

**Daten für eine bessere Versorgung nutzen**  
Machbarkeitsstudie Virtuelles Netzwerk Gesundheitsdaten  
August 2018

*Diese Machbarkeitsstudie wurde im Rahmen eines Kooperationsprojektes von  
Digital Society Institute der ESMT Berlin,  
Dr. med. Ulrike Helbig, Berlin,  
Fraunhofer FOKUS, Berlin,  
Nils Waldeck, LL.M., Berlin,*

*basierend auf dem von Dr. Ulrike Helbig entwickelten Konzept „Netzwerk Gesundheitsdaten“ erstellt.  
Sie wurde durch eine Zuwendung des Bundesministeriums für Gesundheit gefördert.*

*Autorinnen und Autoren: Martin Schallbruch (Projektleitung), Dr. Ulrike Helbig (Stellvertretende  
Projektleitung), Dr. Jörg Caumanns, Marko Harasic, Prof. Dr. Adrian Paschke, Alexandru Todor,  
Nils Waldeck.*

*August 2018*

## Zusammenfassung

Die Studie entwickelt das Modell eines „Virtuelles Netzwerks Gesundheitsdaten“ (NGD), in die Daten zu Zwecken der Forschung im Rahmen eines virtuellen Big-Data-Verbundes für Akteure im Gesundheitswesen, fallweise und für einzelne bestimmte Auswertungen vorübergehend zusammengeführt werden. Das Netzwerk Gesundheitsdaten steht allen Akteuren des Gesundheitswesens zur Nutzung offen, Leistungserbringern ebenso wie gesetzlichen und privaten Krankenversicherungen, öffentlichen und privaten Forschungseinrichtungen, Herstellern oder auch Fachbehörden des Bundes, der Länder und der Kommunen. Gerade mit Blick auf die bestehenden Probleme bei der Überführung von Ergebnissen aus der Forschung in die Versorgung ist ein derart offener Ansatz erstrebenswert. Die Datennutzer und Datenbereitsteller erhalten eine Rückkopplung zu den Ergebnissen von Analysen aus ihren Daten und so ist eine unmittelbare Anwendung dieser Ergebnisse sichergestellt.

Das Netzwerk Gesundheitsdaten wird als unter öffentlich-rechtlicher Verantwortung stehende Einrichtung modelliert, die aus einer Kontaktplattform zur Initiierung und Vorbereitung von Auswertungen, einem Bereich „Contracting und Datenschutz“ zur rechtlichen Ausgestaltung und datenschutzrechtlichen Absicherung der Auswertungen sowie einer Neutralen Stelle besteht, in der unter besonderer Abschottung und höchsten Sicherheitsvorkehrungen die Auswertungen durchgeführt, Ergebnisse übermittelt und anschließend die Daten gelöscht werden.

Die Machbarkeitsstudie untersucht die Anforderungen an ein solches Netzwerk Gesundheitsdaten in medizinfachlicher, rechtlicher und technischer Sicht. An zwei Beispielen Institutionen-übergreifender Auswertungen wird die praktische Durchführung analysiert. Aus den ermittelten Anforderungen an die Umsetzung wird ein Umsetzungsvorschlag entwickelt, der die rechtliche Konstruktion, Grundlagen der Organisation und die technische Architektur eines Netzwerks Gesundheitsdaten beschreibt. Eingeflossen sind Ergebnisse aus einem Stakeholder-Dialog mit relevanten Akteuren des Gesundheitswesens.

Die Machbarkeitsstudie kommt zu folgenden Ergebnissen:

- Das Modell ist für alle relevanten Beteiligten des Gesundheitswesens für die Lösung einzelner forschungs- und versorgungsrelevanter Fragestellungen nutzbar. Die meisten dieser relevanten Beteiligten verfügen bereits über die inhaltlichen und strukturellen Grundvoraussetzungen, an diesem Prozess teilzunehmen.
- Im Rahmen der Nutzung des virtuellen Netzwerks Gesundheitsdaten ist eine Qualitätsentwicklung in den Bereichen Datenerhebung, Datenarchivierung, Datenzusammenführung und Datenanalyse für die beteiligten Netzwerknutzer möglich.
- Voraussetzung für den Erfolg ist eine hohe Transparenz und niedrige Beteiligungsschwelle zur Schaffung eines Anreizes für die Datenerhebenden und Datensammelnden zur Beteiligung.
- Ein Großteil der Tätigkeit des NGD kann datenschutzrechtlich grundsätzlich auf den Erlaubnistatbestand von Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F. gestützt werden. Wegen des umfangreichen nationalen Fachrechts im Gesundheitswesen ist eine nationale gesetzgeberische Absicherung zu empfehlen.
- Eine technische Architektur des NGD kann mit vorhandenen Standard-Systemen (vor allem Hadoop) realisiert werden. In einer solchen Architektur ist ein BSI-Empfehlungen entsprechendes IT-Sicherheitskonzept umsetzbar, das höchsten Anforderungen genügt.
- Durch Festlegung eines Code-Referenzsystems (basierend auf SNOMED CT) können die hohen Anforderungen von NGD an semantische Interoperabilität umgesetzt werden.
- Zur Gewährleistung einer im öffentlichen Interesse organisierten Steuerung der Datenauswertung und zur Erleichterung einer Beteiligung öffentlicher Einrichtungen stehen mit der AöR und einer gGmbH in staatlicher Beteiligung geeignete Rechtsformen zur Verfügung.

**Conclusio:**

Mit der vorliegenden Machbarkeitsstudie ist damit grundsätzlich dargelegt, dass der Ansatz eines virtuellen Netzwerks Gesundheitsdaten fachlich, rechtlich und technisch machbar ist. Er eröffnet die Option, einen im Anwendungsbereich nahezu unbeschränkten Rahmen für die gesundheitspolitisch erwünschte Ausweitung von Big Data-Analysen in der Gesundheitsforschung zu schaffen.

## Inhaltsverzeichnis

ZUSAMMENFASSUNG .....	I
INHALTSVERZEICHNIS.....	III
ABBILDUNGSVERZEICHNIS.....	V
EINLEITUNG .....	1
<b>I. LÖSUNGSMODELL.....</b>	<b>2</b>
<b>A. Grundsätzliche Aufgabe des Netzwerk Gesundheitsdaten (NGD) und Einführung in die organisatorische Struktur.....</b>	<b>2</b>
<b>B. Beteiligte am „virtuellen Netzwerk Gesundheitsdaten“.....</b>	<b>2</b>
1. Nutzer .....	2
2. Netzwerk Gesundheitsdaten.....	2
3. Übergeordnete Governance .....	3
4. Kontrolle und Aufsicht.....	3
5. Personalressourcen .....	3
<b>C. Operative Funktionalitäten des Netzwerk Gesundheitsdaten .....</b>	<b>3</b>
1. Teilbereich 1 – Kontaktaufnahme über Plattform des NGD .....	3
2. Teilbereich 2 – Projektbeschreibung und Definition des benötigten Datensatzes.....	4
3. Teilbereich 3 – Projekt-Akzeptanz und Vertragsabschluss .....	6
4. Teilbereich 4 – Datenübertragung von Netzwerknutzern an Neutrale Stelle des NGD.....	6
5. Teilbereich 5 – Datenanalyse .....	7
6. Teilbereich 6 – Ergebnisübergabe/ Vertragserfüllung.....	8
7. Teilbereich 7 – Datenlöschung.....	8
<b>D. Lösungsmodell NGD.....</b>	<b>9</b>
<b>II. ANFORDERUNGEN AN EINE UMSETZUNG .....</b>	<b>10</b>
<b>A. Modellanwendung auf Beispiele und IST-Analyse.....</b>	<b>10</b>
1. Beispiel 1.....	10
2. Beispiel 2.....	14
<b>B. Zwingende rechtliche Anforderungen, insbesondere des Datenschutzes.....</b>	<b>17</b>
1. Rechtlich-organisatorische Grundlagen.....	17
2. Datenschutzrechtliche Zulässigkeit der Arbeit des NGD.....	18
3. Anforderungen an die organisatorische Ausgestaltung von Datenschutz und IT-Sicherheit.....	25
4. Vertragliche Abbildung der Arbeit des NGD und die Nutzung durch die Nutzer.....	26
5. Erforderlichkeit regulativer Anpassungen .....	27
<b>C. Technische Anforderungen an eine Umsetzung .....</b>	<b>27</b>
1. Begriffsklärung.....	28

2.	Big Data-Plattformen und Datenkataloge.....	29
3.	Interoperabilität und Standards.....	34
<b>D.</b>	<b>Stakeholder-Beteiligung.....</b>	<b>39</b>
<b>E.</b>	<b>Differenzanalyse .....</b>	<b>47</b>
<b>III.</b>	<b>MACHBARKEITSBETRACHTUNG UND UMSETZUNGSVORSCHLAG .....</b>	<b>49</b>
<b>A.</b>	<b>Datenqualität und -verfügbarkeit, Qualitätsentwicklung .....</b>	<b>49</b>
<b>B.</b>	<b>Überlegungen zur Herstellung technischer und semantischer Interoperabilität.....</b>	<b>50</b>
<b>C.</b>	<b>Infrastruktur und Technologie .....</b>	<b>51</b>
1.	Konzept und Architektur des technischen Zielmodells.....	52
2.	Anonymisierungs- und Pseudonymisierungskonzept.....	57
3.	Sicherheitskonzept.....	58
4.	Bewertung .....	62
<b>D.</b>	<b>Governance, Organisation und rechtliche Ausgestaltungsoptionen .....</b>	<b>64</b>
1.	Rollen in der Governance des NGD .....	65
2.	In der Governance abzubildende Kernprozesse.....	66
3.	Rechtsform des NGD und Ausgestaltung der Rechtsbeziehungen.....	67
4.	Rechtlich-organisatorische Ausgestaltung der Empfehlung .....	71
5.	Ausgestaltung des IT-Sicherheits- und Datenschutzmanagements.....	74
6.	Ausgestaltung der Stakeholder-, Gewährleistungs- und Aufsichtsfunktionen.....	76
<b>IV.</b>	<b>WEITERES VORGEHEN .....</b>	<b>77</b>
1.	Stakeholder-Dialog .....	77
2.	Pilotierung .....	78
3.	Feinkonzept .....	79
4.	Aufbau .....	80

## Abbildungsverzeichnis

ABBILDUNG 1: ORGANISATIONSMODELL NETZWERK GESUNDHEITSDATEN	2
ABBILDUNG 2: LÖSUNGSMODELL NGD	9
ABBILDUNG 3: KODIERUNG EINES MEDIZINISCHEN SACHVERHALTS IN UNTERSCHIEDLICHEN STANDARDS	36
ABBILDUNG 4: FUNKTIONALE ARCHITEKTUR NGD	52
ABBILDUNG 5: HADOOP-BASIERTE TECHNISCHE REFERENZARCHITEKTUR	54
ABBILDUNG 6: ARCHITEKTUR DER KONTAKTPLATTFORM	55
ABBILDUNG 7: BEISPIEL FÜR METADATEN-DATENSATZ	56
ABBILDUNG 8: APACHE STANBOL ALS CMS-ERWEITERUNG	56
ABBILDUNG 9: BLOCKBILD FUNKTIONALITÄTEN KONTAKTPLATTFORM	56
ABBILDUNG 10: SICHERHEITSKONZEPT	60
ABBILDUNG 11: EINSATZ VON KNOX	61
ABBILDUNG 12: BEISPIEL SEMANTISCHE INTEROPERABILITÄT	63
ABBILDUNG 13: VORGEHENSVORSCHLAG	77

## Einleitung

Entlang sämtlicher Prozessketten des Gesundheitswesens – von der Prävention über die Versorgung und Leistungsabrechnung bis hin zur klinischen Forschung – werden große Mengen von Daten zu einzelnen Personen über verschiedene im Gesundheitssystem agierende Organisationen und Institutionen erhoben und entsprechend ihres Erhebungszwecks verarbeitet. Schon seit langem wird erkannt, dass diese Fülle an Daten einen Wert an sich darstellt, der umso größer ist, je mehr Daten zu einer Person oder zu einer Fragestellung zusammengeführt und ausgewertet werden können. Durch die Erweiterung des eigenen Datensatzes, können Fragestellungen – auch über die eigenen Organisationsgrenzen hinaus – beantwortet werden, die mit dem eigenen Datenbestand alleine nicht zu beantworten wären oder auch Fragestellungen evident werden, die sich ohne Datensatzerweiterungen nicht gestellt hätten. Daraus kann sich bei geeigneter Ausgestaltung des vorgeschlagenen Konzeptes eine „win-win“-Situation für alle am Datenaustausch Beteiligten ergeben.

Mit diesem Thema hat sich in den letzten Jahren bereits auch das Bundesministerium für Gesundheit befasst und so z.B. in der Studie „Weiterentwicklung der eHealth-Strategie“ mögliche Potenziale von „Big Data“ für das Gesundheitswesen aufgezeigt. Einzelne Einrichtungen (z.B. Deutsches Krebsforschungszentrum in Heidelberg) haben im Kontext von Forschungsprojekten entsprechende Infrastrukturen aufgesetzt und auch in den Ländern (z.B. Bayern) sind Infrastrukturen in Planung, über die sektorenübergreifend Patientendaten im Rahmen der Versorgungsforschung und zur Unterstützung der Versorgung in „Big Data Pools“ zur Auswertung zusammengeführt werden sollen.

Im Rahmen des Nationalen Aktionsplans für Menschen mit Seltenen Erkrankungen (NAMSE) fördert das BMG die Etablierung des „Open-Source-Registersystems für Seltene Erkrankungen (OSSE)“, ein Metadaten-Verzeichnis (Metadata Repository, MDR), das die Interoperabilität verschiedener Register ermöglichen soll. Über den vom BMBF finanzierten Aufbau von Datenintegrationszentren an Unikliniken werden weitere solcher Datenzusammenführungen entstehen, die darauf ausgerichtet sind, möglichst große Kohorten von Patienten zur Erforschung von Krankheiten und ihrer bestmöglichen Behandlung zu gewinnen. Auch die Industrie hat auf diese Entwicklungen reagiert und bietet vielfältige Lösungen für die Auswertung von Gesundheitsdaten an (z.B. SAP HANA, Google DeepMind). Der britische National Health Service (NHS) hat dem US-Unternehmen Google im Gegenzug für die Nutzung einer AI-Lösung zur Datenauswertung weitreichenden Zugang zu den Daten von 1,6 Millionen Patienten gewährt.

Hierbei sichern sich Anbieter von Datenanalyse-Lösungen zuweilen weitreichende Rechte an den Daten und Auswertungsergebnissen. Ob diese dann durchgängig mit den höchsten Sicherheitskriterien und dem Gedanken eines Zugewinns für die individuelle Patientenversorgung einhergehen und sich damit am Gemeinwohl der Bevölkerung ausrichten, kann weder gesteuert noch kontrolliert werden.

Grundsätzlich wird durch verschiedene Ansätze begründet davon ausgegangen, dass durch die institutionsübergreifende Zusammenführung von Daten Mehrwerte sowohl im Bereich der Versorgung als auch im Bereich der Forschung gewonnen werden können.

Eine wesentliche Forschungslücke besteht dabei für die Fragestellung, wie – als Alternative zu gemeinsamen Datenpools oder Echtzeitverknüpfungen – eine temporäre, vorübergehende und flüchtige Zusammenführung von Daten ausgestaltet werden kann, um Versorgungs- und Forschungsinteressen mit Datenschutz und Sicherheit in Einklang zu bringen. Mit der Machbarkeitsstudie wird das von Dr. Ulrike Helbig entwickelte fachliche Modell weiter ausgearbeitet, das eine Zusammenführung von Daten und deren Analyse fallweise und auf Grundlage der an den jeweiligen Interessen der Akteure des Gesundheitswesens orientierten Vereinbarungen vorsieht, um die Potentiale von Big Data zu nutzen, ohne kritische zentrale Datenbestände zu benötigen.

Mit der vorliegenden Studie soll das Virtuelle Netzwerk Gesundheitsdaten als alternatives Modell fachlich, organisatorisch, rechtlich und technisch beschrieben und auf Machbarkeit untersucht werden.

## I. Lösungsmodell

### A. Grundsätzliche Aufgabe des Netzwerk Gesundheitsdaten (NGD) und Einführung in die organisatorische Struktur

Das Modell des virtuellen Netzwerk Gesundheitsdaten (NGD) sieht vor, eine Organisationsstruktur zu schaffen, in die Einrichtungen des Gesundheitswesens die bei ihnen vorgehaltenen personenbezogenen oder nicht personenbezogenen Daten – es werden im allgemeinen Gesundheitsdaten verarbeitet, zu denen auch genetische Daten und Analysen aus Gewebeproben gehören - für gemeinsame Auswertungen über den gesamten Datenbestand passager über das NGD zur Nutzung einbringen können und über die „Big-Data-Analysen“ erfolgen können.

Organisatorisch beinhaltet das NGD den Bereich für Kontaktaufnahme über die digitale Plattform, den Bereich für die Vertragsabschlüsse und den Bereich für die Datenzusammenführung und Datenanalyse („Neutrale Stelle“). Nutzer sind die Institutionen und Organisationen des Gesundheitswesens sowohl aus dem öffentlichen als auch aus dem privaten Sektor. Wegen der Datensensibilität und der Sensibilität des Prozesses über das NGD wäre eine externe Aufsicht wünschenswert (Abbildung 1).

Neben der breiteren Nutzbarmachung durch Datenzusammenführung und Datenanalyse auf eine spezifische Nachfrage hin, erfüllt das Netzwerk Gesundheitsfragen einen gesellschaftlichen Auftrag im Sinne der Versorgungsentwicklung, der auf Nebenerkenntnissen entlang der Gewinnung von Analyseergebnissen beruht.

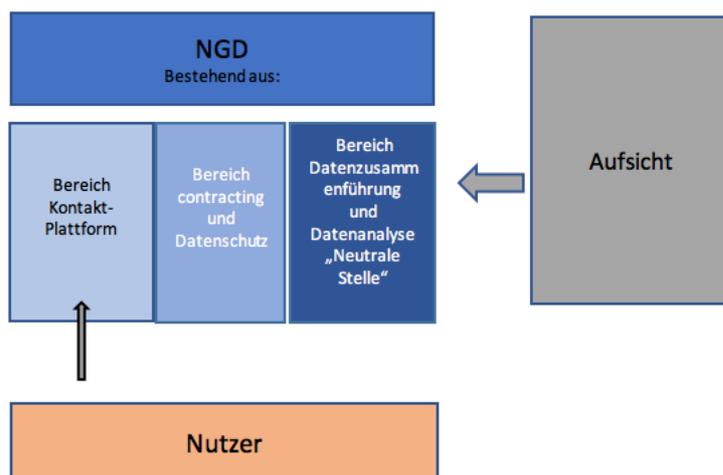


Abbildung 1: Organisationsmodell Netzwerk Gesundheitsdaten

### B. Beteiligte am „virtuellen Netzwerk Gesundheitsdaten“

#### 1. Nutzer

Öffentliche und private Einrichtungen des Gesundheitswesens können Nutzer des NGD sein. Dazu gehören im ersten Schritt deutsche Einrichtungen (wie z.B. das RKI, Krebsregister, KVen, Forschungsinstitutionen), im weiteren Verlauf können aber auch europäische Einrichtungen das NGD nutzen. Je nach Anfrage können sie dabei als Datenbereitsteller (DB) oder Datennutzer (DN) fungieren. Im Modell gehen wir davon aus, dass die Einrichtung einen rechtlichen Sitz in Deutschland hat.

#### 2. Netzwerk Gesundheitsdaten

Das Netzwerk Gesundheitsdaten (NGD) ist die zentrale Organisationsstruktur mit den drei organisatorischen Bereichen

- - Plattform zur Kontaktaufnahme,
- - Bereich für das Contracting und Datenschutz,

- - Bereich für die Datenzusammenführung und Datenanalyse (Neutrale Stelle),

über die alle operativen Schritte durchgeführt, die Nutzer mit ihren Kontaktanfragen verwaltet werden, Daten-Verzeichnisse erstellt und gehalten werden, Auswertungen durchgeführt und dokumentiert werden, der Datenschutz und IT-Sicherheit gewährleistet werden.

### 3. Übergeordnete Governance

Dem Netzwerk Gesundheitsdaten liegt eine übergeordnete Governance zu Grunde, die für das Entstehen und den Bestand des Netzwerks sorgt. Im Hinblick auf die Daseinsvorsorge-Aufgabe des Netzwerks gehen wir von einer öffentlich verantworteten Governance aus.

### 4. Kontrolle und Aufsicht

Die Nutzung von Daten im Netzwerk Gesundheitsdaten unterliegt datenschutzrechtlicher Aufsicht. Auch die Nutzer des NGD unterliegen der Datenschutz-Aufsicht. Wegen der Heterogenität der möglichen Teilnehmer sind grundsätzlich alle Datenschutz-Aufsichtsbehörden betroffen, weshalb die Datenschutz-Aufsicht als eine organisatorische Rolle mit inhaltlich, funktional zu formulierenden Inhalten verstanden wird.

### 5. Personalressourcen

Die Kontaktplattform benötigt speziell geschultes Personal für die Dateneingabe und die Kontrolle der Daten, Webentwickler für die Entwicklung und den Betrieb der Kontaktplattform und zusätzlich IT-Sicherheits- und Datenschutz-Spezialisten zur Gewährleistung der Sicherheit. Zudem benötigt die Kontaktplattform geschultes Personal für den Kundensupport.

Im Bereich Contracting und Datensicherheit müssen Juristen und Datenschutzbeauftragte zur Erstellung von Verträgen und für die Beratung einbezogen oder eingestellt werden.

Die Neutrale Stelle benötigt für die Analyse spezialisiertes Fachpersonal, welches sich in drei Gruppen unterteilt: Datenschutz und Datensicherheitsspezialisten, Fachpersonal zur Betreuung der IT-Infrastruktur, Medizinstatistiker. Bei der personellen Besetzung ist auf besonders qualifiziertes Personal mit hoher Eigenverantwortlichkeit und Selbstmotivation zu achten. Die Bedeutung der Datenauswertung mit Bezug auf den individuellen Schutz und in Perspektive auf die Entwicklung des Allgemeinwohls sollte durchweg präsent sein.

## C. Operative Funktionalitäten des Netzwerk Gesundheitsdaten

Das NGD kann in sieben operative Funktionalitätsbereiche unterteilt werden (siehe Abbildung 2).

### 1. Teilbereich 1 – Kontaktaufnahme über Plattform des NGD

Der erste Bereich bildet als webunterstützte Plattform das Kontaktnetzwerk für Gesundheitsinstitutionen, -organisationen und -unternehmen im Gesundheitssystem (RKI, DIMDI, Krebsregister, KVen, Forschungsinstitute etc.) ab.

