis einfließen zu lassen.

Data4Health.Ruhr greift hierfür nicht nur auf Erfahrungen und Methoden in der algorithmenunterstützten Datenauswertung und -aufbereitung zurück (vgl. *Teilvorhaben-beschreibung „Digital Health Factory Ruhr Starterprojekt Methoden-Entwicklung: Datenwerkstatt“*), sondern bietet gleichzeitig einen vertrags- und datenschutzrechtlichen Rahmen für die Zusammenarbeit der Verbundmitglieder entlang der spezifischen Fragestellungen (= *use cases, s.u.*). Damit entsteht ein nach und nach wachsender Verbund, der anhand konkreter, relevanter Fragestellungen nicht nur Antworten für eine bessere Planung und Versorgung liefert, sondern darüber hinaus auch im Sinne des *DHFR*-Bildes ein Entwicklungsraum für marktfähige digitale Produkte und Dienstleistungen der beteiligten Unternehmen wird.

Anwendungsfelder des Sozialmedizinischen Datenverbunds Gesundheitsmetropole Ruhr

Im Rahmen des Bündnis- und Innovationsmanagements der *Digital Health Factory Ruhr* wurden im ersten Projektquartal Arbeitskreise eingerichtet, die als operative Einheiten der jeweiligen Entwicklungsfelder fungieren und allen interessierten Bündnispartner:innen zur Mitwirkung offenstehen. Das Vorhaben *Data4Health.Ruhr* ist das Ergebnis der Innovationsentwicklungen des Arbeitskreises „Sozialmedizinischer Datenverbund“ im gleichnamigen Entwicklungsfeld. Die Bedarfsbeschreibungen, auf die mit Hilfe von *Data4Health.Ruhr* reagiert werden soll, speisen sich aus den real existierenden Problemlagen hinsichtlich einer sozialraumorientierten Gesundheitsversorgung, die über die Bündnispartner:innen in den Arbeitskreis eingebracht werden.

Abbildung 5: use case Entwicklung in *Data4Health.Ruhr*

Aus eben diesen Problemlagen heraus sollen erste use cases entwickelt werden, die explizite Fragestellungen in den Blick nehmen, mit deren Beantwortung ihrerseits wiederum die Grundlage für die Schaffung von sozialmedizinischen Versorgungsangeboten darstellen. Gleichzeitig stellen diese use cases den anlassbezogenen Rahmen für die Erprobung der dezentral agierenden Algorithmen und des kollaborativ ausgerichteten Lösungsansatzes dar. Erste potenzielle use cases wurden durch die Akteure des

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Organisation von Rettungsdiensten: Eine stadtübergreifende Organisation von Rettungsdiensten stellt sich gerade in polyzentrischen Ballungsräumen wie dem Ruhrgebiet als schwierig dar. Besonders an den städtischen Randgebieten wäre oftmals eine kooperative Planung zielführend, doch fehlen Kenntnisse über mögliche Optimierungsbedarfe in kommunal übergreifender Perspektive. Aufgrund von unterschiedlichen Dokumentationsformen in den Kommunen ist die Vergleichbarkeit derzeit nur bedingt gegeben.

Geotagging bei COVID-19 und Influenza: Mithilfe von Sozialraumdaten der Stadt Dortmund und eigenen Daten stationärer Patienten hat das Klinikum Dortmund bereits im Jahr 2021 epidemiologische Verteilungsmuster im Verlauf der ersten 3 COVID-19 Wellen betrachtet. Möglich war so nicht nur eine örtliche, sondern auch eine zeitliche Zuordnung dieser Daten, um Verläufe und Abstände von potenziellen Infektionsquellen besser zu erkennen. In der Weiterführung dieses Ansatzes wird nun im Rahmen von *Data4Health.Ruhr* angestrebt, die Betrachtung sowohl über die Stadtgrenzen hinaus auszuweiten und vergleichbar zu machen, als auch - aufgrund der Entwicklung hin zu weniger schwerwiegenden Verläufen - auf die ambulante Gesundheitsversorgung sowie die Rettungsdienste (s.o.) auszuweiten. Die Ausweitung auf weitere Indikationen auch über Influenza hinaus ist schrittweise angestrebt.