Diese stellen als NGD-Nutzer ihre zum Datenaustausch relevanten institutionellen Kontaktdaten ein und formulieren in einer Selbstbeschreibung für andere Teilnehmer sichtbar ihr Aufgabenspektrum, die darauf basierenden jeweiligen Datenerhebungen und vorgehaltenen Datensammlungen. Diese Informationen werden dann vom NGD als Index bzw. Register auf der Plattform ebenfalls bereitgestellt. Das Konzept des NGD sieht vor, dass in einem zweiten Schritt auch Institutionen aus dem Europäischen Raum Datenanalyse-Anfragen stellen und auch Daten zur Nutzung bereitstellen können.

Im Bereich der öffentlichen Gesundheitsversorgung agierende Institutionen und Organisationen werden aufgrund ihrer Funktion im gesellschaftlichen Kontext zur Bereitschaft, Daten auf Anfrage zu Analyse Zwecken an das NGD zu übermitteln, verpflichtet. Damit ist verbunden, dass sich Institutionen/Organisationen auf der webunterstützten Plattform des NGD einzutragen haben. Die in der freien Wirtschaft agierenden Unternehmen partizipieren freiwillig und aus eigenem Interesse und können gegen eine für den öffentlichen Bereich der Gesundheitsversorgung angemessene, noch zu ermittelnde,

aufwandsgerechte Gegenleistung für die durch das NGD erbrachten Leistungen nutzen. Eine Anmeldung und ein Eintrag auf der webunterstützten Plattform des NGD ist dafür notwendig.

Von dem anfragestellten Teilnehmer können die Kontaktdaten und das eingetragene Register der verfügbaren Datenbasen bzw. der Index aller Plattformteilnehmer eingesehen werden. Datensätze selbst werden auf dieser Plattform jedoch nicht präsentiert, sondern Datenbeschreibungen. Die Neutrale Stelle als Datenzusammenführungs- und Datenanalysebereich des NGD wird von den Plattformteilnehmern unterstützt, einen internen, nicht für die Nutzer des NGD sichtbaren Datenindex in einem zentralen Daten-Board anzulegen. Die Nutzer der Netzwerks Gesundheitsdaten werden hierzu ebenfalls verpflichtet. Die Neutrale Stelle des NGD ist die operative Einheit des NGD. Das Netzwerk Gesundheitsdaten wahrt die Interessen seiner Vertragspartner, den Nutzern des NGD, und tritt nicht in den Wettbewerb mit ihnen. Das NGD nutzt in einer internen Analyse Prozessergebnisse zur Qualitätsentwicklung und stellt diese Ergebnisse den NGD-Nutzern zur Verfügung, um eine generelle Qualitätsverbesserung zu unterstützen. Dies geschieht unter Kontrolle der Aufsicht führenden Behörde.

Die Intention eines Nutzers ist es, für eine spezifische Fragestellung den eigenen Datensatz um die Daten eines anderen oder mehrerer anderer Plattformnutzer/s zu erweitern, um im Sinne der Fragestellung die Datenanalyse durch die Neutrale Stelle des NGD durchführen zu lassen. Eine Übermittlung der Datensätze selbst (weder des Interessenten noch des Angefragten) ist im ersten Schritt der Kontaktaufnahme nicht notwendig. Der Anfragende weiß aufgrund einer eigenen statistischen Analyse, welche Datensätze und in welchem Umfang diese für die Beantwortung seiner Fragestellung notwendig sind und gibt diese Information für eine Anfrage ausreichend in grober Form über die webunterstützte Plattform an den möglichen Vertragspartner und in für den Informationsgewinn und zur Vertragsgestaltung notwendig detaillierter Form an das NGD. Das NGD kann bei der Suche von Datenbereitstellern (DB) beratend und unterstützend tätig sein.

Über die Kontakt-Plattform des NGD ist nach Anfrage eines interessierten Nutzers eine gezielte Kontaktaufnahme mit anderen Gesundheitsinstitutionen, -organisationen und -unternehmen möglich, in der die zu untersuchende Fragestellung formuliert und der notwendige Datensatz angefragt wird.

Zusammengefasst bietet die webbasierte Kontakt-Oberfläche der Plattform folgende Funktionen für die Plattformnutzer:

- Kontaktadresse der Nutzer mit Unternehmensprofilen
- Informationen zur allgemeinen passageren (= temporären) Datenzusammenführung werden gegeben, sowie Informationen zur Datensicherheit, dem Contracting, der Analyse der zusammengeführten Daten, den Verfahren zur Datenanalyse und der Datenlöschung.
- Kontaktaufnahmen können auf Anfrage über die Kontakt-Plattform (digital und analog) als organisatorische Einheit des NGD durchgeführt werden
- Anfragen zur Datennutzung können über die Kontakt-Plattform des NGD gestellt werden. Eine teilautomatisierte Abwicklung wird angestrebt.

Intern für das NGD:

- Über die Plattform gestellte Anfragen können durch das NGD zu verschiedenen Themengruppen geclustert und entsprechend ausgewertet werden.

## 2. Teilbereich 2 – Projektbeschreibung und Definition des benötigten Datensatzes

Ein Plattform-/ Netzwerk-Nutzer stellt seine Anfrage zum Datenaustausch und zur Datenanalyse über die Plattform des NGD. Im ersten Schritt sind bi- oder multilaterale Kontaktaufnahmen ohne die anderen organisatorischen Bereiche (Contracting und Neutrale Stelle) des NGD möglich. Konkrete Anfragen können an den/die möglichen Vertragspartner adressiert werden. Bei gegenseitigem Interesse wird dann der Bereich Contracting des NGD zur Vertragsformulierung einbezogen. Grundsätzlich ist eine Austausch-Symmetrie zwischen den verschiedenen Plattformnutzern für ein Zustandekommen des Projektes oder zur Vertragskonzeption nicht notwendig.

Die Anfrage eines Plattformnutzers beschreibt eine Fragestellung, die beforscht werden soll, oder das konkrete Forschungsvorhaben. Die Anfrage beschreibt den Inhalt, die benötigten Datenarten und je nach Entwicklungsstadium der Anfrage die Menge der von einem oder mehreren anderen

Plattformnutzern benötigten Daten und die Art der geplanten Auswertung. Zur Datenanfrage werden dafür bereitgestellte Formular-Templates genutzt. Im Laufe der Aktivitäten des NGD werden diese Formular-Templates zur strukturierten Eingabe entsprechend der geclusterten fachlichen Schwerpunktsetzung weiterentwickelt und spezifiziert. Das Formular-Template enthält neben dem freien Eingabefeld für die Anfrage (z.B. hier siehe die ausgewählten Beispiele) folgende Rubriken: anfragendes Institut, Fachbereich, Diagnose(n), Krankheitsstadium, Eingriff, Therapie, Gender, Altersgruppen, Studienart, Studiendauer, Patientenzahl, notwendige/s Institut/e für eine Datenbereitstellung, Datenmenge (siehe hierzu z.B. auch Datensatz klinischer Krebsregister).

Diese Formular-Templates sollen soweit wie möglich strukturiert werden mit Hilfe existierender Ontologien und Standards. Diese werden auch iterativ verbessert durch die gewonnenen Erfahrungen aus zuvor ausgeführten Studien. Dies erlaubt eine bessere Strukturierung und Standardisierung der Formulare, limitiert die Notwendigkeit der Freitextangabe und führt mit der Zeit zu einer zunehmend automatischen oder teil-automatischen Identifizierung von passenden Datensätzen für die Studie.

Zur genaueren Prüfung wird über den Bereich Kontakt-Plattform des NGD eine interne Anfrage bzw. Schlagwortsuche an das Register bzw. den Index der durch die Plattformnutzer bereitgestellten Datenlisten gestellt. Diese Datenlisten sind in Form eines Datenkataloges angelegt.

Anhand der Anfrage mittels des ausgefüllten Template sucht der Bereich Kontakt-Plattform des NGD Vertragspartner, i.e. Datenbereinsteller (DB) über den internen Datenkatalog. Nur das NGD hat Zugriff auf diesen internen Datenkatalog und weiß, wer welche Daten potentiell zur Verfügung stellen kann (z.B. über SNOMED CT, das aber bisher nicht flächendeckend genutzt werden kann). Innerhalb des NGD gibt es ein iteratives Vorgehen entsprechend der Funktionsweise von Call-centers zur Spezifizierung des auszufüllenden Frageformulars (Template). Der intern angelegte Datenkatalog wird gefüllt durch die Plattformnutzer (DN und DB). Entlang des intern des NGD angelegten Datenkataloges wird das auf der webbasierten Plattform sichtbare Datenregister/ Index weiterentwickelt.

Wenn der/die angefragte/n Plattformnutzer und potentielle DB Interesse an einer Annahme der Anfrage hat/haben, kann er/können sie das auf direktem Weg über die webbasierte Kontakt-Plattform eingeben und sich dann durch den hergestellten Kontakt ausführlicher über die Plattform an den Anfrager wenden. Ein teilautomatisierter Prozess ist möglich.

Wenn die Anfrager als potentielle DN und die Adressierten als potentielle DB Interesse an einer passageren Datenzusammenführung und -Analyse haben, wendet sich der anfragende Teilnehmer (DN) an das NGD mit der Bitte um Vertragsgestaltung.

Das Institut – der potentielle DB – wird im nächsten Schritt von der Neutralen Stelle des NGD angefragt zu prüfen, ob der spezifische notwendige Datensatz vorhanden ist. Das NGD kann bei Bedarf weitere Institute/Organisationen zur Ergänzung des benötigten Datensatzes vermitteln.

Datensatz-Formulierung: Der benötigte Datensatz wird vom Anfrage stellenden Nutzer (DN) des NGD selber formuliert, diese Anfrage geht an das NGD. Das NGD kennt den gesamten Datenindex und sucht ggf. ergänzende Institute – potentielle Datenbereinsteller (DB) – heraus. Der DB muss die geforderten Daten selektieren soweit wie möglich, sie in die gewünschten Standards überführen und dem NGD übermitteln, zusammen mit einer Beschreibung des Inhaltes und der Struktur der Daten. Innerhalb der Datenbereinsteller/ Datennutzer optionierte pseudonymisierte Daten müssen ggf. zum Zweck der Zusammenführung und des Matchings von den Datenbesitzern innerhalb der eigenen Institution wieder in den Personenbezug gebracht werden. Damit können sie nach Übersendung an das NGD mit den Daten der weiteren Institutionen, die ggf. dasselbe Verfahren durchlaufen müssen, innerhalb des NGD zusammengeführt werden.

Zusammengeführte Daten (pseudonymisiert/personen-bezogen) werden in der Neutralen Stelle des NGD zusammengeführt und gematcht und dann für die Analyse erneut pseudonymisiert (siehe TMF, Medizinkonsortium). Falls die Neutrale Stelle feststellt, dass die Analyse mit anonymisierten oder pseudonymisierten Daten gemacht werden kann, übermittelt sie den Datenbereinstellern den genauen Prozess der Anonymisierung/Pseudonymisierung, und die Analyse wird nur mit pseudo-/ anonymisierten Daten durchgeführt. Falls dies nicht der Fall sein sollte und personenbezogene Daten für die Integration für die verschiedenen Datensätze notwendig sind, so muss der DB auch diese Daten

zusenden. Die Sicherheit der Daten ist dabei immer durch das Sicherheitskonzept des NGD gewährleistet.

Eine Anfrage bestimmter Daten über das NGD soll vor der Weiterleitung an Datenbereinsteller auf ihre Vereinbarkeit mit dem Zweck des Netzwerks geprüft werden, soweit möglich auch bereits die datenschutzrechtliche Zulässigkeit. Eine abschließende Prüfung der Rechtsgrundlage der Auswertung sowie der Einhaltung weiterer datenschutzrechtlich relevanter Vorschriften erfolgt im Rahmen der Vertragsaushandlung.

### 3. Teilbereich 3 – Projekt-Akzeptanz und Vertragsabschluss

Im weiteren Schritt erfolgt eine teil-automatisierte Vertragsaushandlung mit den Einrichtungen, die Daten für ihre Anfrage nützen werden, und Einrichtungen, die Daten für die Anfrage bereitstellen werden (DB), sowie auch mit dem Netzwerk Gesundheitsdaten, das dort in der organisatorischen Einheit der Neutralen Stelle die gewünschte Analyse durchgeführt werden soll. Voraussetzung für den Vertragsabschluss ist die vorausgegangene Freigabe durch die zuständige Ethikkommission.

Das bedeutet auch, dass die rechtliche Basis der Datenzusammenführung durch das Netzwerk Gesundheitsdaten geprüft wird. Hierbei sind die rechtliche Zweckbindung der Daten und die Beziehung zum Zweck der Auswertung zu prüfen. Im Ergebnis dieser Aushandlung ist die Zweckbindung der Auswertung mit den datenschutzrechtlichen Vorgaben einer Nutzung der einzubeziehenden Daten abgeglichen, die jeweiligen Rechte zur Verwertung der Analyseergebnisse durch alle teilnehmenden Nutzer des NGD (DN und DB) fixiert und ein Vertragsentwurf entsprechend dem Forschungsvorhaben und der dafür erwünschten und notwendigen, passager zusammenzuführenden Daten für die beteiligten Vertragspartner verfasst.

Die Vertragsgestaltung und –abwicklung wird durch den organisatorischen Bereich Contracting und Datenschutz des NGD durchgeführt. Es wird auch vertraglich formuliert, dass der Datenanalyst aus der Neutralen Stelle bei Fragen des DN und Auftraggebers, mit diesem nach Prüfung durch den Bereich Contracting und Datenschutz schriftlich Rücksprache halten kann.

In diesem Schritt erfolgt durch den Bereich Contracting und Datenschutz des NGD eine abschließende Prüfung der datenschutzrechtlichen Zulässigkeit der geplanten Vertragsgestaltung. Für die Ausgestaltung der Neutralen Stelle und das Contracting ist es zwingend erforderlich, dass diese eine rechtsfähige Einrichtung ist, die vertrauenswürdig und fachlich akzeptiert ist und vor allem von keinem Nutzer beherrscht wird. Es muss daher sichergestellt sein, dass keine Einflussmöglichkeit und keine Missbrauchsmöglichkeit von Daten oder Analyseergebnissen möglich ist. Die Teil- bzw. vollständige Automatisierung der Vertragsabwicklung wird dadurch möglich gemacht, dass nach einer bestimmten Laufzeit des Projektes genügend Vertragsmuster gesammelt werden können. Durch die gut annotierten Datensätze und Studienbeschreibungen kann das System ermitteln, wenn eine neue Studie ähnliche Voraussetzungen bezüglich notwendiger Daten sowie Analyseverfahren erfüllt, wie bereits in der Vergangenheit durchgeführter Studien. In diesen Fällen werden den Teilnehmern neuer Studien bereits existierende Vertragsmuster vorgeschlagen.

### 4. Teilbereich 4 – Datenübertragung von Netzwerknutzern an Neutrale Stelle des NGD

Dieses Modell hat den höchsten Schutzbedarf zu erfüllen. Damit die maximale Sicherheit der Datenübertragungen gewährleistet werden kann, wird dies durch die Einhaltung von zwei wichtigen Schritten zur Datenübertragung umgesetzt. Die Datenverschlüsselung wird durch ein Public und Private Key Verschlüsselungsverfahren gewährleistet, in dem die „Neutrale Stelle“ ein Public/Private Key-Paar erstellt und dann den Public Key dem Datenbereinsteller zur Verschlüsselung der Daten zusammen mit einer Beschreibung des Verschlüsselungsprozesses übergibt. Der Private Key bleibt hingegen bei der „Neutralen Stelle“, damit die Daten nur in der gesicherten „Neutralen Stelle“ entschlüsselt werden können. Zusätzlich werden die verschlüsselten Daten auf einen Datenspeicher geschrieben und auf höchstem Schutzbedarf gesichert an die „Neutrale Stelle“ übertragen.

Der Datenbereitsteller muss sicherstellen, dass er eine genaue Kopie der gelieferten Daten im gesetzlich vorgeschriebenen Zeitraum bei sich vorhält, damit die Validität der Beauftragung und der durchgeführten Analyse innerhalb dieses Zeitraumes nachvollziehbar ist.

## 5. Teilbereich 5 – Datenanalyse

Nach Datenzusammenführung der Datenpakete in dem organisatorischen Bereich Neutrale Stelle des NGD erfolgt die Auswertung entsprechend des formulierten Forschungsvorhabens.

Die Zusammenführung benötigt eine Datenintegration. Diese geschieht in zwei Schritten: Zuerst teilautomatisiert durch die Benutzung gleicher Ontologien und Standards von den Datenbereitstellern, danach falls notwendig manuell durch die Medizinstatistiker zu Sicherstellung einer hohen Qualität. Daher ist bei der personellen Besetzung auf besonders hochqualifiziertes Personal mit einem hohen Maß an Eigenverantwortlichkeit und Selbstmotivation zu achten. Die Bedeutung der Datenauswertung mit Bezug auf den Datenschutz und in Perspektive auf die Entwicklung des Allgemeinwohls soll durchweg präsent sein.

Die Neutrale Stelle des Netzwerk Gesundheitsdaten führt überwiegend statistische Analysen numerischer Daten durch. Daten können anonymisiert und pseudonymisiert ausgewertet werden. Um die notwendige Qualität der Analyseergebnisse gewährleisten zu können, muss das System den modernsten Standard erfüllen, Weiterentwicklungsfähigkeit und Anpassungsfähigkeit an Anfragen und externe Anforderungen muss gegeben sein, es muss eine Skalierbarkeit, Gruppierbarkeit von Datensätzen und Anpassung bzw. Weiterentwicklung von erstellten Algorithmen möglich sein. Auch Bilddaten sollten verarbeitbar sein.

Die Neutrale Stelle, führt die Datenpakete, die von den Vertragspartnern vertraglich beschrieben, ausgewählt und an die Neutrale Stelle übermittelt wurden, zusammen, speichert sie passager (temporär) für die Datenanalyse und sichert sie entsprechend den vorgeschriebenen Kriterien, welche sich aus den datenschutz- und IT-sicherheitsrechtlichen Anforderungen ergeben.

Dann folgt die statistische Analyse der numerischen (im Rahmen einer weiteren und späteren Entwicklung können auch Bilddaten analysiert werden) Daten und deren Analyse entsprechend Fragestellung und Vertrag. Das Analyseergebnis wird durch die Neutrale Stelle des Netzwerk Gesundheitsdaten gegen Verfälschung gesichert und so mit Hinweisen auf die eingeflossenen Daten verknüpft, dass eine Nachvollziehbarkeit des Ergebnisses gegeben ist. Die Datensätze, die für die Analyse notwendig waren, werden im Nachgang der Analyse von den Nutzern des NGD (DN und DB) entsprechend der vertraglichen Vereinbarung aufbewahrt, um die Analyse bei Bedarf verifizieren zu können. Außerhalb der Neutralen Stelle des NGD findet keine Datenanalyse statt. Das Netzwerk Gesundheitsdaten muss in der Lage sein, die durch das NGD ausgeführten Analyseergebnisse für eine Metaanalyse zur Gesundheitsversorgung, zur Entwicklung von machine-learning und zur Metaanalyse für eine Qualitätsentwicklung nutzen zu können. Dies wurde im Contracting mit den Nutzern (DN und DB) entsprechend formuliert. Auswertungen sind nur über die zu Vertragsabschluss vorhandenen Daten möglich. Eine prospektive Datenanalyse (prospektiv randomisierte Studie) kann nur so erfolgen, dass die Plattformnutzer die Daten institutsintern sammeln und dann zum Analysezweck auf das NGD zukommen. Eine ongoing-Datenerhebung und ongoing-Analyse im NGD über einen longitudinalen Zeitraum hinweg ist nicht möglich. Um die Analyse durchführen zu können muss die Neutrale Stelle über eine eigene, den hohen Sicherheitsanforderungen entsprechende Big Data-Plattform verfügen.

Die aufgrund der Vertragsabschlüsse durchgeführten Analyseprozesse und die Analyseergebnisse werden von der Neutralen Stelle des NGD gespeichert. Dies ist vertraglich mit den Nutzern vereinbart und wird von der das NGD kontrollierenden Aufsicht geprüft und genehmigt. Sie können von den Vertragspartnern erneut angefragt und repliziert werden und dürfen von der Neutralen Stelle des NGD für Metaanalysen genutzt werden.

Das NGD hat das Recht, die Daten für machine-Learning zu nutzen. Die Gestaltung der technischen Möglichkeiten ist dabei so ausgerichtet, dass übergreifende Erkenntnisse entwickelt und dargestellt werden können.

Wegen der Analyse von u.a. personenbezogenen Daten und der damit verbundenen Datenvulnerabilität, ist für das NGD die höchste Sicherheitsstufe anzusetzen. Das NGD ist daher zu einem besonderen Datensicherungskonzept verpflichtet und muss unter besonderer Aufsicht und Kontrolle stehen.

Um die Prozesse zur Nutzung von personenbezogenen Daten zu vereinfachen und unter der Voraussetzung, dass das NGD mit seiner Neutralen Stelle besondere Sicherheitsvorkehrungen treffen muss und unter regulativer Aufsicht steht, darf das NGD personenbezogene Daten verwenden, ohne nochmals die vorherige Einwilligung dieser Person einholen zu müssen. Weiterführende Erkenntnisse zu einzelnen Personen oder Personengruppen darf das NGD nicht gewinnen. Eine besondere Aufsicht und Regulierung sind notwendig. Dies ist ebenfalls notwendig, da vorgesehen ist, dass die Neutrale Stelle des NGD Metadatenanalysen, zum Beispiel im Interesse der Weiterentwicklung der Versorgungsqualität, durchführen kann und die Ergebnisse eben dieser kontrollierenden Aufsicht vorgestellt werden müssen, bevor sie einer Öffentlichkeit zugänglich gemacht werden.

Die Durchführung der Analyse der Datensätze erfolgt ausschließlich durch die Neutrale Stelle des NGD. Hierfür stehen Standard-Auswertungswerkzeuge zur Verfügung sowie die Möglichkeit des Datennutzers eigene Auswertungswerkzeuge mitzuliefern.

## 6. Teilbereich 6 – Ergebnisübergabe/ Vertragserfüllung

Anschließend übersendet die Neutrale Stelle des NGD die Analyseergebnisse über ein geeignetes und gesichertes System gemäß den vertraglichen Vereinbarungen der Nutzer (DN und DB) sowohl an DN als auch an DB. Dies geschieht so, dass sowohl die Analyseergebnisse als auch die Prozesse dauerhaft nachvollzogen werden können. Der Analyseprozess und das Analyseergebnis werden in der Neutralen Stelle des NGD reproduzierbar gesichert und können unter Kontrolle der Aufsichtführenden Behörde zu Zwecken der Metaanalyse mit Ausrichtung auf die Gesundheitsversorgung, zum machine-learning und zur Qualitätsentwicklung weiterverwendet und ausgewertet werden. Alle ursprünglich von den DN und DB anderen Daten (Rohdaten) werden nicht rückübermittelt, sondern gelöscht.