Berufliche Eingliederung von Arbeitslosen nach Aufenthalt in psychiatrischer Klinik oder Rehabilitation: Sowohl in Rehabilitationskliniken als auch in psychiatrischen Kliniken werden im Rahmen von stationären oder teilstationären Aufenthalten regelhaft umfangreiche Diagnostiken hinsichtlich der Arbeitsfähigkeit der Patient:innen nach Entlassung und ggf. bestehenden Unterstützungsbedarfen erstellt. Vereinzelte Rückmeldungen lassen aber vermuten, dass diese Informationen selten von den betreuenden JobCentern aufgegriffen werden; die Gründe dafür sind weitgehend unbekannt. So ist beispielsweise nicht klar, ob bzw. wie gut die Kontaktaufnahme zwischen den Patient:innen und den JobCentern gelingt und welche Angebote zielführend wären, um die Defizite zu minimieren.

Reha vor Rente: Ein relevanter Anteil von Menschen im Erwerbsminderungsrentenbezug hat zuvor trotz der gesellschaftlichen Maxime „Reha vor Rente“ keine Rehabilitationsleistung in Anspruch genommen. Doch auch wenn den Betroffenen zuvor eine Rehabilitationsmaßnahme zuteilwurde, besteht häufig kein Bezug zwischen Rehabilitations- und Berentungsgrund. Auch hier stellt sich die Frage nach den Ursachen für diese Diskrepanz. Dabei sollte zunächst der Frage nachgegangen werden, ob dieser Umstand auf bestimmte Zielgruppen (z.B. hinsichtlich des sozialökonomischen Stauts) in besonderer Weise zutrifft und wie etwaige Unterschiede in diesen Gruppen definiert sind. Im Ergebnis können dann Lösungen für eine verbesserte Erreichbarkeit von Rehabilitationsangeboten konzipiert werden. (vgl. S. 27 *Konzept zur Digital Health Factory Ruhr – Digitale Lösungen für die Gesundheitsversorgung in Ballungsräumen*)

Bildungssettings & Gesundheit: Die unterschiedlichen Lebenswelten wie Kitas und Schulen nehmen Einfluss auf die Entwicklung und Gesundheit von Kindern und Jugendlichen (Fegeler et al. 2020). Daher sollte der Fokus auf die Schnittstelle zwischen Prävention, Gesundheitsförderung sowie Gesundheitsversorgung und dem Bildungsumfeld in seiner Eigenschaft als Setting gelegt werden. Gleichzeitig sind diese Institutionen in der Lage, früh gesundheitsrelevante

GEFÖRDERT VOM
wir! Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Auffälligkeiten von Kindern und Jugendlichen zu erkennen, Interventionsmöglichkeiten aufzuzeigen und zu vermitteln.

Gemeinsam ist diesen use cases die Notwendigkeit der Verknüpfung von epidemiologischen, medizinischen und abrechnungstechnischen Versorgungsdaten sowie sozialräumlichen und Stadtentwicklungsdaten.

III. Das Potenzial von Region und DHFR-Bündnis für den Verbundansatz

Die Gesundheitsmetropole Ruhr hat nicht nur einen hohen Bedarf an der hier formulierten Perspektive eines Sozialmedizinischen Datenverbunds. Sie stellt auch einen besonders geeigneten Anwendungsraum dar, insofern die polyzentrisch verdichtete Siedlungs- und die soziodemografisch vielschichtig differenzierte Bevölkerungsstruktur quantitativ wie qualitativ eine maximale Generierung von überregional verwendbaren Erkenntnissen und Nutzungsperspektiven erlaubt.

Vor allem aber verfügt sie auch über ein ausgeprägtes Potenzial, eine derartige Perspektive nachhaltig auf der Basis eines föderierten Ansatzes verfolgen zu können. Hierfür steht schon jetzt das – weiter wachsende – WIR!-Bündnis der *Digital Health Factory Ruhr* in seiner gesamten Bandbreite:

- Das *Digital Health Cluster* an der Ruhr verfügt insbesondere im Bereich des Datenmanagements über herausragende Stärken, für die industrielle Unternehmen mit marktführenden Positionen (wie die im Bündnis schon vertretenen opta data Gruppe, CompuGroup Medical, VISUS Health IT GmbH, KAIROS GmbH und G DATA Cyberdefense AG, weiterhin die adesso SE und die Bitmarck GmbH) ebenso stehen wie das Fraunhofer-Institut für Software- und Systemtechnik (ISST), das führend ist in der Erforschung und Entwicklung von Datenökosystemen (s. z.B. International Data Space (Otto et al. 2016) oder (Gaia-X: A Federated Secure Data Infrastructure 2022).
- Fließend sind hier die Übergänge zu den vorhandenen *Akteuren in der Medizin- und Gesundheitsinformatik* wie an der Universitätsmedizin Essen (insbes. mit dem Institut für Medizinische Informatik, Biometrie und Epidemiologie und dem Institut für Künstliche Intelligenz in der Medizin) sowie der Ruhr-Universität Bochum (speziell in der Datensicherheit), der Universität Witten/Herdecke und der Fachhochschule Dortmund. Besonders herauszuheben ist das Landeskrebsregister NRW als größtem medizinischen Register in Deutschland mit Dienstleistungen, die auch zunehmend jenseits der Onkologie zum Einsatz kommen sollen.
- In der *sozialräumlichen und Regionalentwicklung* formiert sich im WIR!-Bündnis der *Digital Health Factory Ruhr* eine für *Data4Health.Ruhr* und sein Datenökosystem hochrelevante Allianz, die vom Institut für Landes- und Stadtentwicklungsforschung gGmbH (ILS) an der TU Dortmund und dem Institut Arbeit und Technik (IAT) der Westfälischen Hochschule über das Community Health Department der Hochschule für Gesundheit und das Institut für Urban Public Health (InUPH) an der Universität Duisburg-Essen bis hin zum Regionalverband Ruhr reicht.
- Der *Regionalverband Ruhr* vertritt im DHFR-Bündnis zugleich seine Mitglieder, d.h. die Kreise und kreisfreien Städte des Ruhrgebietes und bietet so die Möglichkeit zu einer gemeinsamen

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Definition von datenrelevanten Problemlagen und Erkenntnisbedarfen und darauf aufsetzenden Nutzungsperspektiven im kommunalen Gesundheits- und Sozialwesen. Mit einer Vielzahl von Städten (insbesondere den Sozialdezernaten und Gesundheitsämtern) sind darüber hinaus auch bereits direkte Kontakte zum DHFR-Bündnis hergestellt worden (u.a. Essen, Bochum, Dortmund und Gelsenkirchen).

- Über die Kommunen hinaus sind die *Sozialversicherungsträger sowie die Kassenärztlichen Vereinigungen* von besonderer Bedeutung für die Überprüfung und Erschließung des erforderlichen Datenportfolios. Hierfür stehen die schon jetzt im Bündnis vertretenen Einrichtungen (z.Zt. Knappschaft Bahn-See, die Techniker Krankenkasse, die DAK Gesundheit, die BIG direkt gesund und die DRV Rheinland) wie auch die Beziehungen, die über den MedEcon-Verbund und die Bündnisgremien bestehen (z.Zt. schon zu den Allgemeinen Ortskrankenkassen in NRW und der KV Westfalen-Lippe).

Über diese „Stakeholder“ hinaus beinhaltet das DHFR-Bündnis vor allem auch ein breites Spektrum von Einrichtungen, die den Sozialmedizinischen Datenverbund Gesundheitsmetropole Ruhr für jeweils konkrete Zwecke nutzen und mit jeder Nutzung seine Leistungsfähigkeit und seinen Wert erhöhen – ein Spektrum, das von Krankenhäusern bis zu Stadtteilinitiativen, von kommunalen Planungseinrichtungen bis zu innovativen Start-ups reicht. Dieser Zusammenschluss ist explizit offen für die Beteiligung weiterer interessierter Akteure angelegt, die sich sowohl im Bündnis als auch im Entwicklungsschwerpunkt des „Sozialmedizinischen Datenverbunds“ engagieren möchten.