Sollte aufgrund einer im NGD durchgeführten Metaanalyse eine zusätzliche neue Erkenntnis mit möglicher Versorgungsrelevanz entstehen, muss das NGD diese Erkenntnis, nachdem sie durch den Datenschutzbereich geprüft und freigegeben wurde, an die beaufsichtigende Einrichtung geben..

Das Datenschutzmodell sieht vor, dass die Übermittlung der Analyseergebnisse stichprobenartig durch technische und organisatorische Maßnahmen auf eine mögliche Personenbeziehbarkeit überprüft wird. .

## 7. Teilbereich 7 – Datenlöschung

Im letzten Schritt werden die von den Nutzern des NGD übermittelten Datensätze sowie die zur Verschlüsselung notwendigen Verschlüsselungsdaten dauerhaft aus dem Speichersystem der Neutralen Stelle des Netzwerks Gesundheitsdaten gelöscht. Dies ist technisch, organisatorisch und rechtlich sichergestellt sowie vertraglich vereinbart und muss jederzeit auf Verlangen von Datenbereitstellern und Datennutzern belegt werden. Die Aufsichtführende Behörde prüft, dass die Daten tatsächlich vollständig gelöscht wurden.

Für die Löschung der Daten ist vom NGD ein Löschkonzept zu erstellen, welches eine regelmäßige Löschroutine der Datensätze definiert. Die kontrollierende Aufsicht überprüft diese endgültige Löschung der passager zusammengeführten Daten. Zudem müssen vom NGD die in Art. 5 Abs. 2 DSGVO normierten Rechenschaftspflichten eingehalten werden, mit denen auch Dokumentationspflichten im Bereich Neutrale Stelle des NGD einhergehen, welche eine längere Speicherung von Daten der Nutzer des NGD (DN und DB) notwendig machen kann. Dies muss ebenfalls in dem zu erstellenden Löschkonzept berücksichtigt werden.

## D. Lösungsmodell NGD

Im Überblick stellt sich das Lösungsmodell NGD wie folgt dar:

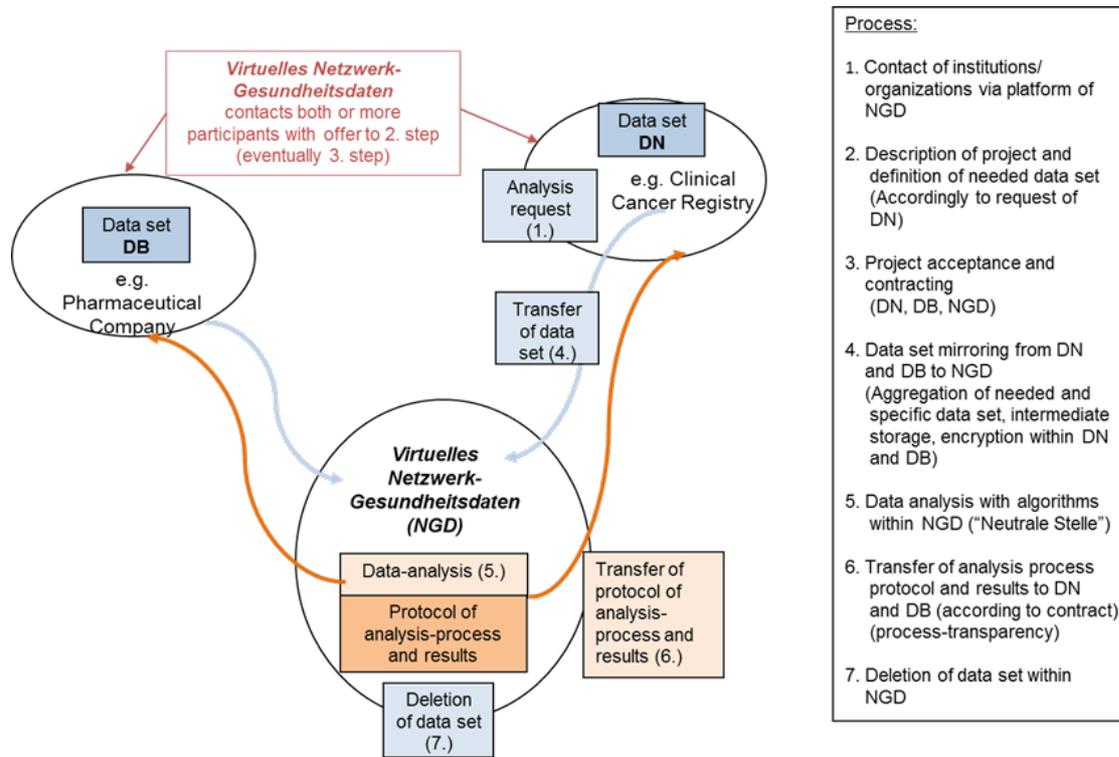


Abbildung 2: Lösungsmodell NGD

## II. Anforderungen an eine Umsetzung

### A. Modellanwendung auf Beispiele und IST-Analyse

Unabhängig davon, dass die Ergebnisse der hier gewählten Fragestellungen inhaltliche Relevanz haben, ging es bei der Auswahl um die Prüfungsmöglichkeit, inwieweit ein Datenaustausch verschiedener Datensätze von verschiedenen Institutionen zusammengeführt werden können. Hier ist von zusätzlicher Bedeutung, dass Daten von verschieden weit auseinander liegenden Erhebungszeitpunkten zusammengeführt werden müssen und auch, dass Daten aus dem Gesundheitsbereich mit Daten außerhalb des Gesundheitsbereichs kombiniert werden sollen. Die Lokalitäten der hier ausgewählten Kliniken sind willkürlich und ohne jede inhaltliche Bedeutung für die Darstellung und Prüfung der Beispiele.

#### 1. Beispiel 1

##### a) Problemstellung

A) Zusammenführung von Daten zur Untersuchung, ob die Durchführung einer Sektio caesarea (elektiv und im bereits laufenden Geburtsprozess) zur Erhöhung der Inzidenz eines Uterussarkoms führt.

- Zusammenführung von Daten aus DIMDI bzw. von Versicherungsdaten, den klinischen Krebsregistern und von Kliniken

B) Zusammenführung von Daten zur Untersuchung, ob der eigene individuelle Geburtsprozess (vaginale Geburt versus Geburt durch eine Sectio caesarea des Kindes) zu einer Erhöhung des eigenen, individuellen Risikos an einer onkologischen Erkrankung zu erkranken, führt.

- Zusammenführung von Daten aus dem Robert Koch-Institut, Versicherungsdaten und klinischen Krebsregisterdaten

##### b) Operative Funktionalitäten des „virtuellen Netzwerks Gesundheitsdaten“ im Bezug auf das 1. Beispiel:

##### i. Teilbereich 1 – Kontaktaufnahme über Plattform des NGD

Zur Frage steht, ob die Inzidenz eines Uterussarkoms nach Sectio caesarea erhöht ist. Diese Fragestellung geht von einer wissenschaftlichen Institution (z.B. der Universitätsklinik XYZ) aus. Diese Frage wird von dem Studienleiter für die Universität XYZ auf der vom NGD für die Einstellung von Fragestellungen potentieller NGD Nutzer erstellte, webunterstützten Plattform eingestellt. Dort sind Institutionen, Organisationen und Unternehmen der Gesundheitsversorgung wie z.B. das RKI, DIMDI, Krebsregister, KVen, Forschungsinstitute institutionell und auch mit einem strukturierten Inhaltsverzeichnis vorhandener Datensätze vertreten. Ihre zum Datenaustausch relevanten institutionellen Kontaktdaten sind eingestellt und eine Selbstbeschreibung ist für andere Teilnehmer wie z.B. in diesem Fall die Universitätsklinik XYZ sichtbar eingestellt. Als Grundlage für die Anfragestellung gibt die Universitätsklinik Lübeck ihr Aufgabenspektrum, welche Grundlage ist für den instituts-/ organisationsbezogenen Aufgabenbereich zur Datenerhebung und Datensammlung sowie die Kontaktdaten an.

Die Fragestellung der Universitätsklinik XYZ zur Inzidenzzunahme des Uterussarkoms nach Sectio caesarea ist für alle Netzwerknutzer auf der NGD-Plattform sichtbar, sodass diese anderen Netzwerknutzer ihre Bereitschaft gegenüber der Universitätsklinik XYZ signalisieren können, dass sie ihre Daten zur Datenverwendung bereitstellen würden (potentieller Datenbereitsteller). Hier kommen z.B. das DIMDI, Krankenversicherer, das RKI, epidemiologische Krebsregister in Frage. Europäische Institutionen können ebenfalls die Bereitschaft, Daten zur Verfügung zu stellen, signalisieren. Sowohl die Einstellung der Anfrage seitens der Universitätsklinik XYZ (als potentiellem Datennutzer) als auch die Bereitschaft seitens der anderen auf der Plattform des NGD vertretenen Institutionen, Daten zur Verfügung zu stellen, werden vom NGD registriert.

Der vom NGD erstellte Index bzw. das Register für die Datensätze der Universitätsklinik XYZ wird über das NGD und für alle NGD-Nutzer sichtbar auf der Plattform eingestellt.

Die Universitätsklinik XYZ, kann anhand des durch das NGD eingestellten Indexes/ Registers über die Datensätze anderer Institutionen, die ebenfalls NGD-Nutzer sind, erkennen, welche Institutionen als Datenbereitsteller für eine Lösung der Anfrage in Frage kommen.

Die Universität XYZ wird ebenfalls von der das NGD beaufsichtigenden Behörde verpflichtet. Das NGD nutzt in einer internen Analyse Prozesse zur Anfrage, zur Kontaktaufnahme (Zeit, Passgenauigkeit des Datenangebotes zur Fragestellung etc.) Prozessergebnisse zur Qualitätsentwicklung und stellt diese Ergebnisse den NGD-Nutzern zur Verfügung, um eine generelle Qualitätsverbesserung zu unterstützen. Die Universitätsklinik XYZ hat in einer Stichprobenerhebung Patientinnen, die mit der Diagnose eines Uterussarkoms zur Hysterektomie kamen, bei Geburten in der Vergangenheit zum Geburtsprozess (vaginale Geburt versus Sektio caesarea) befragt. Auf Grund der seltenen Inzidenz (2/100.000 Frauen/anno) des Tumors ist die Stichprobe zu klein, um eine signifikante Aussage treffen zu können. Es scheint eine erhöhte Inzidenz bei Hypertonie, Adipositas und Diabetes vorzuliegen. Eine exogene Östrogenzufuhr erhöht das Risiko, die Einnahme von Ovulationshemmern vom Kombinationstyp scheint mit einer Hemmung des Risikos einherzugehen. Diese Faktoren mit Wirkung auf die Inzidenzraten müssen bei der statistischen Kalkulation zur Größe der Untersuchungsgruppen berücksichtigt werden. Da das Ergebnis zur Fragestellung von Interesse ist, fragt die Universitätsklinik XYZ über das NGD Daten an. Interessant sind dabei die Datensätze des DIMDI, die Hinweis auf Begleiterkrankungen oder Medikamenteneinnahme geben können und klinische Krebsregisterdaten, die die Tumor-Art exakt benennen und die Daten zum Geburtsprozess, der bei diesen Patientinnen in der Regel bei einem mittleren Erkrankungsalter von 55-60 ca. 25-35 Jahre zurück liegt.

Eine Übermittlung der Datensätze selbst (weder des Interessenten noch des Angefragten) ist im ersten Schritt der Kontaktaufnahme nicht notwendig. Die Universitätsklinik Lübeck hat die statistische Kalkulation institutsintern durchführen lassen und hat auf Basis der Epidemiologie Kenntnis, welche Daten in welchem Umfang zur Beantwortung notwendig sind.

Diese Informationen gibt Universitätsklinikum XYZ entsprechend in grober Form über die webunterstützte Plattform an den möglichen Vertragspartner und in detaillierter Form an das NGD. Das NGD kann bei der Suche nach Datenbereitstellern (DB) wie z.B. nach anderen Klinika beratend und unterstützend tätig sein.

Über die Kontakt-Plattform des Netzwerk Gesundheitsdaten kann die Universitätsklinik XYZ eine gezielte Kontaktaufnahme mit den anderen Gesundheitsinstitutionen, -organisationen und -unternehmen (DIMDI, große GKV, klin. Krebsregister und Kliniken) durchführen, in der die zu untersuchende Fragestellung (s.o.) formuliert und der notwendige Datensatz aus der statistischen Kalkulation angefragt wird.

Intern für das NGD:

- Die über die Plattform gestellte Anfrage zum Uterussarkom in Korrelation zum Geburtsprozess kann durch das NGD zu verschiedenen Themengruppen geclustert (Organ, Diagnose, Bereich etc.) und entsprechend für eine interne Statistik ausgewertet werden.

## ii. Teilbereich 2 – Projektbeschreibung und Definition des benötigten Datensatzes

In diesem Beispiel stellt die Universität Lübeck seine Anfrage zur Datennutzung und zur Datenanalyse über die Plattform des NGD. Im ersten Schritt sind bi- / oder multilaterale Kontaktaufnahmen ohne die anderen organisatorischen Bereiche (Contracting und Neutrale Stelle) des NGD möglich.

Die konkrete Anfrage zur Assoziation von Uterussarkom und Geburtsprozess kann an die möglichen Vertragspartner (DIMDI, RKI, klinische Krebsregister, Kliniken) adressiert werden. Bei gegenseitigem Interesse wird dann der Bereich Contracting des NGD zur Vertragsformulierung dazu geschaltet. Eine Austausch-Symmetrie zwischen Universitätsklinikum XYZ und DIMDI, RKI, klinische Krebsregister und Kliniken ist für ein Zustandekommen des Projektes oder zur Vertragskonzeption nicht notwendig.

Die Anfrage der Universitätsklinik XYZ beschreibt die Fragestellung, die beforscht werden soll (s.o.). Sie definiert entsprechend auch den Inhalt (z.B. die Patientenmerkmale), die benötigten Datenarten (z.B. ICD-Kodes) und je nach Entwicklungsstadium der Anfrage die Menge der von einem oder mehreren anderen Plattformnutzern benötigten Daten und die Art der geplanten Auswertung (s.o.).

Zur Datenanfrage werden dafür vom NGD bereitgestellte Formular-Templates genutzt. Im Laufe der Aktivitäten des NGD werden diese Formular-Templates zur strukturierten Eingabe entsprechend der geclusterten fachlichen Schwerpunktsetzung weiterentwickelt und spezifiziert. Das Formular-Template enthält neben dem freien Eingabefeld für die Anfrage Prüfung der Assoziation zwischen Uterussarkomen nach Sectio Caesarea im Vergleich zu einem vaginalen Geburtsprozess folgende Rubriken:

Anfragendes Institut: Universitätsklinikum XYZ, Fachbereich: Gynäkologie/ Onkologie und Geburtshilfe, Diagnose: Uterussarkom und Sektio Caesarea bzw. vaginale Geburt, Stadium I-IV, Eingriff: PE bzw. Hysterektomie, Therapie: Hysterektomie ggf. Chemotherapie, Gender: Frauen mit Geburten, Altersgruppen: 25-40 und 55-65 Jahre, Studienart: Fall-Kontroll Studie, Studiendauer: X Monate, Patientenzahl: entsprechend statistischer Ermittlung, notwendige/s Institut/e für eine Datenbereitstellung: DIMDI bzw. große GKV, klein Krebsregister, Kliniken, Datenmenge: entsprechend statistischer Powerkalkulation unter Einbeziehung der möglichen Risikofaktoren.

Im Folgenden wird zur genaueren Prüfung über die organisatorische Einheit „Plattform des NGD“ eine interne Anfrage bzw. Schlagwortsuche an das Register bzw. den Index der durch die Plattformnutzer bereitgestellten Datenlisten gestellt.

Anhand der Anfrage der Universitätsklinik Lübeck über das ausgefüllte Template (s.o.) sucht der Bereich Kontakt-Plattform des NGD die passenden Vertragspartner als DB über den internen Datenkatalog

Wenn der/ die angefragte/n Plattformnutzer und potentielle DB (hier das DIMDI, klinische Krebsregister, Kliniken) Interesse an einer Annahme der Anfrage hat/haben, kann er/können sie das auf direkten Weg über die webbasierte Kontakt Plattform eingeben und sich dann durch den hergestellten Kontakt ausführlicher über die Plattform an den Anfrager wenden. Ein teilautomatisierter Prozess ist möglich.

Wenn die Universitätsklinik XYZ als potentielle DN und DIMDI bzw. eine große GKV, klin. Krebsregister und Kliniken als potentielle DB Interesse an einer passageren Datenzusammenführung und -Analyse haben, wendet sich die Universitätsklinik (DN) an das NGD mit der Bitte um Vertragsgestaltung.

Zusammengeführte Daten (pseudonymisiert/ personen-bezogen) werden in der Neutralen Stelle des NGD gemacht und dann für die Analyse pseudonymisiert (siehe TMF, Medizinkonsortium). Falls die Neutrale Stelle des NGD feststellt, dass die Analyse mit anonymisierten oder pseudonymisierten Daten gemacht werden kann, dann übermittelt sie den Datenbereitstellern den genauen Prozess der Anonymisierung/ Pseudonymisierung und die Analyse wird nur mit pseudonymisierten Daten gemacht. Falls dies nicht der Fall sein sollte und personenbezogene Daten für die Integration für die verschiedenen Datensätze notwendig sind, so muss der Datenbereinsteller auch diese Daten zusenden.

Die Anfrage der Universitätsklinik XYZ wird über das Netzwerk Gesundheitsdaten vor der Weiterleitung an die DB auf ihre Vereinbarkeit mit dem Zweck des Netzwerks auch auf die datenschutzrechtliche Zulässigkeit geprüft und in diesem Falle für unproblematisch erachtet. Eine abschließende Prüfung der Rechtsgrundlage der Auswertung sowie der Einhaltung weiterer datenschutzrechtlich relevanter Vorschriften erfolgt im Rahmen der Vertragsaushandlung (s.o.).

### iii. Teilbereich 3 – Projekt-Akzeptanz und Vertragsabschluss

Im weiteren Schritt erfolgt eine teil-automatisierte Vertragsaushandlung mit der Universitätsklinik XYZ (DN) und dem DIMDI, klinischen Krebsregistern und Kliniken (DB), sowie auch mit dem NGD, dass dort in der organisatorischen Einheit der Neutralen Stelle die gewünschte Analyse durchgeführt werden soll. Voraussetzung für den Vertragsabschluss ist die vorausgegangene Freigabe durch die zuständige Ethikkommission.

Das bedeutet auch, dass die rechtliche Basis der Datenzusammenführung durch das NGD geprüft wird. Hierbei sind auch die rechtliche Zweckbindung der Daten und die Beziehung zum Zweck der Auswertung zu prüfen. Im Ergebnis dieser Aushandlung ist die Zweckbindung der Auswertung mit den datenschutzrechtlichen Vorgaben einer Nutzung der einzubeziehenden Daten abgeglichen, die jeweiligen Rechte zur Verwertung der Analyseergebnisse durch alle teilnehmenden Nutzer des NGD

(DN und DB) fixiert und ein Vertragsentwurf entsprechend des Forschungsvorhabens und der dafür erwünschten und notwendigen, passager zusammenzuführenden Daten für die beteiligten Vertragspartner verfasst.

#### iv. Teilbereich 4 – Datenübertragung von Netzwerknutzern an Neutrale Stelle des NGD

Damit die maximale Sicherheit der Datenübertragung gewährleistet werden kann, wird dies durch die Einhaltung von zwei wichtigen Schritten zur Datenübertragung umgesetzt: Die Datenverschlüsselung wird durch ein Public und Private Key Verschlüsselungsverfahren gewährleistet, in dem die Neutrale Stelle ein Public/Private Key Paar erstellt und dann den Public Key der Universitätsklinik XYZ zur Verschlüsselung der Daten zusammen mit einer Beschreibung des Verschlüsselungsprozesses übergibt. Der Private Key bleibt hingegen bei der Neutralen Stelle, damit die Daten nur in der gesicherten Neutralen Stelle entschlüsselt werden können. Zusätzlich werden die verschlüsselten Daten auf einen Datenspeicher geschrieben und mit einem gesicherten analogen Transportmittel mit Personenschutz an die Neutrale Stelle übertragen. DIMDI, KVen, klin. Krebsregister und Kliniken müssen sicherstellen, dass eine genaue Kopie der gelieferten Daten im gesetzlich vorgeschriebenen Zeitraum bei sich vorgehalten werden.

#### v. Teilbereich 5 – Datenanalyse

Nach Datenzusammenführung der Datenpakete aus der Universität XYZ, dem DIMDI, KVen, den klinischen Krebsregistern und Kliniken in dem organisatorischen Bereich Neutrale Stelle des NGD, kann die Auswertung entsprechend des formulierten Forschungsvorhabens erfolgen.

Die Zusammenführung benötigt ferner eine Datenintegration. Diese geschieht in zwei Schritten: Zuerst teilautomatisiert durch die Benutzung gleicher Ontologien und Standards von den Datenbereitstellern, danach falls notwendig manuell durch die Medizinstatistiker zu Sicherstellung einer hohen Qualität. Daher ist bei der personellen Besetzung auf besonders qualifiziertes Personal mit hoher Eigenverantwortlichkeit und Selbstmotivation zu achten. Die Bedeutung der Datenauswertung mit Bezug auf den individuellen Schutz und in Perspektive auf die Entwicklung des Allgemeinwohls sollte durchweg präsent sein.

Die Neutrale Stelle führt die Datenpakete, die von den Vertragspartnern vertraglich beschrieben, ausgewählt, komprimiert und an die Neutrale Stelle übermittelt wurden, zusammen, speichert sie passager für die Datenanalyse und sichert sie entsprechend der vorgeschriebenen Kriterien, welche sich aus den datenschutz- und IT-sicherheitsrechtlichen Anforderungen ergeben. Sie hat die die Expertise zur Erstellung der Ontologie, Interoperabilität, Datenzusammenführung und der Datentransformation,

Dann folgt die statistische Analyse der in diesem Fall numerischen Daten und deren Auswertung entsprechend der Fragestellung und des Vertrages. Das Analyseergebnis wird durch die Neutrale Stelle des Netzwerk Gesundheitsdaten gegen Verfälschung gesichert und so mit Hinweisen auf die eingeflossenen Daten verknüpft, dass eine Nachvollziehbarkeit des Ergebnisses gegeben ist. Die Datensätze, die für die Analyse notwendig waren, werden im Nachgang der Analyse von den Nutzern des NGD (DN und DB) entsprechend der vertraglichen Vereinbarung aufbewahrt, um die Analyse bei Bedarf verifizieren zu können. Außerhalb der Neutralen Stelle des NGD findet keine Datenanalyse statt. Die Analyse erfolgt ausschließlich in der Neutralen Stelle des NGD.

#### vi. Teilbereich 6 – Ergebnisübergabe/Vertragserfüllung

Anschließend übersendet die Neutrale Stelle des NGD die gesicherten Analyseergebnisse über ein geeignetes und gesichertes Speichersystem gemäß den vertraglichen Vereinbarungen von Universitätsklinik XYZ, DIMDI, KVen, klin. Krebsregister und Kliniken an selbige. Dies geschieht so, dass sowohl die Analyseergebnisse als auch die Prozesse dauerhaft nachvollzogen werden können. Daten dürfen nur noch in aggregierter Form vorliegen. Die Analyseergebnisse werden vor Übermittlung durch technische und organisatorische Maßnahmen auf eine mögliche Personenbeziehbarkeit überprüft.

#### vii. Teilbereich 7 – Datenlöschung

Im letzten Schritt werden die von den Nutzern des NGD übermittelten Datensätze dauerhaft aus dem Speichersystem der Neutralen Stelle des Netzwerk Gesundheitsdaten gelöscht. Dies ist technisch, organisatorisch und rechtlich sichergestellt sowie vertraglich vereinbart und es kann jederzeit ein Beleg

durch die Datenbereitsteller und Datennutzer verlangt, sowie durch die Aufsichtführende Behörde geprüft werden, dass die Daten tatsächlich gelöscht wurden.

c) *Beispiel 1 B:*

*Zusammenführung von Daten zur Untersuchung, ob der eigene individuelle Geburtsprozess (vaginale Geburt versus Geburt durch eine Sectio caesarea des Kindes) zu einer Erhöhung des eigenen, individuellen Risikos an einer onkologischen Erkrankung zu erkranken, führt.*

- Zusammenführung von Daten aus dem Robert Koch-Institut, Versicherungsdaten und klinischen Krebsregisterdaten.

Die Datenanfrage, die Vertragsabschlüsse, Datenzusammenführung etc. (7 funktionale Schritte, Kap. 1) erfolgen entsprechend des Beispiels 1 a). Da hier die Korrelation einer erhöhten Inzidenz onkologischer Erkrankungen im Zusammenhang des individuellen, persönlichen Geburtsprozesses (Sektion caesarea vs. vaginale Geburt) geprüft wird, liegt bei einem mittleren Erkrankungsalter von Krebs eine zeitliche Spanne von ca. 60-70 Jahren zwischen den zu korrelierenden Ereignissen (Geburtsprozess und onkologische Diagnose). Da Gesundheitsdaten bei den Krankenversicherern derzeit nicht so lange vorgehalten werden müssen, lässt sich diese Fragestellung aus den bereits vorhandenen Daten nicht beantworten und es müsste in diesem Fall eine prospektive Studie durchgeführt werden.

2. *Beispiel 2*

a) *Problemstellung*

*Zusammenführung von Daten zur Untersuchung, ob es eine Assoziation von M. Parkinson und kühlen, sonnenarmen Klimazonen gibt.*

- Zusammenführung von Daten dem Robert Koch-Institut, Kliniken, DIMDI, bevölkerungsbezogenen Daten des statistischen Bundesamtes und von meteorologischen Daten (Meteorologische Institute, Deutscher Wetterdienst).

b) *Operative Funktionalitäten des „virtuellen Netzwerks Gesundheitsdaten“ im Bezug auf das 2. Beispiel:*

i. *Teilbereich 1 – Kontaktaufnahme über Plattform des NGD*

Das RKI plant den Zusammenhang von M. Parkinson und der Auswirkung der Wettersituation, insbesondere von Tageslicht zu untersuchen, da es bei M. Parkinson Veränderungen des Vitamin D Stoffwechsels gibt.

Diese Anfrage wird auf die Plattform des NGD eingestellt und andere neurologische Kliniken zur Studienteilnahme angesprochen. Benötigt werden Patienten, die neben einer M. Parkinson-Diagnose auch einen nachweislich veränderten Vitamin-D-Stoffwechsel haben. Über die Kontakt-Plattform des NGD wird außerdem angefragt, ob Daten über das DIMDI flächendeckend zur Verfügung stehen.

Zusätzlich wird das NGD gebeten, Daten des Statistischen Bundesamtes heranzuziehen, um den Bezug zur Wohnhaftigkeit der zu analysierenden Bevölkerungsgruppe ziehen zu können und auch die Daten Meteorologischer Institute und des Deutschen Wetterdienstes zu Sonnenstunden in den verschiedenen deutschen Regionen zu erhalten.

Intern für das NGD:

Auch hier werden die über die Plattform gestellten Anfragen durch das NGD zu verschiedenen Themengruppen (Diagnose, Stoffwechsel, Wohnhaftigkeit nach Regionen, Wetter, Prävention) geclustert und entsprechend ausgewertet werden.

ii. *Teilbereich 2 – Projektbeschreibung und Definition des benötigten Datensatzes*

Wissenschaftliche Studien zeigen, dass bei M. Parkinson gehäuft der Vitamin-D-Spiegel bei Patienten gemindert ist und legen eine Assoziation nahe. Da der individuelle Vitamin-D-Spiegel u.a. von Sonnenlichtexposition abhängig ist, will, um in nachgeordneten Schritten präventive Maßnahmen wie eine Vitamin-D-Prophylaxe in sonnenärmeren Regionen einführen zu können, das RKI prüfen, ob es

einen epidemiologischen Zusammenhang von M. Parkinson zum Leben in sonnenärmeren Regionen gibt.

Die Festlegung des Datensatzes (Menge und Inhalte) erfolgen über den Statistiker im eigenen Institut.

Daten von Patienten mit M. Parkinson Diagnose und Vitamin-D-Spiegel Bestimmung aus neurologischen Kliniken und aus dem DIMDI bei ggf. fehlenden Vitamin-D-Spiegel Bestimmungen werden angefragt. Diese Patientendatensätze werden mit Postleitzahlen angefragt, um den Regionen-Bezug herstellen zu können.

Zusätzlich wird das NGD gebeten, Daten des Statistischen Bundesamtes heranzuziehen, um den Bezug zur Wohnhaftigkeit der zu analysierenden Bevölkerungsgruppe ziehen zu können und auch die Daten der Meteorologischen Institute und des Deutschen Wetterdienstes zu Sonnenstunden pro Tag und Region in den verschiedenen deutschen Regionen zu erhalten.

Zur Datenanfrage werden die vom NGD dafür bereitgestellte Formular-Templates genutzt. Im Laufe der Aktivitäten des NGD werden diese Formular-Templates zur strukturierten Eingabe entsprechend der geclusterten fachlichen Schwerpunktsetzung weiterentwickelt und spezifiziert. Das Formular-Template enthält neben dem freien Eingabefeld für die Anfrage (z.B. hier siehe die ausgewählten Beispiele) folgende Rubriken:

- Anfragendes Institut; RKI, Fachbereich: Neurologie, Allgemein-/ Innere Medizin, Diagnose: M. Parkinson, Vit. D Mangel, Stadium, Therapie: M. Parkinson Therapie und Vit. D Substitution, Gender: m und w, Altersgruppen: 60-80, Studienart: Fall-Kontroll, Epidemiologisch, Studiendauer: X Mon, Patientenzahl: nach Statistik, notwendige/s Institut/e für eine Datenbereitstellung: Robert Koch-Institut, Kliniken, DIMDI, statistisches Bundesamt, Meteorologische Institute, Deutscher Wetterdienst), Datenmenge: Statistik.

Das Formular-Template wird soweit wie möglich strukturiert beschrieben mit Hilfe existierender Ontologien und Standards. Diese werden auch iterativ verbessert durch die gewonnenen Erfahrungen aus zuvor ausgeführten Studien. Zur genaueren Prüfung wird über die organisatorische Einheit Plattform des NGD eine interne Anfrage bzw. Schlagwortsuche an das Register bzw. den Index der durch die Plattformnutzer bereitgestellten Datenlisten gestellt. Diese Datenlisten sind in Form eines Datenkataloges angelegt.

Anhand der Anfrage über das ausgefüllte Template sucht der Bereich Kontakt-Plattform des NGD Vertragspartner i.e. Datenbereinsteller über den internen Datenkatalog. Der intern angelegte Datenkatalog wird gefüllt durch die Plattformnutzer (DN und DB). Entlang des intern des NGD angelegten Datenkataloges wird auch in diesem Fall das auf der webbasierten Plattform sichtbare Datenregister/Index weiterentwickelt. Das RKI kann die Anfrage auf direktem Weg über die webbasierte Kontakt Plattform eingeben und sich dann durch den hergestellten Kontakt ausführlicher über die Plattform an den Anfrager wenden. Ein teilautomatisierter Prozess ist möglich.

Wenn die Anfrager als potentielle DN (hier RKI) und die Adressierten als potentielle DB (hier Kliniken, DIMDI, Met. Inst., DWD, StBA), Interesse an einer passageren Datenzusammenführung und -Analyse haben, wendet sich das RKI an das NGD mit der Bitte um Vertragsgestaltung.

Das Institut wird im nächsten Schritt von der Neutralen Stelle des NGD angefragt zu prüfen, ob der spezifische notwendige Datensatz vorhanden ist. Das NGD kann bei Bedarf weitere Institute/Organisationen zur Ergänzung des benötigten Datensatzes hinzuziehen.

Datensatz-Formulierung: Der benötigte Datensatz wird vom anfragstellenden DN selber formuliert, diese Anfrage geht ans NGD. Das NGD kennt den gesamten Datenindex und sucht die ergänzenden Institute heraus. Der DB muss die geforderten Daten selektieren soweit wie möglich, sie in die gewünschten Standards überführen und dem NGD übermitteln zusammen mit einer Beschreibung des Inhaltes und der Struktur der Daten. Für diesen Mehraufwand erhält der DB eine Vergütung.

Zusammengeführte Daten (pseudonymisiert/ personen-bezogen) werden in der Neutralen Stelle des NGD gematcht und dann für die Analyse pseudonymisiert (siehe TMF, Medizinkonsortium). Falls die Neutrale Stelle feststellt, dass die Analyse mit anonymisierten oder pseudonymisierten Daten

gemacht werden kann, dann übermittelt sie den Datenbereitstellern den genauen Prozess der Anonymisierung/ Pseudonymisierung und die Analyse wird nur mit pseudonymisierten Daten gemacht. Falls dies nicht der Fall sein sollte und personenbezogene Daten für die Integration für die verschiedenen Datensätze notwendig sind, so muss der Datenbereinsteller auch diese Daten zusenden.

Auch hier wird die Anfrage bestimmter Daten über das NGD vor der Weiterleitung an Datenbereinsteller auf ihre Vereinbarkeit mit dem Zweck des Netzwerks geprüft, soweit möglich auch bereits die datenschutzrechtliche Zulässigkeit. Eine abschließende Prüfung der Rechtsgrundlage der Auswertung sowie der Einhaltung weiterer datenschutzrechtlich relevanter Vorschriften erfolgt erst nach der Aushandlung.

### iii. Teilbereich 3 – Projekt-Akzeptanz und Vertragsabschluss

Im weiteren Schritt erfolgt eine teil-automatisierte Vertragsaushandlung mit den Einrichtungen, die Daten für ihre Anfrage nützen werden (RKI) und Einrichtungen, die Daten für die Anfrage bereitstellen werden (DB) (Kliniken, DIMDI, Met. Instit, DWD, Stat. Bundesamt), sowie auch mit dem NGD, dass dort in der organisatorischen Einheit der Neutralen Stelle die gewünschte Analyse durchgeführt werden soll. Voraussetzung für den Vertragsabschluss ist die vorausgegangene Freigabe durch die zuständige Ethikkommission.

Der Vertragsabschluss erfolgt hier wie auch in Beispiel 1 entsprechend des im 1. Kapitel beschriebenen Procedere.

### iv. Teilbereich 4 – Datenübertragung

Damit die maximale Sicherheit gewährleistet werden kann, wird die Datenübertragung von RKI, Kliniken, DIMDI, bevölkerungsbezogenen Daten des StBA und von meteorologischen Daten (Meteorologische Institute, DWD) an die Neutrale Stelle des NGD durch die Einhaltung von zwei wichtigen Schritten zur Datenübertragung umgesetzt: Die Datenverschlüsselung wird durch ein Public und Private Key-Verschlüsselungsverfahren gewährleistet, in dem die Neutrale Stelle ein Public/Private Key-Paar erstellt und dann den Public Key der Universitätsklinik Lübeck zur Verschlüsselung der Daten zusammen mit einer Beschreibung des Verschlüsselungsprozesses übergibt. Der Private Key bleibt hingegen bei der Neutralen Stelle, damit die Daten nur in der gesicherten Neutralen Stelle entschlüsselt werden können. Zusätzlich werden die verschlüsselten Daten auf einen Datenspeicher geschrieben und mit einem gesicherten analogen Transportmittel mit Personenschutz an die Neutrale Stelle übertragen. DIMDI, klein. Krebsregister und Kliniken müssen sicherstellen, dass eine genaue Kopie der gelieferten Daten im gesetzlich vorgeschriebenen Zeitraum bei sich vorgehalten werden.

### v. Teilbereich 5 – Datenanalyse

Nach Datenzusammenführung der Datenpakete aus Robert Koch-Institut, Kliniken, DIMDI, bevölkerungsbezogenen Daten des statistischen Bundesamtes und von meteorologischen Daten (Meteorologische Institute, Deutscher Wetterdienst) in dem organisatorischen Bereich Neutrale Stelle des NGD, kann die Auswertung entsprechend des formulierten Forschungsvorhabens erfolgen.

Die Zusammenführung benötigt ferner eine Datenintegration. Diese geschieht in zwei Schritten: Zuerst teilautomatisiert durch die Benutzung gleicher Ontologien und Standards von den Datenbereitstellern, danach falls notwendig manuell durch die Medizinstatistiker zu Sicherstellung einer hohen Qualität.

Die Neutrale Stelle führt die Datenpakete, die von den Vertragspartnern vertraglich beschrieben, ausgewählt und an die Neutrale Stelle übermittelt wurden, zusammen, speichert sie passager für die Datenanalyse und sichert sie entsprechend der vorgeschriebenen Kriterien, welche sich aus den datenschutz- und IT-sicherheitsrechtlichen Anforderungen ergeben.

Dann folgt die statistische Analyse der in diesem Fall numerischen Daten und deren Auswertung entsprechend der Fragestellung und des Vertrages. Das Analyseergebnis wird durch die Neutrale Stelle des Netzwerk Gesundheitsdaten gegen Verfälschung gesichert und so mit Hinweisen auf die eingeflossenen Daten verknüpft, dass eine Nachvollziehbarkeit des Ergebnisses gegeben ist. Die Datensätze, die für die Analyse notwendig waren, werden im Nachgang der Analyse von den Nutzern des NGD (DN und DB) entsprechend der vertraglichen Vereinbarung aufbewahrt, um die Analyse bei

Bedarf verifizieren zu können. Außerhalb der Neutralen Stelle des NGD findet keine Datenanalyse statt. Die Analyse erfolgt ausschließlich in der Neutralen Stelle des NGD durch Medizinstatistiker. Auswertungen sind nur zu dem vertraglich gesicherten Fall möglich.

#### vi. Teilbereich 6 – Ergebnisübergabe/ Vertragserfüllung

Anschließend übersendet die Neutrale Stelle des NGD die gesicherten Analyseergebnisse über ein geeignetes und gesichertes Speichersystem gemäß der vertraglichen Vereinbarungen von Robert Koch-Institut, Kliniken, DIMDI, bevölkerungsbezogenen Daten des Statistischen Bundesamtes und von meteorologischen Daten (Meteorologische Institute, Deutscher Wetterdienst) an die Vertragspartner. Dies geschieht so, dass sowohl die Analyseergebnisse als auch die Analyseprozesse dauerhaft nachvollzogen werden können.

#### vii. Teilbereich 7 – Datenlöschung

Im letzten Schritt werden von die Robert Koch Institut, Kliniken, DIMDI, statistischem Bundesamt und den Meteorologischen Instituten/ dem Deutschen Wetterdienst übermittelten Datensätze dauerhaft aus dem Speichersystem der Neutralen Stelle des Netzwerk Gesundheitsdaten gelöscht. Dies ist technisch, organisatorisch und rechtlich sichergestellt sowie vertraglich vereinbart und es kann jederzeit ein Beleg durch die Datenbereitsteller und Datennutzer verlangt, sowie durch die Aufsicht führende Behörde geprüft werden, dass die Daten tatsächlich gelöscht wurden.

## B. Zwingende rechtliche Anforderungen, insbesondere des Datenschutzes

### 1. Rechtlich-organisatorische Grundlagen

Dem Modell liegt der Gedanke eines rechtlich verfassten virtuellen Netzwerks von Datenbereitstellern (DB) und Datennutzern (DN) im Gesundheitswesen zu Grunde („Virtuelles Netzwerk Gesundheitsdaten“, NGD), das zu Zwecken der Gesundheitsforschung und Gesundheitsversorgung beliebige medizin-fachliche Auswertungen personenbezogener und nicht personenbezogener Daten auf grundsätzlich aller bei den Datenbereitstellern verfügbaren Datenbestände ermöglicht. Organisatorisch umfasst das NGD eine digitale Plattform zur Kontaktaufnahme, den Bereich für die Vertragsabschlüsse und den Bereich für die Datenzusammenführung und Datenanalyse (Neutrale Stelle). Das NGD schafft die rechtlichen, organisatorischen, fachlichen und technischen Voraussetzungen, damit die Auswertungen stattfinden können und die fachlichen, vertraglichen sowie gesetzlichen Vorgaben eingehalten werden. Kernfunktionen des NGD sind:

- RAHMEN: Schaffung des übergeordneten rechtlichen, organisatorischen, fachlichen und technischen Rahmens für das Netzwerk
- NUTZER-MANAGEMENT: Incentivierung und Zulassung von Datennutzern zum Netzwerk, Ausschluss von Nutzern, Schaffung der Rahmenbedingungen für die Beteiligung an Auswertungen
- AUSWERTUNGS-KONFORMITÄT: Prüfung eines Auswertungsvorhabens auf Vereinbarkeit mit Regeln und Gesetzen, Sicherstellung von datenschutzrechtlicher Konformität und Unverfälschbarkeit der Auswertungsergebnisse, Sicherstellung von Löschung der Daten nach Auswertung, Dokumentation des Prozesses
- AUSWERTUNG: Durchführung von Auswertungen durch die Neutrale Stelle

Im Hinblick auf die datenschutzrechtlichen Anforderungen wird das datenschutzrechtliche Modell auf Basis folgender Prämissen bzw. Einschränkungen erstellt:

1. Das Modell soll auf Basis geltenden europäischen Rechts umgesetzt werden können. Dabei sollen Änderungen im nationalen Recht, insbesondere die Nutzung von Öffnungsklauseln der DSGVO für das Gesundheitswesen, möglich sein.
2. Alle Auswertungen sollen grundsätzlich einwilligungsfrei möglich sein, also auf Grundlage der DSGVO oder eine europarechtlich zulässige nationalen Rechtsgrundlage.

3. Das Modell soll für Fälle, in denen das nicht möglich ist, die Einholung von Einwilligungen berücksichtigen.
  4. Das Modell soll Auswertungen sowohl auf personenbezogenen Daten als auch auf pseudonymisierten und anonymisierten Daten berücksichtigen und unterstützen.
  5. Ergebnisse von Auswertungen sollen nur anonymisierte, statistische Daten sein, keine personenbezogenen oder pseudonymisierten Daten.
2. Datenschutzrechtliche Zulässigkeit der Arbeit des NGD
- a) *Grundsatz der Rechtmäßigkeit der Verarbeitung von Gesundheitsdaten durch das NGD und die Nutzer*

Art. 5 Abs. 1 lit. a DSGVO normiert den Grundsatz der Rechtmäßigkeit. Demnach ist für jeden Datenverarbeitungsvorgang eine Rechtsgrundlage erforderlich ist (vgl. auch Erwägungsgrund (40) DSGVO). Die aus diesem Datenschutzgrundsatz resultierende Gesetzesstruktur wird im nationalen Recht als „*Verbot mit Erlaubnisvorbehalt*“ bezeichnet.

Der Grundsatz der Rechtmäßigkeit findet sich vor allem in den Generalklauseln der Art. 6 und 9 DSGVO wieder. Eine Verarbeitung personenbezogener Daten ist grundsätzlich verboten bzw. nur dann zulässig ist, wenn sie auf einen der dort abschließend aufgelisteten Erlaubnistatbestände gestützt werden kann. Weitere Erlaubnistatbestände zur Verarbeitung personenbezogener Daten können sich ferner aus dem nationalen Recht (bzw. dem Recht der Mitgliedstaaten) ergeben, sofern die DSGVO hierfür in den unterschiedlichen Fällen Öffnungsklauseln vorsieht.

b) *Verantwortlichkeit für die Datenverarbeitung im NGD*

Der Grundsatz der Rechtmäßigkeit gilt für alle Verantwortlichen. Verantwortlicher i.S.d. DSGVO kann jede natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle sein, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet (vgl. Art. 4 Nr. 7 DSGVO).

Davon abzugrenzen sind die Verarbeitung im Auftrag (vgl. Art. 28 und 29 DSGVO) und die gemeinsame Verantwortung (vgl. Art. 26 DSGVO). Wichtigstes Abgrenzungsmerkmal ist dabei die Entscheidungsbefugnis über den Zweck der Verarbeitung von personenbezogenen Daten und somit die Entscheidung über das Ob, Wofür und Wieweit einer Datenverarbeitung. Insofern sind die Weisungsgebundenheit und auch der Umfang und das Recht der Beaufsichtigung bzw. der Kontrollen von großer Bedeutung. Aufgrund des komplexen Verarbeitungsumfelds, muss das NGD eine rechtsfähige Einrichtung sein, die in der Verantwortung des Bundesministeriums für Gesundheit liegt, von keinem DB oder DN beherrscht wird und einer öffentlich verfassten Kontrolle unterliegt. Somit wird unter diesen Kautelen die letzte Entscheidungsbefugnis über das Ob, Wofür und Wieweit der Verarbeitung von Gesundheitsdaten beim NGD liegen. Aufgrund der Komplexität ist es zudem wichtig, dass die Rollen und Verantwortlichkeiten im NGD leicht zuzuordnen sind, um sicherzustellen, dass eine unpraktikable Verteilung der Verantwortung nicht dazu führt, dass die Wirksamkeit des Datenschutzrechts beeinträchtigt wird.<sup>1</sup> Verantwortlicher im Sinne der DSGVO ist insofern sowohl der DB (dies kann bei Datenbereitstellung gleichzeitig auch der DN selbst sein) als auch das NGD.

Sofern der Datennutzer selbst keine Daten bereitstellt, ist er im Kontext des NGD kein Verantwortlicher. Dies ergibt sich dann insbesondere aus dem Umstand, dass der DN zu keinem Zeitpunkt personenbezogenen Daten erhält, in keiner sonstigen Art und Weise Entscheidungsgewalt über die Verarbeitung personenbezogener Daten der DB besitzt und allein der DB einen angefragten Datensatz zur Verarbeitung freigeben kann.