IV. NATIONALE UND INTERNATIONALE INITIATIVEN ALS BASIS DES VERBUNDANSATZES

Initiativen zur Überwindung von Grenzen in der Datenverfügbarkeit

In der Gesundheitsversorgungslandschaft bestehen Wertschöpfungsketten aus verschiedenen Akteuren, Organisationen und Datenbeständen, die das gemeinsame Interesse eint, sich entlang einer Zielsetzung über ihre sich gegenseitig ergänzenden Tätigkeitsbereiche hinweg zu orchestrieren. An verschiedenen Stellen werden hierzu dezentrale Ansätze und Werkzeuge diskutiert, die zur nachhaltigen Einführung und zum Betrieb eines sicheren und vertrauenswürdigen Ökosystems beitragen sollen. Zu nennen ist hier die deutsch-französische Gaia-X Initiative, mit deren Instrumenten (Specifications, Open-Source Software Codes, Labels) die Voraussetzungen für sichere digitale, föderierte Ökosysteme geschaffen werden sollen (Gaia-X: A Federated Secure Data Infrastructure 2022). Weiterhin wurde mit dem European Health Data Space eine Verordnung auf den Weg gebracht, die eine solche Infrastruktur für die Mitgliedsländer der EU gesetzgeberisch verpflichtend machen soll (Europäische Kommission 2022). Auch das Konzeptpapier der Telematikinfrastruktur 2.0 setzt auf föderierte und dezentrale Werkzeuge und Mechanismen (Telematikinfrastruktur | gematik 2022).

Die genannten Initiativen sollen über den Aufbau dezentraler Infrastrukturen organisationsübergreifende Datennutzung entlang europäischer Werte ermöglichen. Um die Service-Ebene möglichst früh erlebbar zu machen, können schon heute dezentrale

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Datenverarbeitungsverfahren genutzt werden. Eingesetzt werden dabei Lernalgorithmen, durch die die Entwicklung datengestützter Intelligenz sowie die dezentrale Verfügbarkeit und Verfügbarmachung von Auswertungswerkzeugen dargestellt wird. Gleichzeitig werden die technische und organisatorische Sicherstellung der Datenschutzinteressen im Sinne der Europäischen Datenschutzgrundverordnung (DSGVO) ermöglicht. Um die Integration beliebiger Daten von beliebigen Dateneigentümern zu erleichtern und gleichzeitig Datenschutzinteressen anzuerkennen, verfolgen Federated-, Fog- und Swarm Learning einen dezentralen Ansatz für maschinelles Lernen, die Edge Computing, Peer-to-Peer-Netzwerke und Koordination miteinander verbindet und dabei die Vertraulichkeit wahrt (Li et al. 2020) (Hosseinalipour et al. 2020) (Warnat-Herresthal et al. 2021).

Der Ansatz des „Federated Learnings“ als Instrument des Sozialmedizinischen Datenverbunds Gesundheitsmetropole Ruhr

> *However, there is an increasing divide between what is technically possible and what is allowed, because of privacy legislation. Here, to facilitate the integration of any medical data from any data owner worldwide without violating privacy laws, we introduce Swarm Learning-a decentralized machine-learning approach that unites edge computing, blockchain-based peer-to-peer networking and coordination while maintaining confidentiality without the need for a central coordinator, thereby going beyond federated learning.*
>
> *(Warnat-Herresthal et al. 2021)*

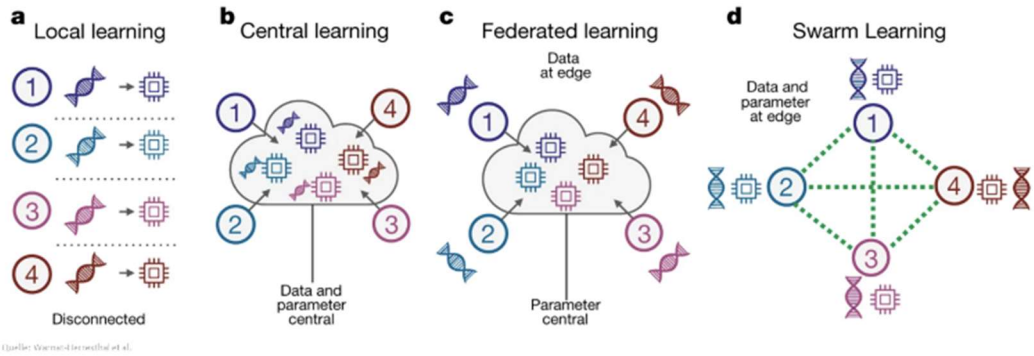

Abbildung 6: Concept of Swarm Learning (Warnat-Herresthal et al. 2021)

Dezentrales Lernen (a) bewahrt große (und private) Datensätze lokal beim Dateneigentümer auf; (b) erfordert keinen Austausch von Rohdaten, wodurch der Datenverkehr reduziert und die Effizienz erhöht wird; (c) bietet ein hohes Maß an Datensicherheit; (d) garantiert eine sichere, transparente und faire Aufnahme dezentraler Mitglieder in das Netz (Warnat-Herresthal et al. 2021).