<sup>1</sup> vgl. *Hartung* in Kühling/Buchner (Hrsg.), DSGVO, 2. Aufl. 2018, Art. 4 Nr. 7 Rn. 13; Arbeitspapier 169 der Art.-29-Datenschutzgruppe, S. 21 ff.

c) *Anforderung an die Festlegung der Verarbeitungszwecke*

i. *Grundsatz der Zweckbindung (Art. 5 Abs. 1 lit. b DSGVO)*

Gem. Art. 5 Abs. 1 lit. b DSGVO gilt der Grundsatz der strengen Zweckbindung. Demnach dürfen personenbezogene Daten grundsätzlich nicht für andere als die bei der Erhebung festgelegten Zwecke weiterverarbeitet werden. Demnach entscheidet der verfolgte Zweck einer Verarbeitung personenbezogener Daten und das Verhältnis zum ursprünglichen Zweck über deren Zulässigkeit. Die Erlaubnistatbestände des Art. 9 DSGVO knüpfen implizit jeweils an die festgelegten, eindeutigen und legitimen Verarbeitungszwecke an.

Grundsätzlich müssen die Verarbeitungszwecke bereits zum Zeitpunkt der Erhebung der personenbezogenen Daten feststehen (Erwägungsgrund (39) S. 6 DSGVO). Dies wird im Falle der Verarbeitung personenbezogener Daten im NGD i.d.R. nicht der Fall sein. Allerdings sieht Art. 5 Abs. 1 lit. b Hs. 2 DSGVO eine Ausnahme vor, bei der die Weiterverarbeitung von Daten für wissenschaftliche Forschungszwecke oder für statistische Zwecke gemäß Art. 89 Abs. 1 DSGVO nicht als unvereinbar mit den ursprünglichen Zwecken gilt. Unter den Voraussetzungen der Zulässigkeit der Verarbeitung der Daten gemäß Art. 9 Abs. 2 lit. j DSGVO, zu wissenschaftlichen Forschungszwecken oder zu statistischen Zwecken, ist also in der Regel eine Vereinbarung der Zwecke anzunehmen. Da diese Zwecke bei der Weiterverarbeitung kompatibel mit dem Zweck der Erstverarbeitung sind, können sich die DB und das NGD als Rechtsgrundlage erneut auf die Rechtsgrundlage stützen, die bereits für die Erstverarbeitung der Nutzer galt.

Es ist darüber hinaus erforderlich, dass der festgelegte Zweck „eindeutig“ ist. Unklare Umschreibungen des Verarbeitungszwecks aber auch sprachlich eindeutige Verarbeitungszwecke, die jedoch eine weite Auslegung zulassen und somit im Ergebnis eine Verarbeitung zu vielen unterschiedlichen Zwecken ermöglichen, sind rechtswidrig. Bei der Ausgestaltung des NGD wird es darauf ankommen, dass eine präzise und nachvollziehbare Definition der Zwecke der Auswertungen erfolgt.

Zusätzlich muss die Verarbeitung der personenbezogenen Daten zur Erreichung des verfolgten Zwecks „erforderlich“ sein, sodass immer auch eine Bewertung der Zweck-Mittel-Relation notwendig ist. Hier knüpft die weitere Voraussetzung des Grundsatzes der Datenminimierung nach Art. 5 Abs. 1 lit. c DSGVO an, wonach die verarbeiteten personenbezogenen Daten für die verfolgten Zwecke, zu denen sie verarbeitet werden, „angemessen“ und „erheblich“ sowie auf das für die Zwecke ihrer Verarbeitung notwendige Maß beschränkt sein müssen (vgl. Erwägungsgrund (39) S. 7 DSGVO). Das Gebot der Datenminimierung verlangt sowohl die Anzahl der verarbeiteten Daten („Angemessenheit“), als auch den Umfang der Datennutzung („Erheblichkeit“), auf ein Minimum zu reduzieren. Der Grundsatz der Datenminimierung stellt somit einen Unterfall des Grundsatzes der Zweckbindung dar (vgl. Erwägungsgrund (39) S. 8 DSGVO).

Die Erforderlichkeit der Daten für eine bestimmte Auswertung sowie die Einhaltung des Grundsatzes der Datenminimierung sind vom NGD vor Durchführung einer Auswertung zu prüfen.

ii. *Sonderregelung für Forschungszwecke und statistische Zwecke (Art. 89 Abs. 1 DSGVO)*

Der strenge Zweckbindungsgrundsatz wird für die Verarbeitung zu statistischen Zwecken und Forschungszwecken (Art. 9 Abs. 2 lit. j DSGVO) unter Verweis auf die Erforderlichkeit gem. Art. 89 Abs. 1 DSGVO aufgeweicht. Art. 89 Abs. 1 DSGVO stellt keinen eigenständigen Erlaubnistatbestand dar, sondern regelt die Anforderungen, zu denen eine Weiterverarbeitung von Daten durch das NGD und deren Nutzer für wissenschaftliche Forschungszwecke oder für statistische Zwecke als mit den ursprünglichen Zwecken vereinbar gilt.

Dabei beziehen sich die Voraussetzungen des Art. 89 Abs. 1 DSGVO insbesondere auf die Erbringung von Garantien, die die Sicherheit der Daten gewährleisten und die Einhaltung des Grundsatzes der Datenminimierung sicherstellen. Hierfür sind vom NGD entsprechende technische und organisatorische Maßnahmen zu implementieren.

### iii. Übergeordneter Verarbeitungszwecke im NGD

Das NGD dient dem übergeordneten Zweck der Förderung des öffentlichen Gesundheitswesens i.S.d. der im öffentlichen Interesse liegenden Gesundheitsversorgung, Versorgungs- und allgemeinen Gesundheitsforschung. Sowohl öffentliche Betriebe, Anstalten etc. als auch privatrechtlich organisierte Unternehmen (z.B. öffentliche/private Krankenhäuser, Versicherungen, Register, Pharmaunternehmen etc.) können unter den vorgenannten übergeordnete Zwecken statistische Auswertungen durchführen und wissenschaftliche Forschung betreiben, sofern diese der Förderung der Gesundheitsversorgung, der Versorgungsforschung sowie der allgemeinen Gesundheitsforschung dienen.

Im Folgenden wird zur Klarstellung des Begriffs der Versorgungsforschung auf das Papier „Definition und Abgrenzung Versorgungsforschung“ des Arbeitskreises „Versorgungsforschung“ beim Wissenschaftlichen Beirat der Bundesärztekammer verwiesen, welches verabschiedet und am 08.09.2004 veröffentlicht wurde:<sup>2</sup>

„Der Begriff der Gesundheitsversorgung umschließt alle Formen sowohl Individuums-bezogener als auch populationsbezogener Prävention und Gesundheitsförderung, die durch medizinische und nicht-medizinische Gesundheitseinrichtungen und -fachkräfte erbracht werden (Institutions- und Professionsprinzip); dies schließt auch die Förderung der Selbstkompetenz und Selbstversorgung ein.“

„Uneingeschränkt zur Versorgungsforschung zählen:

- die Grundlagenforschung in versorgungsrelevanten Bereichen (z.B. kognitive und kommunikative Prozesse, Compliance, Arzteinstellungen, Determinanten von Arztentscheidungen)
- Untersuchungen zur Arbeitszufriedenheit und zu den Arbeitsbedingungen der Ärzte und der anderen Gesundheitsberufe
- Untersuchungen zur Arzt-Patient-Beziehung
- Qualitätsforschung (z.B. Qualitätssicherung, -entwicklung, -management) zu Versorgungsleistungen
- Untersuchungen Ökonomischer Aspekte der Versorgung
- Inanspruchnahmeforschung, Bedarfsforschung, sonstige Untersuchungen zur Versorgungsepidemiologie
- Methodische Entwicklungen im Bereich der Versorgungsforschung (z.B. risk adjustment, multivariate Modelle, Erhebungsinstrumente)
- klinische Studien, Metaanalysen und systematische Reviews zur effectiveness
- HTA, sofern sie effectiveness untersucht
- Arzneimittel- und Medizinprodukte-Studien unter Alltagsbedingungen (sog. Phase IV-Studien)

Bedingt zur Versorgungsforschung zählen, abhängig davon, ob ein Versorgungsbezug gegeben ist, u.a. die Fächer:

- Public Health
- Gesundheitssystemforschung
- Gesundheitsökonomie
- Organisationsforschung

Nicht zur Versorgungsforschung zählen:

- HTA, Metaanalyse und systematische Reviews, sofern diese nur efficacy untersuchen
- klinische Studien zur efficacy (die also unter Idealbedingungen durchgeführt werden)
- Arzneimittelstudien der Phasen I-III
- *Studien, die nicht explizit gesundheitsbezogene Dienstleister untersuchen, die analytische Epidemiologie zur Ursachen- und Risikoermittlung, die sich nicht auf spezifische Versorgungsformen bezieht.*

Im Folgenden wird zur Klarstellung des Begriffs der Gesundheitsforschung auf die gemeinsame Definition der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. und der Gesellschaft für Datenschutz und Datensicherheit e.V. verwiesen:<sup>3</sup>

„[Der allgemeine Begriff der Gesundheitsforschung] umschließt die systematische Suche nach neuen Erkenntnissen sowie deren Dokumentation und Veröffentlichung, wobei Suche sowohl im Bereich der Grundlagenforschung als auch der angewandten Forschung erfolgen kann. Die Ergebnisse der Suche müssen darauf abzielen, dass die Erkenntnisse

<sup>2</sup> vgl. Bundesärztekammer, Arbeitskreis Versorgungsforschung – Definition Gesundheitsversorgung, S. 1 ff., Link: <http://www.versorgungsforschung-deutschland.de/Definition.pdf>.

<sup>3</sup> vgl. Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO), Link: [http://ds-gvo.gesundheitsdatenschutz.org/download/forschung\\_ds-gvo.pdf](http://ds-gvo.gesundheitsdatenschutz.org/download/forschung_ds-gvo.pdf).

- a) dem öffentlichen Interesse im Bereich der öffentlichen Gesundheit dienen oder
- b) der Verbesserung der Lebensqualität zahlreicher Menschen oder der Verbesserung der Effizienz der Sozialdienste dienen oder
- c) der klinischen Prüfung therapeutischer Maßnahmen dienen oder
- d) der Registerforschung dienen.

Die privat finanzierte Forschung ist dabei der öffentlichen Forschung gleichgestellt.“

d) *Gesetzlicher Erlaubnistatbestände zur Verarbeitung von Gesundheitsdaten und genetischen Daten (Art. 9 DSGVO)*

i. *Zulässigkeit der Verarbeitung von besonderen Kategorien personenbezogener Daten*

Die vom NGD zu verarbeitenden Gesundheitsdaten gelten als „*besondere Kategorien personenbezogener Daten*“ i.S.d. Art. 9 Abs. 1 DSGVO. Art. 9 DSGVO normiert besondere restriktive Zulässigkeitsvoraussetzungen für die Verarbeitung solcher Daten, da die Verarbeitung dieser Arten von Daten ihrem Wesen nach besonders sensibel sind und erhebliche Risiken für die Grundrechte und Grundfreiheiten der Betroffenen auftreten können (vgl. Erwägungsgrund (51) S. 1 DSGVO). Art. 9 Abs. 1 DSGVO definiert diejenigen Datenkategorien, die als besonders schützenswert einzustufen sind. Zu den besonderen Kategorien personenbezogener Daten zählen Gesundheitsdaten sowie auch genetische Daten die zur eindeutigen Identifizierung einer natürlichen Person führen können. Im Folgenden soll aus Gründen der Lesbarkeit der Begriff Gesundheitsdaten ebenfalls die genetischen Daten umfassen. Sofern eine Unterscheidung bei der Prüfung der Machbarkeit notwendig erscheint, wird auf die Anforderungen der entsprechenden Datenkategorie gesondert eingegangen.

Die Verarbeitung der besonderen Kategorien von Daten ist grundsätzlich untersagt. Art. 9 Abs. 2 DSGVO normiert besondere gesetzliche Erlaubnistatbestände, nach denen die Verarbeitung dieser Daten rechtmäßig ist. Dabei enthält der Abs. 2 sogenannte Öffnungsklauseln, die es den Mitgliedsstaaten gestatten abweichende Regelungen im nationalen Recht zu treffen.

Für die Auslegung der Öffnungsklauseln bestehen zusätzlich zahlreiche Möglichkeiten der Konkretisierung über spezifisches nationalstaatliches Recht, da die Erlaubnistatbestände des Abs. 2 eine Vielzahl unbestimmter Rechtsbegriffe enthalten, die sich typischerweise an der jeweiligen Rechtskultur des Mitgliedstaates orientieren werden.

ii. *Gesetzliche Erlaubnistatbestände für die Verarbeitung von Gesundheitsdaten und genetischen Daten im europäischen Recht*

Mögliche gesetzliche Erlaubnistatbestände für die Verarbeitung von Gesundheitsdaten im Rahmen des NGD ergeben sich aus Art. 9 Abs. 2 lit. h bis j DSGVO. Danach ist die Verarbeitung von Gesundheitsdaten zu den folgenden übergeordneten Zwecken erlaubt:

- Gesundheitsfürsorge und Arbeitsmedizin (Art. 9 Abs. 2 lit. h DSGVO);
- Öffentlichen Gesundheit, schwerwiegende Gesundheitsgefahren (Art. 9 Abs. 2 lit. i DSGVO);
- Archiv-, Forschungs- und statistische Zwecke (Art. 9 Abs. 2 lit. j DSGVO).

Vor dem Hintergrund der strengen Zweckbindung und den vom NGD verfolgten Zwecken, entspricht das NGD den Voraussetzungen des Art. 9 Abs. 2 lit. h, i, j jeweils in folgendem Maße:

(1.) *Gesundheitsfürsorge und Arbeitsmedizin (Art. 9 Abs. 2 lit. h DSGVO)*

Art. 9 Abs. 2 lit. h DSGVO ist auf die Verarbeitung von Gesundheitsdaten im Rahmen der Versorgung und Behandlung im Gesundheits- und Sozialbereich gerichtet und berücksichtigt den Umstand und die Notwendigkeit, dass Gesundheitsdaten in großem Umfang verarbeitet und zwischen den verschiedenen Akteuren ausgetauscht werden. Die Verarbeitung von Gesundheitsdaten ist dabei zulässig, wenn es für das Erreichen gesundheitsbezogener Zwecke im Interesse einzelner natürlicher Personen und der Gesellschaft insgesamt erforderlich ist, insbesondere im Zusammenhang mit der Verwaltung der Dienste und Systeme des Gesundheits- oder Sozialbereichs (vgl. Erwägungsgrund (53) S. 1 DSGVO).

Art. 9 Abs. 3 DSGVO schränkt den Erlaubnistatbestand des Art. 9 Abs. 2 lit. h DSGVO insoweit ein, dass die Verarbeitung von Gesundheitsdaten zu den in lit. h genannten Zwecken ausschließlich Fachpersonal oder andere Personen gestattet ist, die nach dem Unionsrecht oder dem Recht eines

Mitgliedstaats oder den Vorschriften nationaler zuständiger Stellen dem Berufsgeheimnis unterliegt (vgl. § 203 SGB und §§ 53, 53a StPO).

Dieser Erlaubnistatbestand setzt zudem eine konkretisierende Rechtsgrundlage im Unionsrecht oder dem Recht einzelner Mitgliedstaaten oder einen Behandlungsvertrag voraus

Auswertungen des NGD erfolgen auch zum Zwecke der Gesundheitsversorgung. Der Erlaubnistatbestand des Art. 9 Abs. 2 lit. h DSGVO bezieht sich allerdings primär auf die Verarbeitung von Gesundheitsdaten im Zusammenhang mit der Verwaltung der Dienste und Systeme des Gesundheits- oder Sozialbereichs und somit primär auf infrastrukturelle bzw. systemische Aspekte des Gesundheitswesens. Unter Berücksichtigung und Abgrenzung des lit. h zu lit. i und j ist nicht auszuschließen, dass dieser zumindest einen Teil der medizinischen Forschung und Entwicklung, wie es das Modell vorsieht, umfasst (vgl. Erwägungsgrund (53) S. 1 DSGVO). Dies wird allerdings die Ausnahme sein, sodass Art. 9 Abs. 2 lit. h DSGVO im Ergebnis keine umfassend nutzbare Rechtsgrundlage für die Verarbeitung der Daten im NGD darstellt.<sup>4</sup>

## (2.) Öffentlichen Gesundheit, schwerwiegende Gesundheitsgefahren (Art. 9 Abs. 2 lit. i DSGVO)

Art. 9 Abs. 2 lit. i DSGVO bezieht sich u.a. auf die Verarbeitung von Gesundheitsdaten im öffentlichen Gesundheitswesen und die Gewährleistung hoher Qualitäts- und Sicherheitsstandards bei der Gesundheitsversorgung und bei Arzneimitteln und Medizinprodukten. Tatbestandsvoraussetzung ist, dass die Verarbeitung im öffentlichen Interesse liegt. Der Erlaubnistatbestand verweist gleichfalls auf eine nötige Konkretisierung durch das Unionsrecht und das Recht einzelner Mitgliedstaaten.

Der Begriff der „öffentlichen Gesundheit“ ist im Sinne der Verordnung (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates auszulegen. Dieser umfasst alle Elemente im Zusammenhang mit der Gesundheit wie den Gesundheitszustand einschließlich Morbidität und Behinderung, die sich auf diesen Gesundheitszustand auswirkenden Determinanten, den Bedarf an Gesundheitsversorgung, die der Gesundheitsversorgung zugewiesenen Mittel, die Bereitstellung von Gesundheitsversorgungsleistungen und den allgemeinen Zugang zu diesen Leistungen sowie die entsprechenden Ausgaben und die Finanzierung und schließlich die Ursachen der Mortalität (vgl. Erwägungsgrund (54) S. 3 DSGVO).

Für das „öffentliche Interesse“ enthält die Verordnung keine Legaldefinition. Es handelt es sich um einen unbestimmten Rechtsbegriff, der sich auf die Belange des Gemeinwohls bezieht und einem übergeordneten gesellschaftlichen Ziel dient. Was ein im öffentlichen Interesse liegendes Ziel ist, liegt im Ermessen des nationalen Gesetzgebers.

Art. 9 Abs. 2 lit. i DSGVO wird dadurch eingeschränkt, dass eine solche Verarbeitung von Gesundheitsdaten aus Gründen des öffentlichen Interesses nicht dazu führen darf, dass Dritte, unter anderem Arbeitgeber oder Versicherungs- und Finanzunternehmen, solche personenbezogene Daten zu anderen Zwecken verarbeiten (vgl. Erwägungsgrund (54) S. 4 DSGVO).

Beispielhaft kann hier § 9 Infektionsschutzgesetz (IfSG) als eine nationale für den vorgenannten Zweck dienende Verarbeitungsnorm angeführt werden. Das IfSG erlaubt es dem RKI, im gesetzlich festgelegten Rahmen Gesundheitsdaten zu den Zwecken des Art. 9 Abs. 2 lit. i DSGVO zu verarbeiten.

Soweit die Datenbereiter oder das NGD die Auswertung zum Zwecke der Sicherstellung der öffentlichen Gesundheitsversorgung durchführt, ließe sich die Verarbeitung der Gesundheitsdaten auf Art. 9 Abs. 2 lit. i DSGVO stützen. Damit könnte eine datenschutzrechtliche Zulässigkeit für einen Teil der angestrebten Auswertungen hergestellt werden, jedoch nicht für alle NGD-Auswertungen. Insbesondere können sich Dritte dann nicht auf diese Erlaubnisnorm stützen, wenn Gesundheitsdaten zu anderen, nicht im öffentlichen Interesse liegenden, Zwecken verarbeitet werden sollen. NGD müsste dementsprechend ein Teil des öffentlichen Gesundheitswesens und entsprechend verfasst sein.

Die Nutzung dieser Erlaubnisnorm setzt zudem voraus, dass der nationale Gesetzgeber eine entsprechende konkretisierende Regelung schafft. Dies tut der deutsche Gesetzgeber mit der Regelung zur Verarbeitung besonderer Kategorien personenbezogener Daten in § 22 BDSG n.F. unter anderem

<sup>4</sup> vgl. Weichert in Kühling/Buchner (Hrsg.), DSGVO, 2. Aufl. 2018, Art. 9 Rn. 93.

auch für die Ermächtigung des Art. 9 Abs. 2 lit. h und i DSGVO zur Verarbeitung von Gesundheitsdaten für öffentliche und nicht-öffentliche Stellen.

(3.) Archiv-, Forschungs- und statistische Zwecke (Art. 9 Abs. 2 lit. j DSGVO)

Art. 9 Abs. 2 lit. j DSGVO bezieht sich explizit auf die Verarbeitung von Gesundheitsdaten für wissenschaftliche Forschungszwecke oder für statistische Zwecke unter Verweis auf die Erforderlichkeit gem. Art. 89 Abs. 1 DSGVO. Zudem verweist der Erlaubnistatbestand gleichfalls auf eine nötige Konkretisierung durch das Unionsrecht oder das Recht einzelner Mitgliedstaaten.

*Wissenschaftliche Forschungszwecke und Informationen aus Registern*

Erwägungsgrund (159) DSGVO stellt klar, dass die Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken weit auszulegen ist und die Verarbeitung für beispielsweise die technologische Entwicklung und die Demonstration, die Grundlagenforschung, die angewandte Forschung und die privat finanzierte Forschung einschließen. Die wissenschaftlichen Forschungszwecke sollten auch Studien umfassen, die im öffentlichen Interesse im Bereich der öffentlichen Gesundheit durchgeführt werden. Um den Besonderheiten der Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken zu genügen, sollten spezifische Bedingungen insbesondere hinsichtlich der Veröffentlichung oder sonstigen Offenlegung personenbezogener Daten im Kontext wissenschaftlicher Zwecke gelten.

Erwägungsgrund (157) DSGVO stellt klar, dass durch die Verknüpfung von Informationen aus Registern Forscher neue Erkenntnisse von großem Wert in Bezug auf weit verbreiteten Krankheiten wie Herz-Kreislauferkrankungen, Krebs und Depression erhalten können. Durch die Verwendung von Registern können bessere Forschungsergebnisse erzielt werden, da sie auf einen größeren Bevölkerungsanteil gestützt sind. Durch Register erhaltene Forschungsergebnisse bieten solide, hochwertige Erkenntnisse, die die Basis für die Erarbeitung und Umsetzung wissenschaftlich gestützter politischer Maßnahmen darstellen, die Lebensqualität zahlreicher Menschen verbessern und die Effizienz der Sozialdienste verbessern können. Zur Erleichterung der wissenschaftlichen Forschung können daher personenbezogene Daten zu wissenschaftlichen Forschungszwecken verarbeitet werden, wobei sie angemessenen Bedingungen und Garantien unterliegen, die im Unionsrecht oder im Recht der Mitgliedstaaten festgelegt sind.

*Statistische Zwecke*

Erwägungsgrund (162) DSGVO stellt klar, dass das Unionsrecht oder das Recht der Mitgliedstaaten in den Grenzen dieser Verordnung den statistischen Inhalt, die Zugangskontrolle, die Spezifikationen für die Verarbeitung personenbezogener Daten zu statistischen Zwecken und geeignete Maßnahmen zur Sicherung der Rechte und Freiheiten der betroffenen Personen und zur Sicherstellung der statistischen Geheimhaltung bestimmen sollte. Unter dem Begriff „statistische Zwecke“ ist jeder für die Durchführung statistischer Untersuchungen und die Erstellung statistischer Ergebnisse erforderliche Vorgang der Erhebung und Verarbeitung personenbezogener Daten zu verstehen. Diese statistischen Ergebnisse können für verschiedene Zwecke, so auch für wissenschaftliche Forschungszwecke, weiterverwendet werden. Im Zusammenhang mit den statistischen Zwecken wird vorausgesetzt, dass die Ergebnisse der Verarbeitung zu statistischen Zwecken keine personenbezogenen Daten, sondern aggregierte Daten sind und diese Ergebnisse oder personenbezogenen Daten nicht für Maßnahmen oder Entscheidungen gegenüber einzelnen natürlichen Personen verwendet werden.

(4.) Nationale Regelung auf Grundlage der Öffnungsklausel des Art. 9 Abs. 2 lit. j DSGVO

Mit der Regelung zur Verarbeitung von Daten zu wissenschaftlichen Forschungszwecken und zu statistischen Zwecken in § 27 Abs. 1 BDSG n.F. konkretisiert der deutsche Gesetzgeber unter anderem auch die Ermächtigung des Art. 9 Abs. 2 lit. j DSGVO zur Verarbeitung von Gesundheitsdaten. Er gilt für die öffentliche und private Forschung durch öffentliche und nicht-öffentliche Stellen und schafft eine

Regelung im nationalen Recht für die Verarbeitung von Gesundheitsdaten zu wissenschaftlichen Forschungszwecken und zu statistischen Zwecken.<sup>5</sup>

Die Verarbeitung nach § 27 Abs. 1 BDSG n.F. setzt dabei immer auch das Vorliegen einer Rechtsgrundlage nach Art. 6 Abs. 1 DSGVO voraus. Ergänzend zu Art. 9 Abs. 2 DSGVO kann die Verarbeitung von Gesundheitsdaten zur öffentlichen und privaten Forschung durch öffentliche und nicht-öffentliche Stellen somit nach Durchführung einer Interessenabwägung zulässig sein, sofern die Verarbeitung zu diesen Zwecken erforderlich ist und die Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen.

**Folglich kann ein Großteil der Arbeit des NGD auf Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F. gestützt werden.** Unter Berücksichtigung des Erwägungsgrunds (157) DSGVO gilt dies auch für Register, deren Datenbestände so im Rahmen der Arbeit des NGD genutzt werden könnten.

Da im Rahmen der „Nutzung“ des NGD die Zwecke der Verarbeitung je nach Nutzer variieren können, muss die Einschätzung der rechtlichen Zulässigkeit gleichwohl stets am konkreten Fallbeispiel vorgenommen werden. Aufgrund der strengen Zweckbindung (vgl. Art. 5 Abs. 1 lit. b DSGVO) und der Sensibilität der Daten wird es nicht ausreichen, pauschal ein übergeordnetes öffentliches Interesse als Zweck anzuführen.

Vielmehr kann zwar eine grundsätzliche Nutzbarkeit der Rechtsgrundlage aus Art. 9 Abs. 2 lit. j angenommen werden, es muss jedoch, wie zuvor beschrieben, gem. § 27 Abs. 1 BDSG n.F. jeweils eine Interessenabwägung für den Einzelfall durchgeführt werden.

#### *e) Sicherstellung der Zweckbindung durch das NGD*

Der konkrete Zweck, zu dem Gesundheitsdaten der Datenbereitsteller verarbeitet werden, bestimmt sich durch die einzelne Auswertungsanfrage des Datennutzers (z.B. eine Datenanfrage einer Krankenversicherung an Datenbereitsteller zum Zwecke der Gewinnung von Erkenntnissen über örtliche Krankheitsschwerpunkte) und wird ggf. im Rahmen von Verhandlungen zwischen Datennutzer und Datenbereitsteller modifiziert und konkretisiert.

Das NGD prüft entsprechende Anfragen von Datennutzern auf Kompatibilität mit den übergeordneten Zwecken und stellt sicher, dass sich diese von den übergeordneten Zwecken des Verbunds ableiten lassen. Entsprechend der unterschiedlichen Reichweite der Erlaubnistatbestände bewertet das NGD darüber hinaus, welche datenschutzrechtlichen Grundlage für die einzelne Auswertung entsprechend ihres jeweiligen konkreten Zwecks gegeben ist.

Zudem ist zu berücksichtigen, dass sich aus spezialgesetzlichen Vorschriften im nationalen Recht Beschränkungen der ursprünglichen Zwecke für die Erhebung, Speicherung und Nutzung von Gesundheitsdaten ergeben, denen die Datenbereitstellern unterliegen. Die in Art. 9 Abs. 4 DSGVO enthaltene Öffnungsklausel erlaubt es den Mitgliedsstaaten zusätzliche Bedingungen, einschließlich Beschränkungen, einzuführen oder aufrechtzuerhalten, soweit die Verarbeitung von genetischen, biometrischen oder Gesundheitsdaten betroffen ist. Solche Beschränkungen der Verarbeitungszwecke im nationalen Recht sind somit nach Art. 9 Abs. 4 DSGVO zulässig. So ist es denkbar, dass Daten der Datenbereitsteller auf Basis nationalen Rechts mit beschränkterer Zweckbindung gespeichert werden (z.B. in den Sozialgesetzbüchern, in den krankenhausspezifischen Bundes- und Landesgesetzen sowie den Spezialgesetzen des RKI, DIMDI und der Krebsregister). In einem solchen Fall wird sich die Privilegierung der Forschung in Bezug auf die Kompatibilität der Zwecke, wie dies Art. 5 Abs. 1 lit. b DSGVO vorsieht, nicht über diese zulässige Einschränkung hinwegsetzen. Dies ist vom NGD bei der Prüfung der Kompatibilität der Zwecke mit zu berücksichtigen. Auf den damit zusammenhängenden notwendigen gesetzgeberischen Anpassungsbedarf wird im Folgenden noch näher eingegangen.

<sup>5</sup> vgl. Entwurf eines Gesetzes zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680), S. 99.

*f) Einwilligungserklärung nach Art. 9 Abs. 2 lit. a DSGVO*

Die Weitergabe der Daten durch die Datenbereitsteller an NGD und die Durchführung von Auswertungen im NGD stützen sich in der Regel auf Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F. Insofern kann davon ausgegangen werden, dass die datenschutzrechtlich ebenfalls denkbare Möglichkeit der Datenverarbeitung im NGD mit Einwilligung der Betroffenen nicht Gebrauch gemacht werden muss. Für den Fall, dass dies gleichwohl erfolgen soll, gilt folgendes:

Voraussetzung für eine wirksam erteilte (Patienten-)Einverständniserklärung zur Verarbeitung von Gesundheitsdaten durch NGD ist, dass diese neben der ausdrücklichen Erklärung den Wirksamkeitsvoraussetzungen des Art. 7 DSGVO entspricht. Zudem muss der Patient nach Maßgabe des Art. 12 Abs. 1 S. 1 DSGVO über alle Informationen, die sich auf die Verarbeitung beziehen, in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache zu informiert werden. Die zu erteilenden Informationen ergeben sich bei einer Erhebung beim Patienten aus Art. 13 Abs. 1 und 2 DSGVO.

Da der Zweck der Verarbeitung personenbezogener Daten für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Erhebung der personenbezogenen Daten i.d.R. nicht vollständig angegeben werden, ist es zulässig, dass die Betroffenen ihre Einwilligung für bestimmte Bereiche wissenschaftlicher Forschung erteilen, wenn dies unter Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung geschieht. Den Betroffenen ist es somit möglich, ihre Einwilligung nur für bestimmte Forschungsbereiche oder Teile von Forschungsprojekten in dem vom verfolgten Zweck zugelassenen Maße zu erteilen (Erwägungsgrund (33) DSGVO).

Weil NGD im Modell in keiner direkten Beziehung zu den Betroffenen steht und weil der Zweck der Auswertung im NGD für die dortige Datenverarbeitung ebenso bestimmend ist wie für die Datenübermittlung durch die Datenbereitsteller, müsste NGD für die Fälle der Notwendigkeit der Einholung der Einwilligung den Datenbereitstellern vorausgefüllte Einwilligungserklärungen bereitstellen, die den Zweck für bestimmte Forschungsbereiche oder Teile von Forschungsprojekten in dem zugelassenen Maße benennt und die weiteren Informationspflichten erfüllt.

3. Anforderungen an die organisatorische Ausgestaltung von Datenschutz und IT-Sicherheit

*a) Grundsätze in Bezug auf Datenschutz- und IT-Sicherheit*

Die Anforderungen an Datenschutz- und IT-Sicherheit orientieren sich grundsätzlich an den Schutzziele der Wahrung der Integrität und Vertraulichkeit (Art. 5 Abs. 1 lit. f DSGVO). Daneben konkretisiert Art. 25 sowie 32 bis 35 DSGVO die Anforderungen an die Ausgestaltung des Datenschutzmanagements durch Datenschutz durch Technikgestaltung, datenschutzfreundliche Voreinstellungen, zu treffende technischen und organisatorischen Maßnahmen, Meldepflichten bei Datenschutzverstößen sowie durch das Instrument der Datenschutz-Folgenabschätzung.

Gemäß dem Grundsatz der Vertraulichkeit und Integrität müssen personenbezogene Daten in einer Weise verarbeitet werden, die eine angemessene Sicherheit der personenbezogenen Daten gewährleistet, einschließlich Schutz vor unbefugter oder unrechtmäßiger Verarbeitung und vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung durch geeignete technische und organisatorische Maßnahmen.

*b) Datenschutz-Folgenabschätzung nach Art. 35 DSGVO*

Durch die DSGVO wird das bereits bekannte Konzept der Vorabkontrolle durch das Konzept einer Datenschutz-Folgenabschätzung (DSFA) weitergeführt und ausgebaut. Eine DSFA ist ein Verfahren, anhand dessen die Verarbeitung beschrieben, ihre Notwendigkeit und Verhältnismäßigkeit bewertet und die Risiken für die Rechte und Freiheiten natürlicher Personen, die die Verarbeitung personenbezogener Daten mit sich bringt, durch eine entsprechende Risikoabschätzung und die Ermittlung von Gegenmaßnahmen besser kontrolliert werden sollen.

Für die umfangreiche Verarbeitung von Gesundheitsdaten (vgl. besondere Kategorien von personenbezogenen Daten gem. Art. 9 Abs. 1 DSGVO) durch die Neutrale Stelle ist die Durchführung

einer DSFA vorgeschrieben (vgl. Art. 35 Abs. 3 lit. b DSGVO). Die Datenschutz-Folgenabschätzung muss gemäß Art. 35 lit. a bis d DSGVO Folgendes beinhalten:

- a) eine systematische Beschreibung der geplanten Verarbeitungsvorgänge und der Zwecke der Verarbeitung (gegebenenfalls einschließlich der von dem Verantwortlichen verfolgten berechtigten Interessen);
- b) eine Bewertung der Notwendigkeit und Verhältnismäßigkeit der Verarbeitungsvorgänge in Bezug auf den Zweck (Zweck-Mittel-Relation);
- c) eine Bewertung der Risiken für die Rechte und Freiheiten der betroffenen Personen unter Berücksichtigung der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung und
- d) die zur Bewältigung der Risiken geplanten Abhilfemaßnahmen, einschließlich Garantien, Sicherheitsvorkehrungen und Verfahren, durch die der Schutz personenbezogener Daten sichergestellt und der Nachweis dafür erbracht wird, dass diese Verordnung eingehalten wird, wobei den Rechten und berechtigten Interessen der betroffenen Personen und sonstiger Betroffener Rechnung getragen wird.

Im Rahmen der Folgenabschätzung kann es erforderlich sein, dass das NGD bei der Verarbeitung zur Erfüllung einer im öffentlichen Interesse liegenden Aufgabe, einschließlich der Verarbeitung zu Zwecken der sozialen Sicherheit und der öffentlichen Gesundheit, die zuständige Aufsichtsbehörde vor der Verarbeitung konsultiert und deren vorherige Genehmigung einholt (vgl. Art. 36 Abs. 5 DSGVO).

Zudem wird es aufgrund der Sensibilität der Daten und des Umfangs der Datenverarbeitung des NGD erforderlich sein eine regelmäßige Überprüfung der Risikobewertung durchzuführen. Dies gilt zumindest, wenn hinsichtlich des mit den Verarbeitungsvorgängen verbundenen Risikos Änderungen eingetreten sind (vgl. Art. 35 Abs. 11 DSGVO).

#### *c) Technische und organisatorische Maßnahmen nach Art. 32 DSGVO*

Der Sektor des Gesundheitswesens ist gemäß § 2 Abs. 10 BSI-G eine kritische Infrastruktur. Jedoch fällt die von NGD erbrachte Dienstleistung nicht in den Katalog der vom IT-Sicherheitsgesetz regulierten Dienstleistungen des Gesundheitswesens (§ 6 BSI-Kritis-VO). Sie müsste vom Verordnungsgeber explizit aufgenommen werden. Sofern NGD in größerem Umfang genutzt werden wird, könnte sich eine kritische Funktion innerhalb des Gesundheitswesens ergeben. Dem sollte bei der rechtlichen Ausgestaltung des IT-Sicherheitsmanagements, ggf. Änderung der BSI-Kritis-VO, Rechnung getragen werden.

#### 4. Vertragliche Abbildung der Arbeit des NGD und die Nutzung durch die Nutzer

Das Netzwerk Gesundheitsdaten (NGD) soll bestimmte Grundanforderungen erfüllen:

- Vorbildlicher Umgang mit den sensiblen Daten
- Hohes Vertrauen der Datenbereiter, Datennutzer und der Öffentlichkeit
- Besondere Steuerungsmöglichkeit für die öffentliche Hand

Die Erfüllung dieser Anforderungen ist nur durch eine langfristig angelegte, institutionalisierte Einrichtung möglich, in deren Kern das NGD steht.

Es wird im Folgenden davon ausgegangen, dass das NGD eine rechtsfähige Einrichtung ist, die von keinem Datenbereiter oder Datennutzer beherrscht wird und einer öffentlich verfassten Kontrolle unterliegt. Aus dem Modell können die nachfolgenden vertraglichen Ausgestaltungsnotwendigkeiten extrahiert werden:

1. Nutzung „Netzwerk Gesundheitsdaten“ (NGD)
  - 1.1. Teilnahmebedingungen für Datennutzer (DN)
  - 1.2. Teilnahmebedingungen für Datenbereiter (DB)
2. Anfrage über den Bereich Plattform des NGD
  - 2.1. Datennutzer stellt eine allgemeine Anfrage an den Bereich „Plattform“
  - 2.2. Datennutzer stellt eine konkrete Anfrage direkt an mögliche Datenbereiter

### 3. Annahme der Anfrage

- 3.1. Annahme der Anfrage durch den Datenbereitsteller
- 3.2. Verweigerung der Anfrage durch den Datenbereitsteller
- 3.1.1. Bei Annahme: Vertragsgestaltung durch den Bereich Contracting und Datenschutz des NGD
- 3.1.1.1. Rechtliche Prüfung der Datenzusammenführung durch das NGD
- 3.1.1.2. Vertragsaushandlung zwischen dem Datennutzer und den Datenbereitstellern
- 3.1.1.3. Vertragsaushandlung jeweils zwischen dem Datennutzer und dem NGD bzw. dem/ den Datenbereitsteller/n und dem NGD

Die Anforderungen an die jeweilige Ausgestaltung der Rechtsbeziehungen können an dieser Stelle aufgrund der starken Abhängigkeit von der Governance des Verbundes und der Rechtsform des NGD nur oberflächlich dargestellt werden. Eine Diskussion und Konkretisierung der detaillierten rechtlichen Ausgestaltung erfolgt in Kapitel III. Insgesamt kann jedoch festgehalten werden, dass die rechtliche Ausgestaltung der Beziehungen zwischen NGD, Datenbereitsteller und Datennutzer zweigeteilt werden muss in:

- die rechtliche Ausgestaltung der Verträge zwischen den Datennutzern (DN) und den Datenbereitstellern (DB) zur Durchführung der Studie,
- die rechtliche Ausgestaltung über die gemeinsame Studiendurchführung und temporäre Datenzusammenführung zur Datenanalyse zwischen Datennutzer (DN) und Datenbereitsteller (DB) und NGD.

## 5. Erforderlichkeit regulativer Anpassungen

Grundsätzlich lassen sich die legitimen Zwecke für die Verarbeitung von Gesundheitsdaten aus den Rechtsgrundlagen des Art. 9 Abs. 2 lit. i und j DSGVO i.V.m. §§ 22 und 27 BDSG n.F. ableiten. Allerdings regeln im nationalen Raum für die verschiedenen möglichen Datennutzer und Datenbereitsteller des NGD, Sondervorschriften die zulässigen Zwecke der Verarbeitung sowie zusätzliche Beschränkungen. Dies macht es zwingend erforderlich, dass im nationalen Raum regulativen Anpassungen vorgenommen werden. Diese im nationalen Raum erforderlichen regulativen Anpassungen beziehen sich auf zwei Aspekte:

1. Erstellung eines gesetzlichen Auftrags für die Arbeit des NGD  
(wie bspw. die DaTraV für das DIMDI; das BDBOS-Gesetz für die BDBOS; das KRG für die Krebsregister etc.)
2. Datenschutzrechtliche Zulässigkeit der Bereitstellung der Daten durch die einzelnen Nutzer im nationalen Raum  
(bspw. Anpassung SGB I, V, X; Anpassung landesspezifischer Krankenhausgesetze etc.)

Im Rahmen der Machbarkeitsbetrachtung (Kapitel III) werden für die genannten Beispielfälle konkrete Umsetzungsvorschläge genannt.

## C. Technische Anforderungen an eine Umsetzung

Eine Flut von Daten wird täglich im Gesundheitswesen erzeugt. Um diese großen Datenmengen zu verwalten und zu analysieren braucht es eine technische IT-Lösung, die für solche Herausforderungen des „Big Data“ konzipiert worden ist. Darüber hinaus müssen personenbezogene Daten besonders geschützt werden, was eine hohe IT-Sicherheit impliziert. In Fällen, wo es möglich ist, müssen diese anonymisiert oder pseudonymisiert werden. In anderen Fällen müssen besondere Maßnahmen getroffen werden, damit das System auch sicher mit personenbezogenen Daten arbeiten kann. Eine andere Anforderung an das technische System ist eine Mediatorplattform als technische Unterstützung der Neutralen Stelle zwischen interessierten Parteien, die bei der Aushandlung der rechtlichen und technischen Rahmenbedingungen hilft und die Datenanalyse durchführt.

Im Folgenden wird zunächst eine Begriffsklärung vorgenommen, um auf dieser Basis die Anforderungen anhand des Stands der Technik hinsichtlich technischer Big Data-Plattformen, Datenkataloge und Interoperabilitätsstandards im Gesundheitswesen zu spiegeln.

## 1. Begriffsklärung

### a) *Big Data und Smart Data*

Unter Big Data versteht man Datensätze, die mit herkömmlicher Software nur schwer oder gar nicht bearbeitet werden können. Big Data Datensätze beinhalten die Erfassung von Daten, Datenspeicherung, Datenanalyse, Suche, Freigabe, Übertragung, Visualisierung, Abfrage, Aktualisierung, Datenschutz und Datenquelle.

Im Gegensatz zu Big Data konzentriert sich Smart Data auf den Wert der Daten. Die meisten Big Data Datensätze sind heterogen und schlecht annotiert. Ohne eine klare Struktur sind diese Daten für die meisten Anwendungen unbenutzbar. Unter Smart Data versteht man Big Data, welche entweder mit einer klaren semantischen Struktur erzeugt werden oder eine Strukturierung im Nachhinein erfolgt, um schnelle und nachvollziehbare Entscheidungen für Data Analytics Anwendungen zu ermöglichen.

### b) *Big Data-Plattform*

Eine Big Data-Plattform ist eine Ansammlung von mehreren Big Data Anwendungen und Dienstprogrammen, die in einer einzigen Lösung vereint werden. Es handelt sich um IT-Plattformen, die für den Einsatz in großen Unternehmen gedacht sind, die sogenannte Enterprise-Klasse. Solche IT-Plattformen ermöglichen es Unternehmen und anderen Organisationen, eine Big Data Infrastruktur zu entwickeln, bereitzustellen, zu betreiben und zu verwalten.

### c) *Big Data Analytics*

Data Analytics ist der Prozess der Analyse von Datensätzen mit Hilfe von spezialisierter Hardware und Software, um Rückschlüsse auf die darin enthaltenen Informationen ziehen zu können. Es umfasst die Entdeckung, Interpretation und Kommunikation von aussagekräftigen Mustern in Daten und setzt Algorithmen und Verfahren aus Informatik, Mathematik und Statistik ein.

Im Gegensatz zu traditionellen Data Analytics bringt Big Data neue Herausforderungen für Data Analytics. Hier werden spezielle Algorithmen eingesetzt, die besser mit großen Datenmengen, heterogenen Datensätzen und nicht ausreichend gelabelten Daten umgehen können.

### d) *Open Data*

Open Data ist eine gesellschaftliche Bewegung, hinter der die Idee steckt, dass in einer datengetriebenen Welt manche Datensätze, vor allem aus dem Bereich der öffentlichen Hand, zur freien Verfügung stehen müssen, und zwar ohne Einschränkungen durch Urheberrechte, Patente oder andere Kontrollmechanismen, um Forschung und Gesellschaft voranzutreiben.

Die Ziele der Open Data-Bewegung sind ähnlich derer anderer "offener" Bewegungen wie Open Source, Open Hardware, Open Content, Open Government, Open Access und Open Science. Viele Regierungen sind sich der gesellschaftlichen Wichtigkeit dieser Bewegung bewusstgeworden, wie man - durch die Gewinnung an Popularität von Open Data - Regierungsinitiativen wie Data.gov, Data.gov.uk und Open-Data.europa.eu hervorgerufen hat und auch wie man mit Hilfe einer Vielzahl von Gesetzen öffentliche Einrichtungen dazu zwingen kann, ihre Daten als Open Government Data zur Verfügung zu stellen.