Data4Health.Ruhr strebt einen dezentralen Ansatz über föderierte Schnittstellen an. Über diese Schnittstellen sollen Dienste (Softwaredienste und Datenangebote) auffindbar, die Authentizität und Konformität dieser Dienste dokumentiert und überprüfbar sowie die tatsächliche Verfügbarkeit von Daten und Anwendung von Softwarediensten (Algorithmen) realisiert werden. Als Vorbereitung für die Analyse der dezentral vorliegenden Daten werden geeignete Algorithmen für die Anwendung

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

maschinellen Lernens identifiziert. Hierfür muss die Struktur und das Format der später verwendeten Daten beschrieben und z.B. durch Dummy-Daten plausibel gemacht werden. Der ausgewählte und konfigurierte Algorithmus wird ggf. unter Ergänzung einer Vorverarbeitungspipeline in einer Anwendung verpackt, die bei den beteiligten Organisationen installiert, konfiguriert und angewendet wird. Somit ist es nicht notwendig, sensible Daten außerhalb der Einrichtungen zu verarbeiten.

V. VORGEHENSWEISE & UMSETZUNGSPLAN

Es sind verschiedene Etappen hin zum *Sozialmedizinischen Datenverbund Gesundheitsmetropole Ruhr* geplant, die jeweils in eigenen formalen Projektanträgen realisiert werden sollen. In diesem Sinne bildet die Projektfamilie als Gesamtes die operative Umsetzung des in dieser Rahmenskizze dargestellten Vorhabens.

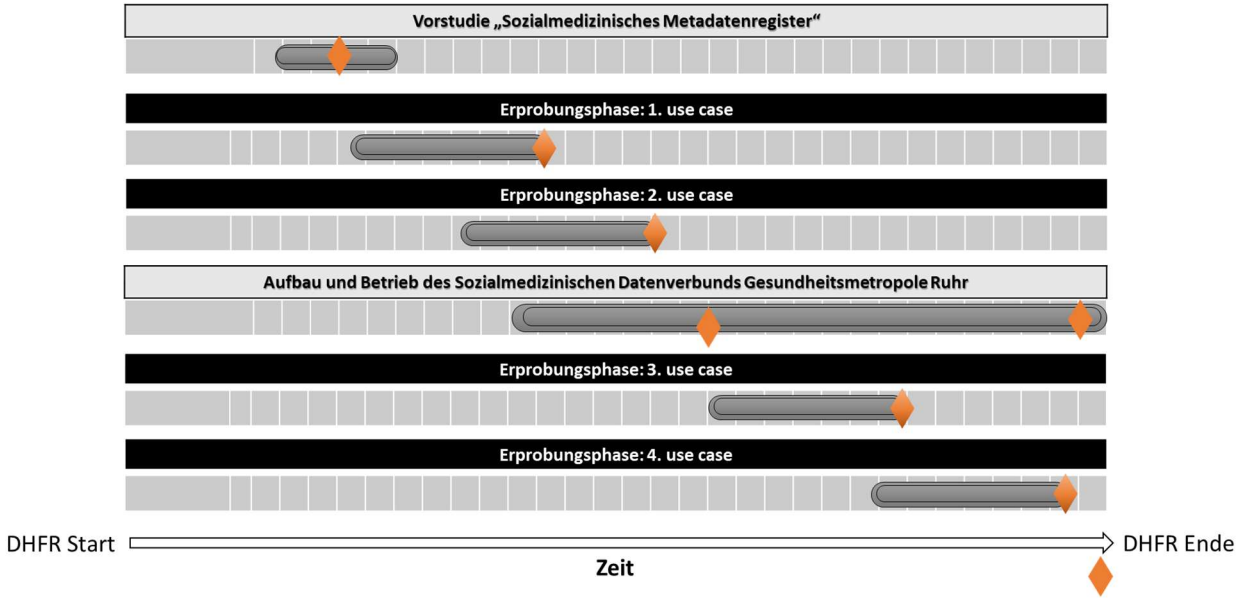

Abbildung 7: : Darstellung des Arbeitsplans in der Projektfamilie *Data4Health.Ruhr*

Vorstudie „Sozialmedizinisches Metadatenregister“: Da bei den einzelnen Bündnispartnern, die sich im Arbeitskreis engagieren, je nach beruflichem und einrichtungsbezogenem Kontext ihres Wirkungskreises unterschiedlich tiefe Kenntnisse über Umfang und Inhalte der verschiedenen relevanten Datenbestände bestehen, soll in einem ersten Schritt eine Zusammenstellung der entsprechenden Informationen erfolgen. Durch eine systematische Aufbereitung und Darstellung wird so ein umfassender Überblick über die aktuelle Situation in der derzeit fragmentiert gehaltenen Daten- und Informationslandschaft gegeben, von der alle Mitglieder im DHFR-Bündnis dauerhaft profitieren können. Im Ergebnis steht somit eine Metadatenbank, die gleichzeitig als Fundierung für die ersten im Rahmen der Projektfamilie umzusetzenden use cases dient.

Erprobungsphase: Use cases 1 & 2: Als Grundlage für die use case Entwicklung wurden bereits im Rahmen der Arbeitskreisarbeit verschiedene Fragestellungen erörtert und diskutiert, von deren

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Beantwortung aus Sicht der beteiligten Akteure relevante Erkenntnisse für die Entwicklung der eingangs beschriebenen Lösungsansätze zu erwarten sind (s.o.). Da die Umsetzung der in den use cases zu beschreibenden Evaluationen in *Data4Health.Ruhr* mittels des *Federated Learning Ansatzes* erfolgen wird, werden in einer ersten Erprobungsphase zunächst use cases adressiert, die in ihren Anforderungen an die wissenschaftliche Auswertung einen beherrschbaren Komplexitätsgrad aufweisen. Die Entwicklung der ersten use cases erfolgt iterativ sowie in enger Einbindung der (Zwischen-)Erkenntnisse aus der Vorstudie und stellt in ihren Ergebnissen die Voraussetzung für die Schaffung des Datenverbunds *Data4Health* dar.

Konzipierung und Entwicklung des Sozialmedizinischen Datenverbunds Gesundheitsmetropole Ruhr: Auf Basis der vorangegangenen Entwicklungsarbeiten wird nun konsequenterweise der Aufbau des Datenverbundes *Data4Health.Ruhr* prototypisch umgesetzt. Hierzu zählen die Erweiterung und Weiterentwicklung des Metadatenregisters, bei der maßgeblich öffentlich zugängliche Quellen und Forschungsdatenbanken auf Landes und Bundesebene in den Blick genommen werden, und die prototypische Entwicklung eines dem Geschäftsmodell zugrundeliegenden Dienstleistungs-portfolios.

Erprobungsphase: Use cases 3 & 4: Weitere use cases werden im Verlauf des Projektes durch den Arbeitskreis identifiziert und im Rahmen des Innovationsmanagements zu eigenständigen formalen Projekten ausdifferenziert. Diese use cases werden nun innerhalb des formalstrukturellen Rahmens des Datenverbundes umgesetzt, der somit operationell erprobt werden kann. In Weiterentwicklung zu den use cases der ersten Erprobungsphase unterscheiden sich die use cases der zweiten Erprobungsphase hinsichtlich ihrer Komplexität sowie der einzubeziehenden Datenbestände und Datenlieferanten.

Anmerkung: Grundsätzlich ist die Anzahl der use cases im Vorfeld nicht vorgeben oder limitiert. Da jedoch zum Zeitpunkt der Rahmenskizzenerstellung davon ausgegangen wird, dass vier use cases eine realistische Größe hinsichtlich Umsetzbarkeit und notwendigen Erkenntnisgewinns darstellen, wurde der Übersichtlichkeit halber mit dieser Anzahl in der Darstellung gearbeitet.