### e) *Linked Data*

Eine der Grundansätze des World Wide Webs ist die Verknüpfung zwischen Dokumenten, die es ermöglicht, auch durch große Dokumentensammlungen zu navigieren. Eines der Probleme, die mit der Offenlegung großer Datenmengen aus verschiedenen Domänen entstanden sind, war die Heterogenität und die Zusammenführung dieser Daten. Um dieses Problem zu lösen, wurden die Grundideen des WWW eingesetzt. Sie sind im Rahmen des Semantic Webs weiterentwickelt worden, um die Datensätze miteinander zu verknüpfen. Dies führte zu der Erschaffung von verknüpften Offenen Daten oder Linked

Open Data, Linked Data, sowie auch das Semantic Web, bauen auf Standard Webtechnologien wie HTTP, XML/RDF und URIs auf, aber sie haben nicht die Erschaffung von Webseiten für menschliche Nutzer im Fokus, sondern Informationen auf eine Weise zu beschreiben, dass sie von Computern automatisch gelesen und verarbeitet werden können.

#### f) *Datenkataloge*

Ein Datenkatalog stellt Funktionen bereit, mit denen Datenquellen entdeckt und genutzt werden können. Benutzer sind Datenanalysten, Entwickler oder andere Wissenschaftler. Typischerweise beinhaltet ein Datenkatalog eine Crowdsourcing Komponente, die es den Benutzern erlaubt, Datenquellen beizutragen. Zusätzlich muss aber auch ein durchdachtes Metadatenmodell, mit dem die Datensätze annotiert werden, eingehalten werden.

#### g) *Data Warehouse*

Eine große, strukturierte und spezialisierte Datenbank, die Informationen aus vielen verschiedenen Anwendungen sammelt und ineinander integriert, um diese Daten (typischerweise als ein mehrdimensionaler Datenwürfel) dann für die Verwendung in Business Intelligence und Datenanalyseanwendungen geeignet zu machen.

#### h) *Data Lake*

Ein sehr großer Datenspeicher von unstrukturierten Daten. Dieser Datenspeicher beinhaltet Daten aus sehr vielen unterschiedlichen internen und externen Quellen im Originalformat.

#### i) *IT-Sicherheit*

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) definiert IT-Sicherheit als „Zustand, in dem Vertraulichkeit, Integrität und Verfügbarkeit von Informationen und Informationstechnik durch angemessene Maßnahmen geschützt sind“<sup>6</sup>. Damit nennt die Definition bereits die wichtigsten Schutzziele:

- Vertraulichkeit der Daten (Confidentiality)
- Integrität der Daten (z.B. Schutz vor Verändern) (Integrity)
- Verfügbarkeit der Daten (Availability)

Weitere Schutzziele sind z.B.:

- Zurechenbarkeit und Nichtabstreitbarkeit: Wissen, wer (Personen, Rollen, Systeme) die Daten erzeugt, geändert oder gelöscht hat.
- Authentifizierung: Fähigkeit, den Nachweis zu führen, dass eine Person, eine Rolle oder ein System tatsächlich diejenige ist, die sie vorgibt zu sein.
- Autorisierung: Die Fähigkeit, eine Person, eine Rolle oder ein System nur zu bestimmten Aktionen zu berechtigen.

## 2. *Big Data-Plattformen und Datenkataloge*

In diesem Abschnitt werden wir einen Überblick über Big Data und seine definierenden Merkmale sowie über die Auswirkung dieser Merkmale bei der Anwendung auf das Gesundheitswesen geben. Wir werden uns auch mit dem aktuellen Stand der Technik auf Big Data-Plattformen und deren Verwendung für das Gesundheitswesen befassen.

### a) *Big Data – Herausforderungen*

Folgende Herausforderungen sind typische Merkmale von Big Data, und werden auch als die (englischen) Vs von Big Data benannt:

- **Volume:** Die Daten müssen so groß sein, dass herkömmliche Datenverwaltungssoftware Probleme beim Speichern und Verarbeiten der Daten haben.

<sup>6</sup> [https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/Glossar/glossar\\_node.html](https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/Glossar/glossar_node.html), Zugriff am 27. Juli 2018

- **Velocity:** Diese Eigenschaft behandelt die Geschwindigkeit, mit der neue Daten erstellt werden und die Wichtigkeit, mit der diese Daten (typischerweise Realzeitdaten) schnell verarbeitet werden müssen.
- **Variety:** Wie bei den meisten Organisationen sind die meisten Daten nicht in einem einzigen Format, sondern in einer breiten Palette von Formaten wie Text, Datenbanken, Tabellenkalkulationen, Bildern oder Video.

In den letzten Jahren Zeit sind weitere Aspekte von Big Data wichtig geworden, wie z.B.

- **Validity:** Gewährleistet ein gewisses Maß an Datenqualität, sowie Veracity, die sich mit der Herkunft und Reputation der Daten befasst. Diese Eigenschaft ist sehr wichtig aufgrund der Tatsache, dass große Datenmengen auch eine große Menge an Daten niedriger Qualität zur Folge haben.
- **Value:** Dieses Merkmal beschreibt, wie wertvoll bestimmte Daten für Unternehmen oder Organisationen sind. Nicht alle Daten sind gleich wertvoll, gut strukturierte und beschriebene Daten mit guter Validität und Richtigkeit können schnell gehandelt und somit wertvoller werden.
- **Volatility:** Beschreibt, wie lange auf die Daten in ihrem aktuellen Zustand in verschiedenen Quellen zugegriffen werden kann und wie lange sie gespeichert werden. Einige Daten müssen für eine lange Zeit an der Quelle gespeichert werden, damit die Analyseergebnisse verifizierbar sind, andere Daten, wie die, die in Echtzeit verarbeitet werden, müssen möglicherweise gar nicht gespeichert werden.

Für den Einsatz im Gesundheitswesen stellen diese Merkmale unterschiedliche Anforderungen dar. Während Volume, Velocity und Variety eine Anforderung für die Infrastruktur des Systems darstellen, eine Infrastruktur, die mit immer größer werdenden Datenmengen verschiedener Arten von Datenformaten umgehen muss, die immer schneller in das System eingegeben werden, sind andere wie Validität und Veracity, Value und Volatility auch Aufgaben für das rechtliche und operative System.

#### *b) Hadoop Ecosystem als Big Data-Plattform*

Obwohl es keine Standarddefinition dafür gibt, was genau eine "Big Data-Plattform" ist, hat sich das Hadoop als der De-facto-Standard für die Verarbeitung von Big Data durchsetzen können. Fast alle großen Softwareanbieter wie IBM, SAP, Microsoft oder die Deutsche Telekom benutzen es. Apache Hadoop<sup>7</sup> ist ein in Java geschriebenes Framework, welches dafür optimiert ist, große Datenmengen dezentralisiert zu speichern und zu analysieren. Es zeichnet sich durch seine hervorragende horizontale Skalierbarkeit aus; so kann es sehr kostengünstig um neue dezentrale Knoten erweitert werden, sollte sich nach Inbetriebnahme zeigen, dass mit größeren Datenmengen zu arbeiten ist, als am Anfang geplant war. Es kann z.B. unproblematisch auf Hunderte oder sogar Tausende Knoten hochskaliert werden, welche die Rechenaufgaben unter sich intelligent aufteilen.

Der Kern von Apache Hadoop ist Open Source Software, was bedeutet, dass es kostenlos genutzt werden kann, da der Quellcode offen ist und jedem zur Verfügung steht. Ein weiterer Vorteil, der sich hieraus ergibt, ist die Unabhängigkeit von einem konkreten Software-Hersteller - jedem Benutzer steht frei, die Software nach Belieben zu modifizieren und weiterzugeben, solange man sich dabei an ein paar Spielregeln hält. Viele Softwarelieferanten jedoch bündeln mehrere Komponenten aus dem Hadoop Ecosystem zusammen mit ihrer eigenen Software und bieten es als eine Big Data-Plattform an.

Hadoop ist mit seiner Reifezeit von über 10 Jahren heute ein Top-Level-Projekt der Apache Software Foundation, bei der auch eine Reihe anderer berühmter Open Source-Projekte angesiedelt sind.

Neben der bereits erwähnten horizontalen Skalierbarkeit zeichnet sich Hadoop durch seine Umsetzung des Datenlokalität-Prinzips aus: Hadoop überträgt die Daten nicht zu den Servern, wo sie dann analysiert werden würden, sondern es verteilt den auszuführenden Programmcode auf genau die Knoten eines Hadoop-Clusters, auf welchen sich die benötigten Daten bereits befinden. Wie man sich

<sup>7</sup> <http://hadoop.apache.org/>, Zugriff am 27. Juli 2018

vorstellen kann, führt dies zu massiven Zeitersparnissen, da es einfach nicht praktikabel ist, Petabytes an Daten kontinuierlich über ein Netzwerk von einem Knoten zum anderen zu verschieben.

Die Hadoop-Plattform selbst besteht aus folgenden Komponenten:

i. **HDFS,**

dem verteilten Dateisystem (Distributed Filesystem). Die Aufgabe von HDFS ist das Speichern von Daten auf eine solche Art, dass auf diese von den Hadoop-Knoten zugegriffen werden kann, wobei eine gute Verfügbarkeit und Fehlertoleranz im Fokus stehen.

ii. **Hadoop YARN,**

dem Ressourcen-Management-Layer, welches die konkreten Rechenaufgaben (Jobs) über das Cluster verteilt.

Die Art, wie Apache Hadoop arbeitet, basiert im Kern auf einem Programmiermodell namens "MapReduce", ursprünglich eingeführt durch Dean and Ghemawat von Google Inc. MapReduce ist ein Datenverarbeitungsschema, welches folgendermaßen vorgeht:

Um eine Aufgabe zu bearbeiten, wird diese zunächst in zwei Teile getrennt:

- Der "**Mapping**"-Teil beschäftigt sich damit, Daten aus dem verteilten Dateisystem (HDFS) zu lesen, sie vorzubearbeiten und eine Art Zwischenergebnisse zu produzieren, mit welchen weitergearbeitet werden kann.
- Der "**Reduktion**"-Teil besteht darin, die im Mapping-Teil entstehenden Zwischenergebnisse zu einem Endergebnis zu aggregieren. Dieses Endergebnis wird zum Schluss wieder in das HDFS gespeichert.

Der Sinn einer solchen Aufteilung besteht darin, dass das Mapping innerhalb einer Aufgabe sich ganz anders parallelisieren lässt als die Reduktion - wenn man diesen Umstand nutzt, erreicht man bei der Datenverarbeitung signifikante Beschleunigungen.

Kombiniert mit dem Prinzip der Datenlokalität sorgt das MapReduce-Paradigma für eine Effizienz, ohne die Datenanalysen im Bereich Big Data nur schwer möglich wären.

iii. **Apache Hive:**

Eine Data-Warehouse-Infrastruktur, die Datenzusammenfassung und Ad-hoc-Abfragen ermöglicht. Es ist ein System, das Benutzern die Tools zur Verfügung stellt, um leistungsstarke Abfragen zu erstellen und Ergebnisse häufig in Echtzeit zu erhalten.

iv. **Apache Spark<sup>8</sup>:**

Eine Engine für In-Memory verteiltes Rechnen, sie ermöglicht eine Datenanalyse auf sehr großen Datenmengen. Spark baut auf HDFS auf, umgeht jedoch MapReduce und verwendet stattdessen ein eigenes Datenverarbeitungsframework. Häufige Anwendungsfälle für Apache Spark sind Echtzeit-Abfragen, Verarbeitung von Ereignisströmen, iterative Algorithmen, komplexe Operationen und maschinelles Lernen.

Hinter Spark steht der Spark Core, welcher die grundlegenden Infrastruktur-Funktionalitäten bereitstellt. Spark ist dank Parallelisierung sehr leistungsfähig und mit umfangreichen Bibliotheken und Schnittstellen ausgestattet. Es ist ein Allzweck-Tool zur Datenverarbeitung, eine sogenannte Data Processing Engine. Genutzt wird das Framework von Data Engineers und Data Scientists, um äußerst schnelle Datenabfragen auf große Datenmengen im Terabyte-Bereich auszuführen. Der große Vorteil gegenüber des MapReduce-Algorithmus von Hadoop ist, dass Spark sehr gut iterative Schleifen verarbeiten kann, die für maschinelle Lernalgorithmen sehr wichtig sind. Die vorher genannte In-Memory-Technologie ermöglicht die besonders schnelle Auswertung von Daten – im Gegensatz zum MapReduce-Algorithmus, der zwar ebenfalls eine Parallellisierung hat, allerdings pro Arbeitsschritt Daten von einer Festplatte zur nächsten schreibt. Daher ist die Performance von Spark durch die Nutzung des Arbeitsspeichers hinsichtlich Zugriffszeiten und Schreiben von Daten unschlagbar.

<sup>8</sup> Eine alternative zu Spark ist Apache Flink, dass aber aktuell eine wesentlich kleinere ML Bibliothek bietet.

MMLib ist eine von Spark bereitgestellte Bibliothek mit der Aufgabe, maschinelles Lernen einfach und skalierbar zu machen. Es werden viele verschiedene Anwendungsmöglichkeiten bereitgestellt, darunter passende Modelle, Datenerfassung und Datenkennzeichnung, Featureextraktion und Featuretransformation, Modell-Tuning und Modell-Bewertung. Während in der Praxis oft Probleme mit vielen unterschiedlichen Bibliotheken gelöst werden müssen, was das zusammenführen sehr umständlich werden lässt, beinhaltet MMLib alles in einer Bibliothek und sorgt somit für Kompatibilität unter den unterschiedlichen Teilbereichen des maschinellen Lernens. Verwendet man zusätzlich auch weitere Komponenten von Apache Spark, erhält man eine einheitliche Lösung um bestmöglich gute Resultate zu erzielen. Auch wenn die maschinellen Lernanwendungen extrem vielfältig sind, bietet MMLib die grundlegenden Bausteine und ermöglicht dem Entwickler so das Erweitern der High-Level-API, zur Erstellung seiner Software.

v. **Apache Ambari:**

Eine Managementsoftware für Hadoop Cluster, es bietet eine Management- und Konfigurationsinterface für viele der Werkzeuge im Hadoop Ecosystem wie Hive, HBase, Pig, Sqoop oder Zookeeper.

vi. **Apache PIG:**

Eine Abfragesprache für Hadoop, damit können komplexe Abfragen über große Datenmengen gemacht werden. Es funktioniert zusammen mit Hadoop Map-Reduce oder Apache Spark.

vii. **Hadoop-Sicherheit: Apache Knox**

Apache Knox ist ein Projekt, welches sich ausschließlich dem Thema Sicherheit bei Apache Hadoop widmet. Zwar wird für Sicherheit bei Hadoop auf mehreren Ebenen gesorgt, aber Apache Knox adressiert zusätzliche Bereiche wie die feingranulare Rechteverwaltung für erfolgreich angemeldete Benutzer und ähnliche sicherheitsrelevante Aspekte von Hadoop.

viii. **Hadoop-Distributionen**

In der Praxis kommt es oft vor, dass das Projekt, bei welchem man Hadoop anwenden möchte, einem bestimmten Bereich zugeordnet werden kann, wie z.B. Sicherheit, Unternehmens-Workflow oder Governance-Integration. Solche Bereiche haben oft spezifische Anforderungen an Hadoop, deren Umsetzung sehr zeitaufwendig sein kann. Genau in solchen Fällen schaffen Hadoop-Distributionen Abhilfe. Zum Kernsystem von Hadoop (Hadoop Core) gehören nur der MapReduce-Algorithmus und HDFS. Wenn jedoch von "Hadoop" die Rede ist, ist meistens das ganze Hadoop-Ökosystem inkl. der aufeinander basierten Distributionen gemeint. Zu den Vorteilen solcher Distributionen gehört nicht nur der größere Funktionsumfang oder die Einfachheit beim Aufsetzen, sondern auch der Support, den Unternehmen für Hadoop-Distributionen erhalten. Zu den wichtigsten Distributionen zählen Amazon Web Services (AWS), Cloudera, Hortonworks, IBM, MapR Technologies, Teradata etc. Einige der Distributionen sind im Folgenden aufgelistet und näher beschrieben.

*Hortonworks*

Hortonworks wurde aus dem ursprünglichen Yahoo-Hadoop-Team gegründet. Bis heute ist Hortonworks die treibende Kraft der Hadoop-Core-Projekte. So fließen die meisten Entwicklungen von Hortonworks auch direkt wieder in das Open Source-Projekt der Apache Software Foundation ein. Diese Distribution trägt den offiziellen Namen "Hortonworks Data Platform" (HDP). Es ist die einzige vollständig quelloffene Distribution auf dem Markt. Das Aufsetzen dieser Distribution wird dadurch einfacher gemacht, dass man sie direkt von der offiziellen Projektseite herunterladen kann und sie auch als eine HDP-Sandbox vorliegt, eine vorkonfigurierte Umgebung für VMware, Hyper-V und Virtualbox.

Sie orientiert sich sehr stark am Hadoop-Standard. Sie ist als einzige Hadoop-Lösung auch in Windows HDP für Windows verfügbar. Die Installation von HDP erfolgt manuell und ist sehr gut dokumentiert. Zu prominenten HDP-Nutzern gehören Xing, WD oder Bloomberg.

### *Cloudera*

Die Cloudera Hadoop-Distribution CDH zeichnet sich dadurch aus, "enterprise ready" zu sein. Sie verwendet CDH YARN für das Workload Management und HDFS oder HBase als Storage Engine, sowie MapReduce, Hive und Pig für die Batch-Verarbeitung. CDH hat Apache Spark für die Zwecke von Echtzeitanalyse, Stream-Processing sowie Machine Learning integriert.

Ganz viele Sicherheitstechniken bringt Cloudera auch von Haus aus mit, so z.B. das Apache-Projekt Sentry, welches ein auf Rollen basierendes Authorisierungssystem für den Datenzugriff innerhalb eines Hadoop-Clusters implementiert.

Apache Accumulo, eine Key/Value-Datenbank, ist auch ein Bestandteil der Cloudera-Distribution.

Noch viel wichtiger ist jedoch, dass Cloudera über ein eigenes Installationsprogramm und einen proprietären Cloudera Manager sehr bequeme Cluster-Administration ermöglicht.

Cloudera hat drei Enterprise-Editionen: Basic, Flex und Data Hub. Die Hub Edition stellt sämtliche Tools zur Verfügung wie HBase, Spark, sowie alle Backup-Funktionen.

Zu prominenten Anwendern gehören Ebay, Netapp oder Samsung.

### *Amazon Elastic Map Reduce (EMR)*

Amazon stellt seine eigene Hadoop-Distribution Amazon Elastic MapReduce (Amazon EMR) als Web-Service zur Verfügung. Als Datenspeicher kann EMR HDFS, Amazon S3 oder Amazon DynamoDB verwenden. Der entscheidende Vorteil dieser Lösung ist, dass der Benutzer nichts selbst aufsetzen muss - alles läuft auf den Servern von Amazon, und es steht ein Tutorial für die Inbetriebnahme zur Verfügung. Die Kunden können sehr sorgenfrei die Anzahl der verwendeten Knoten bis zu mehreren Tausend erhöhen oder verringern.

### *c) Datenkataloge für das Gesundheitswesen*

Eine der wichtigsten Komponenten einer Big Data-Plattform ist der Data Lake. Im Gegensatz zur traditionellen Data Warehouse, fließen in ein Data Lake mehrere Arten von Rohdaten ein, von Texten zu Bildern und Videos, ohne dass sie völlig ineinander integriert werden, um diese Daten danach für Analytics Zwecke gut gebrauchen zu können. Um die Benutzung der Daten aus dem Data Lake den Nutzern leichter zu machen, sollten Datenkataloge folgende Funktionalitäten bereitstellen:

- Einfügung: die Nutzer können neue Datensätze einfügen und alte Datensätze editieren.
- Strukturierte Metadaten: Der Katalog muss eine gute Metadaten-Ontologie bereitstellen, mit der die Datensätze einheitlich beschrieben werden können.
- Suche: eine intuitive, semantische, auf den Metadaten basierende Suche muss es den Nutzern erlauben, Datensätze schnell zu finden oder zu entdecken.
- Access Management: Nutzer können sich registrieren, der Zugang zu verschiedenen Datensätzen ist nur für bestimmte Benutzer erlaubt. Die Datensätze und die Benutzer müssen auf Authentizität geprüft werden.
- Services: Beispiele von leichtgewichtigen Anfragen auf den Daten sollten beschrieben sein, um den Nutzern die Benutzung der Daten leichter verständlich zu machen.
- Abstammung: die Verfahren und Quellen, die bei der Erzeugung eines Datensatzes benutzt wurden, müssen klar beschrieben werden.
- Datenschutz: Eine Liste aller nötigen Methoden für Zugangsbeschränkung, Verschlüsselung, Maskierung oder Pseudonymisierung müssen für jeden Datensatz beschrieben werden.
- Datenqualität: Regeln sollten ausgehandelt und beschrieben werden, um die Genauigkeit, Konsistenz, Vollständigkeit und Aktualität der Daten zu sichern.

Um besser zu verstehen, wie ein Datenkatalog für das Gesundheitswesen aussehen kann, können wir uns die meist verbreiteten Datenkataloge ansehen. Diese sind die Open Data-Kataloge. Zwar sind diese größtenteils nur einfache Register und haben nicht alle Funktionalitäten, die ein Datenkatalog für ein Data Lake idealerweise bräuchte, diese stellen aber die Inspiration für Data Lake Datenkataloge dar.

Es gibt eine Vielzahl von Open Data-Kataloge, sie können aber in 3 Kategorien aufgeteilt werden. Einfache Register mit limitierten Annotationen, die nur Datensätze auflisten, die an anderen Stellen

zugänglich sind, zum Beispiel Global Health Exchange. Datenkataloge von Regierungsdaten, die von höherer Qualität sind und mehrere Funktionalitäten bieten wie bessere Metadaten, Suche und Navigation. In diesem Fall werden auch die Daten direkt auf der Plattform gehostet, in manchen Fällen gibt es auch einen Zugangsverwaltungsmechanismus (Beispiel Healthdata.gov). Die dritte Kategorie sind die kommerziellen Datenkataloge, die über die Funktionalitäten von Regierungsportalen hinaus auch noch Dataanalyse -Funktionalität mit anbieten, und auch eine gute Crowdsourcing-Komponente beinhalten (Beispiel Kaggle).

#### i. [Global Health Data Exchange](#)

GHDx ist ein Katalog mit weltweiten Gesundheits- und demographischen Daten, der vom Institute for Health Metrics and Evaluation (IHME) verwaltet wird. IHME ist ein unabhängiges Forschungszentrum, dessen Ziel es ist, die besten Strategien zum Aufbau einer gesünderen Welt zu identifizieren.

Der Katalog bietet eine Auflistung von Datensätzen für eine große Anzahl von Ländern. Zuerst muss ein Land ausgewählt werden und dann können alle verfügbaren Datensätze, die in Verbindung mit der Gesundheit der Bevölkerung stehen, eingesehen werden. Die Datensätze sind nicht auf der Webseite gehostet, lediglich ein Link und eine Beschreibung mit folgenden Informationen ist angeboten: Anbieter, Geographie, geographischer Typ, Zeit/ Periode, Datentyp, Kurzbeschreibung und Schlüsselwörter.