VI. NACHHALTIGKEIT & VERWERTUNGSPERSPEKTIVE

Hinsichtlich der Verwertungsperspektiven von *Data4Health.Ruhr* ist zu unterscheiden zwischen den Verwertungsperspektiven des Verbundes selbst und denjenigen der verschiedenen Teilnehmer- bzw. Nutzergruppen, die miteinander in Abhängigkeit stehen:

1. Die durch *Data4Health.Ruhr* ermöglichten konkreten Nutzungsoptionen (= use cases) in der Gestaltung der sozialräumlichen Gesundheitsförderung und -versorgung von Ballungsräumen begründen die Verwertungsperspektiven der den Verbund nutzenden Einrichtungen. Dies gilt sowohl für die Einrichtungen und Unternehmen am Ende der Wertschöpfungskette bzw. des Gesundheits- und Sozialwesens wie auch für die zuliefernden Gewerbe des Digital Health Clusters.

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

2. Die informationstechnischen, industriellen und wissenschaftlichen Partner erhalten mit dem Verbund und dem ihm zugrunde liegenden föderierten Ansatz zudem einen Erprobungs- und Anwendungsraum, in dem sie ihre Algorithmen lernen lassen und weiterentwickeln können – und dies auch in Kommunikation mit den Verbundpartnern aus dem Gesundheits- und Sozialwesen. Im Kontext von NextGenerationEU (von der Leyen 2020) haben die Verbundpartner so einen entscheidenden Einstiegsvorteil (= First Mover).

3. Zwischen den Nutzungsperspektiven in der Gestaltung der sozialräumlichen Gesundheitsförderung und -versorgung von Ballungsräumen und den Entwicklungsperspektiven in Bezug auf die Algorithmen und die damit verbundene Qualität der föderierten Datenverarbeitung besteht eine Wechselwirkung, deren Dynamik für die Nachhaltigkeit des Gesamtvorhabens von entscheidender Bedeutung ist:

   - Die Systematik und Vorteile des *Federated Learnings* werden für die Nutzer erfahrbar gemacht. Allein diese Möglichkeit, die Strukturen und Arbeitsweisen des Verbundes mitzugestalten sowie föderierte Datengewinnungs- und Auswertungsverfahren zu erlernen wird für viele Bündnispartner eine attraktive Perspektive darstellen.

   - Immer einfacher und schneller durchzuführende Auswertungen mit kooperierenden Partnereinrichtungen werden verbesserte Angebotsallokation, effizienteren Mitteleinsatz und optimierte Planungsprozesse an der Ruhr ermöglichen. *Data4Health.Ruhr* bietet hierfür einen geschützten und vertrauensvollen Kooperationsraum.

   - Damit eröffnet sich die Perspektive eines iterativen Aufbaus und einer „kritischen Masse“. Mit jedem erfolgreich eingebrachten Verbundbaustein – handele es sich um eine relevante Datenressource oder um einen datenbasierten use case - nimmt der Nutzen des Verbundes selbstverstärkend und nachhaltig zu.

4. Eine derartige Dynamik geht mit einer kollaborativen Verwertungsperspektive und einem Geschäftsmodell für den Datenverbund selbst einher (vgl. Abb. 7), das über den Förderhorizont der *Digital Health Factory Ruhr* hinausreicht und damit auch als eine maßgebliche Konkretisierung der im DHFR-Konzept niedergelegten Strategieentwicklung (siehe dort S.16/17) anzusehen ist. Mögliche Elemente dieser Verwertungsperspektive sind:

   - die Bildung einer den Datenverbund langfristig tragenden und betreibenden Einrichtung – dies voraussichtlich in Verbindung mit der *Digital Health Factory Ruhr* (und möglicherweise in einer genossenschaftlichen Form),

   - ein Beitrags- und Gebührenmodell für die Beteiligung an dem Datenverbund und die – anlass- und konditionsabhängige – Zurverfügungstellung und Nutzung von Datenbeständen,

   - ein Vergütungsmodell für die Unterstützung und Betreuung insbesondere anspruchsvoller use cases durch den Datenverbund bzw. die ihn betreibenden Einrichtung (dies im Abgleich mit den Werkstätten der DHFR),

GEFÖRDERT VOM
wir! Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

- die wirtschaftliche Beteiligung des Datenverbundes bzw. der ihn betreibenden Einrichtung an der Neu- oder Weiterentwicklung datenverarbeitender Dienstleistungen und Produkte in Kooperation mit einzelnen Verbundpartnern.

Mittelfristig und im Projektverlauf soll der *Sozialmedizinische Datenverbund Gesundheitsmetropole Ruhr* im Referenzraum Ruhrgebiet erprobt und eingeführt werden. Langfristig soll sich der Datenverbund für weitere Nutzungsszenarien, Partner und Initiativen in deutschland- bzw. EU-weiter Perspektive öffnen.

GEFÖRDERT VOM
Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung

Literaturverzeichnis

Döring, Nicola; Bortz, Jürgen; Pöschl, Sandra; Werner, Christina S.; Schermelleh-Engel, Karin; Gerhard, Carla; Gäde, Jana C. (2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Aufl. 2016. Berlin, Heidelberg: Springer Berlin Heidelberg (Springer-Lehrbuch). Online verfügbar unter http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1624548.

Europäische Kommission (2022): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über den europäischen Raum für Gesundheitsdaten. Online verfügbar unter https://eur-lex.europa.eu/resource.html?uri=cellar:dbfd8974-cb79-11ec-b6f4-01aa75ed71a1.0003.02/DOC_1&format=PDF, zuletzt geprüft am 20.12.2022.

Fegeler, Ulrich; Jäger-Roman, Elke; Rodens, Klaus (Hg.) (2020): Praxishandbuch der pädiatrischen Grundversorgung. 2. Auflage. München: Elsevier. Online verfügbar unter http://shop.elsevier.de/978-3-437-21282-6.

Gaia-X: A Federated Secure Data Infrastructure (2022). Online verfügbar unter https://gaia-x.eu/, zuletzt aktualisiert am 14.12.2022, zuletzt geprüft am 14.12.2022.

Hosseinalipour, Seyyedali; Brinton, Christopher G.; Aggarwal, Vaneet; Dai, Huaiyu; Chiang, Mung (2020): From Federated to Fog Learning: Distributed Machine Learning over Heterogeneous Wireless Networks. In: *IEEE Commun. Mag.* 58 (12), S. 41–47. DOI: 10.1109/MCOM.001.2000410.

Landeszentrum Gesundheit NRW (2022): Aufgaben und Ziele der Gesundheitsberichterstattung in NRW. Online verfügbar unter https://www.lzg.nrw.de/ges_bericht/aufgaben/index.html, zuletzt aktualisiert am 14.12.2022, zuletzt geprüft am 14.12.2022.

Leyen, Ursula von der (2020): Rede zur Lage der Union.

Li, Tian; Sahu, Anit Kumar; Talwalkar, Ameet; Smith, Virginia (2020): Federated Learning: Challenges, Methods, and Future Directions. In: *IEEE Signal Process. Mag.* 37 (3), S. 50–60. DOI: 10.1109/MSP.2020.2975749.

Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (2020): Sozialbericht NRW 2020. Armuts- und Reichtumsbericht - Kurzfassung -. Online verfügbar unter https://broschuerenservice.mags.nrw/files/d/d/dd4318bdf0950f251fff65d23c7d3ef3.pdf, zuletzt geprüft am 20.12.2022.

Otto, B.; Jürjens, J.; Schon, J.; Auer, S.; Menz, N.; Wenzel, S.; Cirulliles, J. (2016): Industrial Data Space. Digitale Souveränität über Daten (White Paper).

Telematikinfrastruktur | gematik (2022). Online verfügbar unter https://www.gematik.de/telematikinfrastruktur, zuletzt aktualisiert am 20.12.2022, zuletzt geprüft am 20.12.2022.

Warnat-Herresthal, Stefanie; Schultze, Hartmut; Shastry, Krishnaprasad Lingadahalli; Manamohan, Sathyanarayanan; Mukherjee, Saikat; Garg, Vishesh et al. (2021): Swarm Learning for decentralized and confidential clinical machine learning. In: *Nature* 594 (7862), S. 265–270. DOI: 10.1038/s41586-021-03583-3.

GEFÖRDERT VOM
wir! Wandel durch Innovation in der Region
Bundesministerium für Bildung und Forschung