#### ii. [HealthData.gov](#)

HealthData.gov ist eine Webseite der US-Regierung mit dem Ziel hochqualitative Gesundheitsdaten verfügbar zu machen für Forscher, Unternehmer und Politiker, um eine Verbesserung des Gesundheitswesens zu erzielen und gleichzeitig die Wirtschaft anzukurbeln.

Die meisten Datensätze stammen von dem U.S. Department of Health and Human Services (HHS). Die Datensätze beinhalten Informationen von der Qualität der Behandlung von Patienten bis hin zu Medikamenten und Kosten. Eines der Merkmale des Datenkatalogs ist die Bereitstellung von APIs, die rigorose Annotierung mit Ontologien und die besondere Sorge für die Anonymisierung der Daten. Die Daten sind per Stichwort durchsuchbar, können aber auch explorativ durchforstet werden. Die Datensätze selber sind in mehreren Maschinen-lesbaren Formaten verfügbar, wie CSV, RDF, JSON oder XSL und eine Vorschau der Daten ist auch möglich.

#### iii. [Kaggle](#)

Kaggle ist eine Plattform für Data Analytics Wettbewerbe, in der Daten- Wissenschaftler konkurrieren, um die besten Modelle für die Vorhersage und Beschreibung der von Unternehmen und Benutzern hochgeladenen Datensätze zu erstellen. Bei Kaggle Data Sets hat ein Unternehmen oder eine Organisation relativ einfach die Möglichkeit Daten online zu veröffentlichen, zu sehen, wie sie verwendet werden, und mit den Benutzern der Daten zu interagieren. Eine intelligente API zur Datenveröffentlichung und Aktualisierung wird zur Verfügung gestellt für diesen Zweck.

### 3. [Interoperabilität und Standards](#)

Eine wesentliche Herausforderung für das NGD ist die Herstellung von Interoperabilität über die zu Auswertungen herangezogenen heterogenen Quellen. Die erforderlichen Daten müssen identifiziert, aus den Quellsystemen extrahiert, in ein weiterverarbeitbares Format transformiert und in die zur Datenauswertung genutzte Plattform eingespielt werden (ETL = Extract, Transfer and Load).

Fragen der Interoperabilität treten dabei entlang des gesamten Ablaufs auf:

- Für eine Auswertung erforderliche Daten können nur gefunden (Schritt 1) und aus umfangreichen Datensammlungen extrahiert (Schritt 2) werden, wenn die angefragte Quelle die Anfrage entsprechend verarbeiten kann, d. h. weiß, welche Daten überhaupt nachgefragt sind.
- Daten aus mehreren Quellen können nur dann übergreifend ausgewertet werden, wenn sie technisch und semantisch interoperabel sind, d. h. in Bezug auf das verwendete Format und die zugrundeliegende Bedeutung geeignet miteinander verknüpft und ausgewertet werden können (Schritt 3 und 4)

Die Anforderungen an Interoperabilität in den ersten Schritten können auch organisatorisch gelöst werden, da davon auszugehen ist, dass der Verantwortliche einer datenbereitstellenden Organisation notfalls sogar mit einer in Prosa gehaltenen Beschreibungen der benötigten Daten entscheiden kann, ob er diese Daten anbieten kann und in welchen Quellsystemen sich diese befinden.

In folgenden liegt der Fokus daher auf Fragen der Herstellung syntaktischer und semantischer Interoperabilität heterogener Daten, die – aus verschiedenen Quellen stammend – zu einer gemeinsamen Auswertung herangezogen werden sollen.

#### a) *Syntaktische Interoperabilität*

Auf Big Data Analysen ausgelegte Plattformen sind grundsätzlich in der Lage, Daten in heterogenen Formaten zu verarbeiten. Da viele der im Gesundheitswesen gängigen Standards jedoch durchaus auch Semantik über Datenstrukturen (z.B. Sections und Entries in HL7 CDA) und Strukturelemente (z.B. Umsetzung von Messwerten über separate Elemente für Art der Messung, Wert der Messung, Einheit des Messwerts) abbilden, ist es sinnvoll, dass die Daten vor der Auswertung in ein einheitliches Format überführt werden.

Hierzu grundsätzlich geeignet erscheint die gesamte Versionshistorie der HL7 Standards<sup>9</sup>:

- HL7v2: dominierendes Format zum Austausch von Ereignis-Nachrichten (z.B. Aufnahme, Dokumentenfreigabe) im stationären Sektor. Viele konforme aber wenige interoperable Systeme, da der Standard viele Umsetzungsspielräume lässt, die von den Herstellern und Nutzern auch ausgeschöpft werden. Reines Austauschformat, wobei sich einzelne HL7v2 Objektkodierungen in vielen anderen Standards/Profilen wiederfinden, die auch zur Datenspeicherung genutzt werden.
- HL7v3 (insb. CDA): auf einem Referenzmodell (RIM) basierender Standard, der sich sowohl für Nachrichtenkommunikation als auch die Umsetzung strukturierter Dokumente (Clinical Document Architecture) eignet. Trotz vieler Spezifikationen für verschiedene Dokumenttypen (Arztbrief, Medikationsplan, etc.) geringe Verbreitung in der Praxis des deutschen Gesundheitswesens; gute Verbreitung in Österreich (ELGA Leitfäden) und partiell ONC-Vorgabe in den USA (Consolidated CDA). Vorrangig Austauschformat, wobei einzelne Hersteller auch Lösungen zur Speicherung von RIM-basierten Daten anbieten.
- HL7 FHIR: Auf dem Modell vernetzter, granularer Ressourcen basierender Standard, der sowohl zum Austausch als auch zur Speicherung von Gesundheitsdaten geeignet ist. Bindings für verschiedene Formate (XML, JSON, etc.). Breite Unterstützung über Anwendungsszenarien und Sektoren hinweg (RKI DEMIS, Apple Patientenakte, AMTS-Schnittstelle der KBV, Ultrakurzformat für Bundesmedikationsplan, etc.). Strukturierte Möglichkeiten zur Profilierung bieten hohe Flexibilität aber auch Risiken für die Interoperabilität.

Andere (nationale) Formate wie z.B. xDT kommen zwar grundsätzlich auch in Frage, decken aber nicht das komplette Spektrum medizinischer Daten ab und/oder sind stark auf einzelne Szenarien fokussiert.

Von den genannten Standards bietet sich aus mehreren Gründen vor allem HL7 FHIR als universelles Format zur Übertragung von Daten aus Quellsystemen zur Analysedatenbank des NGD an:

- Über die standardisierten Ressourcen lässt sich ein breites inhaltliches Spektrum abdecken.
- Über das JSON-Binding können auch große Datenmengen effizient und ggf. als Stream übertragen werden.
- Durch die Möglichkeit, eigene Profile und Datensatzdefinitionen mit den Sprachmitteln von FHIR zu beschreiben, können auch Erweiterungen automatisiert verarbeitet werden.
- Datenobjekte können über mehrere Datenbanken hinweg vernetzt und gespeichert werden.

<sup>9</sup> <http://hl7.de/>, Zugriff am 27. Juli 2018

Hinzu kommt, dass sich in den Quellsystemen in spaltenorientierten Datenbanken verwaltete Daten sehr einfach auf das HL7 FHIR Format abbilden lassen, d. h. eine Transformation in dieses Format aus den Quellsystemen heraus als unproblematisch zu sehen ist.

### b) Semantische Interoperabilität

Medizinische Daten liegen in den Quellsystemen im Idealfall kodiert vor, d. h. werden über Konzepte ausgedrückt, deren Semantik über einen Standard definiert ist. Während beispielsweise ein Begriff wie „Fieber“ unterschiedliche Interpretationen zulässt (Google bietet zwischen 37,6° als 38,0° als untere Grenze an), kann mit einer Kodierung eine Definition des Temperaturbereichs verknüpft werden, der als „Fieber“ zu verstehen ist. Nur wenn Daten aus unterschiedlichen Quellsystemen auf ein solches gemeinsames Verständnis bezogen werden können, ist es möglich, auf diesen zu suchen, diese zu filtern und diese mit anderen Datensätzen zu verknüpfen.

Im deutschen Gesundheitswesen sind unzählige Codesysteme im Einsatz. Viele davon sind nur in einzelnen Teilbereichen des Gesundheitswesens gängig und noch mehr sind schlichtweg proprietär, d.h. wurden z.B. von einem Hersteller definiert und sind nur in den Systemen dieses Herstellers im Einsatz.

Im Rahmen der Empfehlungen für die im schweizerischen Patientendossier zu verwendenden Standards wurde eine einfache Fallstudie ausgearbeitet, bei der eine Patientin mit unterschiedlichen

Textbausteine aus dem Fallbeispiel	Datemodul: Probleme						Datemodul: Behandlung				
	ePA-AC	RAI-HC Problembereich	RAI-HC Auslöser/Alarmzeit	NANDA-I	ICNP	SNOMED CT	Leistungskatalog Spitex (RAI-HC)	LEP Nursing	ICNP	SNOMED CT	
Gemäss Überweisungsrapport benötigt Frau Muster noch Hilfe beim Gehen.	E3_I_0003 gering beeinträchtigte Fähigkeit sich fortzubewegen	1 Aktivitäten des täglichen Lebens (BADL) und Rehabilitationspotential	H2c Gehen	00088 Beeinträchtigte Gehfähigkeit	10001046 beeinträchtigt Gehvermögen	228158008 Walking disability	10505 Hilfe beim Gehen	L_22810 Wegstrecke begleiten	10038986 Assisting With Walking	62013009 Ambulating patient	
Weil sie den Unterkörper und den Rücken nicht alleine waschen kann, benötigt sie auch bei der Körperpflege Hilfe.	E3_I_0108 stark beeinträchtigte Fähigkeit die Körperpflege im Bereich des Unterkörpers durchzuführen	1 Aktivitäten des täglichen Lebens (BADL) und Rehabilitationspotential	H2i Persönliche Hygiene	00102 Selbstversorgungsdefizit Körperpflege	10000987 Beeinträchtigte Fähigkeit Körperpflege durchzuführen	228153004 Personal hygiene disability	10104 Teilwäsche am Lavabo (inkl. Intimpflege)	L_23078 Teilkörperwäsche durchführen	10044803 Waschen des Patienten	225429000 Personal hygiene interventions (procedure)	
Eine Schulung zur Sturzprävention zu Hause wird als weiterführende Massnahme im Überweisungsrapport empfohlen.	E0_I_080 Sturzrisiko	15 Stürze	K5 Stürze K6a Unsicherer Gang	00155 Sturzgefahr	10015122 Risiko zum Sturz	129839007 At risk for falls	10909 Pflegeanleitung/Beratung Klientin oder Angehörige	L_23331 Sturzberatung durchführen	10040253 unterrichten über Sturzprävention	710580007 Education about fall prevention	

Abbildung 3: Kodierung eines medizinischen Sachverhalts in unterschiedlichen Standards

Sektoren des Gesundheitswesens in Kontakt kommt. Die nachfolgende Abbildung stellt dar, wie derselbe Sachverhalt über unterschiedliche semantische Standards ausgedrückt werden kann<sup>10</sup>:

Die Situation in Deutschland ist vergleichbar. Einheitliche Kodierungen sind nur da zu finden, wo diese für die Leistungsabrechnung vorgegeben sind (z.B. ICD-10, OPS, KBV-Schlüsseltabellen) oder historisch in die Systeme „hineingewachsen“ sind. Für das NGD muss daher davon ausgegangen werden, dass von wenigen Ausnahmen abgesehen alle medizinischen Informationen in hochgradig heterogenen Kodierungen aus den Quellsysteme bereitgestellt werden.

Bei Betrachtungen dieses Themas liegt der Fokus zumeist sehr stark auf „prominenten“ Fragestellungen wie z.B. Diagnosen, Prozeduren, Medikationen, Allergien, etc. die sich in jeder Patientenakte und jedem Primärsystem finden lassen. Neben diesen gibt es jedoch auch noch unzählige Scores und Schlüsseltabellen aus den einzelnen medizinischen Fachdisziplinen, die oftmals nur wenige Werte beinhalten, wo aber seitens der Hersteller und/oder Anwender eine große Kreativität an den Tag gelegt wurde, wie sich eine solche Handvoll von Werten kodieren lässt. Die Zahl dieser Codesysteme – die alle irgendwo im praktischen Einsatz sind und sich in irgendwelchen Datenbanken wiederfinden – liegt alleine in Deutschland sicherlich im fünfstelligen Bereich.

### c) Anforderungen an ein Referenz-Codesystem

Das Referenz-Codesystem (bzw. die dieses aufspannenden Codesysteme) sollten idealerweise auf vier Ebenen Interoperabilität herstellen können:

<sup>10</sup> aus eHealth Suisse: Fallbeispiel zur semantischen Interoperabilität von Pflegedaten. 2016. Onlione unter [https://www.e-health-suisse.ch/ch/fileadmin/user\\_upload/Dokumente/2017/D/170213\\_Fallbeispiel:Interoperabilitaet\\_Pflege\\_D.pdf](https://www.e-health-suisse.ch/ch/fileadmin/user_upload/Dokumente/2017/D/170213_Fallbeispiel:Interoperabilitaet_Pflege_D.pdf)

#### i. Domänenmodell

Es muss davon ausgegangen werden, dass Daten in unterschiedlicher Detailtiefe vorliegen oder dass benötigte Daten nur durch Ableitung aus anderen Informationen gewonnen werden können. Das Domänenmodell greift hier auf Ontologien zurück, die über Standard-Beziehungen (Generalisierung/Spezialisierung, Aggregation, etc.) die semantischen Zusammenhänge zwischen Konzepten beschreiben.

- Beispiel: Für eine Auswertung sollen Daten von Patienten zusammengeführt werden, die einen Knochenbruch an einer Extremität hatten. Das Domänenmodell kann die Anfrage semantisch erweitern, so dass auch Daten erfasst sind die granularer als Knochenbruch des linken/rechten Arms/Beins ausgezeichnet sind.

#### ii. Konzeptkoordination

Aus prä-kooordinierten Konzepten lassen sich weitere Informationen zur Semantik des Konzepts ableiten. Dies erleichtert ein automatisiertes Mapping und verbessert die Qualität der Mappings. Über eine post-Koordination können neue Konzepte gebildet werden. Dies vereinfacht die Erweiterung des Referenz-Systems.

- Beispiel ICD-10: Die Konzepte für Frakturen können um .0 oder .1 erweitert werden, um eine offene oder geschlossene Fraktur zu kennzeichnen (einfache Post-Koordination).
- Beispiel LOINC: Konzepte sind über bis zu 6 Achsen definiert, z.B. kann hat derselbe Erreger über unterschiedliche Konzepte repräsentiert sein, je nachdem welches Nachweisverfahren genutzt wurde (Prä-Koordination)

#### iii. Konzeptdefinition

Konzepte sind implizit oder explizit mit einer klaren Definition versehen, die sie von anderen Konzepten abgrenzt und Hinweise zur Nutzung gibt. Erst hierdurch wird erkennbar, ob zwei Codes mit demselben Bezeichner auch wirklich auf dasselbe Konzept referenzieren:

- Beispiel Schlüsseltabellen der KBV: Schlüsseltabellen der KBV sind oftmals einem Formular oder einer ärztlichen Tätigkeit zugeordnet. Hierdurch ergibt sich (implizit) der Kontext in dem die enthaltenen Konzepte zu interpretieren und zu nutzen sind.

#### iv. Spezifität/Sensitivität

Konzepte sollten soweit differenziert sein, dass sich daraus auch zumindest qualitative Aussagen zur Spezifität und/oder Sensitivität der damit kodierten Versorgungsdaten ausdrücken lassen. Hierdurch können Daten aus externen Systemen mit Qualitätsindikatoren versehen werden, die sich aus der Quelldatenbank ableiten.

- Beispiel LOINC: Konzepte zu Erregern enthalten auch immer eine Information zum Untersuchungsgegenstand und zum Nachweisverfahren.

Wie diese Aufstellung zeigt, erscheinen insbesondere mehrachsige, prä-kooordinierte Codesysteme als Referenzsysteme geeignet. Diese haben darüber hinaus den Vorteil, dass die durch ihre Mehrachsigkeit sehr ausdrucksmächtig in Bezug auf die Definition intensionaler Wertemengen (Value Sets) sind, bei denen für einen Anwendungsfall zulässige Werte über Regeln auf Achsen und Konzeptbeziehungen (Relationen) definiert werden (z.B. „Alle Unterkonzepte von [Fraktur] bei denen der Wert der Achse/Eigenschaft [Lokalisierung] ein Unterkonzept von [Extremität] ist“).

Standardisierte Codesysteme von signifikantem Umfang, die die benannten Eigenschaften erfüllen, sind lediglich SNOMED CT und LOINC. Darüber hinaus bieten verschiedene Hersteller eigene Referenz-Systeme an, die die benannten Eigenschaften besitzen (z.B. DIACOS von ID Berlin).

#### d) SNOMED CT

Das kontrollierte Vokabular *SNOMED CT*<sup>11</sup> (*Systematized Nomenclature of Medicine - Clinical Terms*) basiert auf den Konstrukten für multilinguale Thesauri und leichtgewichtige Ontologien. Es bietet

<sup>11</sup> <https://www.snomed.org/snomed-ct>, Zugriff am 27.7.2018

medizinische Begriffe für Krankheiten, Befunde, Substanzen etc. eingeordnet unter z.Z. etwa 18 Top-Level-Konzepten.

Struktur	polyhierarchische Taxonomie Standard- und domänenspezifische Konzeptrelationen Bietet eine Vielzahl von Konzeptrelationen der Domäne.
Lexikalische Elemente	Preferred Terms Synonyme
Ausdrücke	Prä- und Post-koordinierte Ausdrücke Hohe Mächtigkeit durch Angabe von Ausdrücken einer Untermenge der Description Logic (bzw. des OWL-Profiles EL). Dies macht Formulierungen wie "Die Krankheit Schnupfen ist definiert als eine Krankheit mit dem Erreger virus122 an dem Ort Nase und ..." möglich.
Umfang	mehr als 300.000 Konzepte und 1,3 Mio. Relationen.
Lizenzen	Kostenpflichtig über IHTSDO.

#### e) LOINC

Die Codeliste LOINC<sup>12</sup> (*Logical Observation Identifiers Names and Codes*) wird vor allem zur Identifikation von Test- und Untersuchungsergebnissen aus Laboreinrichtungen und Kliniken sowie zur Auszeichnung von Elementen in strukturierten Dokumenten genutzt.

Struktur	Konzeptliste keine Konzeptrelationen Streng genommen ist ein Konzept bei LOINC gleich einem Bezeichner für einen Tupel aus einem mehrdimensionalen Raum, weshalb als Standardrepräsentation eine einfache Datenbanktabelle verwendet wird. LOINC verwendet mindestens die sechs Dimensionen (Achsen): Komponente, Messgröße, Zeitangaben, Art der Probe, Skalentyp, Methode. Z.B. (GLUCOSE, SCNC (Substanzkonzentration), 24H, SER (Serum), QN (quantitativ), AGGL (Agglutination)).
Lexikalische Elemente	Preferred Terms Synonyme
Ausdrücke	Prä-koordinierte Ausdrücke LOINC-Konzepte können als prä-koordinierte Ausdrücke angesehen werden, wenn die Werte der Dimensionen als Konzepte interpretiert werden.
Umfang	ca. 60.000 Konzepte
Lizenzen	Frei verfügbar über loinc.org bzw. Regenstrief als Datenbanktabelle.

#### f) Data Dictionary

Codesystemen und (über ein Referenz-System realisierte) Mappings spannen ein Netzwerk aus Konzepten, Konzeptbeziehungen (relations) und Konzepteigenschaften (properties) auf. Konzeptbeziehungen bilden dabei semantische Bezüge ab, sind aber häufig zu unkonkret, um über Identitätsbeziehungen hinausgehende Transformationen zu erfassen.

Beispiel: In einer Analyse zu einer Fragestellung aus dem Bereich „Diabetes“ werden aus verschiedenen Quellen Blutzuckermesswerte und Angaben zur zugeführten Nahrung zusammengeführt. Je nach Quelle liegen die Blutzuckerwerte in der Einheit mMol/l oder mg/dl vor ( $\text{mmol/l} \times 18,02 = \text{mg/dl}$ ). Die Angaben zur Nahrung liegen in Kohlenhydraten, Kohlenhydrateinheiten (1 KE = 10 g

<sup>12</sup> <https://loinc.org/>, Zugriff am 27.7.2018

Kohlenhydrat) oder Broteinheiten (1 BE = 12 g Kohlenhydrat) vor. Aus dem Referenzsystem ist zunächst nur ableitbar, dass z.B. KH, KE und BE zur Beschreibung desselben Sachverhalts genutzt werden.

Um diese Daten jedoch gemeinsam auswerten zu können, ist eine Umrechnung auf eine Referenzeinheit erforderlich. Referenzeinheiten und Umrechnungsformeln können über ein Data Dictionary (z.B. gemäß ISO 11179) verwaltet werden. Dieses kann an das Referenzsystem angebunden werden und dieses um automatisch ausführbare Transformationsregeln erweitern.

#### D. Stakeholder-Beteiligung

In verschiedenen Vorgesprächen mit Stakeholdern (Krankenversicherung, Klinisches Krebsregister), auch teilweise bereits vor Einreichung des Antrages auf Unterstützung der Machbarkeitsstudie, wurde deutlich, dass es eine niederschwelligere Möglichkeit der Datenzusammenführung, wie es das Konzept des Netzwerk Gesundheitsdaten (vormals healthnet.org) vorsieht, hilfreich und notwendig für eine effektivere und bessere Gesundheitsversorgung ist. Im Rahmen eines Stakeholder-Meetings wurde seitens der Stakeholder nochmals das Besondere des Konzeptes betont, nämlich Daten im Sinne einer „Datenbörse“ zugänglich zu machen, ohne dass einzelne Institutionen die Daten dauerhaft abgeben müssten und ohne dass die Daten komplett auf einem zentralen „Mega-Server“ gespeichert werden müssten. Der „Charme“ des Lösungsansatzes wurde darin gesehen, dass bereits existierende Daten verarbeitende Institutionen wie RKI oder DIMDI die Daten weiterhin im Rahmen der eigenen Aufgabenstellung weiterverarbeiten können und damit bereits vorhandene und gut entwickelte Strukturen weiterhin Bestand haben können. An diesem Stakeholder-Meeting beteiligt war neben den Erstellern der Machbarkeitsstudie Vertreter des RKI, DIMDI, Klinischer Krebsregister und aus der Krankenversicherung.

Relevant bei einer Involvierung der Stakeholder, bzw. der künftig das NGD nutzenden Institutionen ist die rechtliche Formulierung und der Auftrag über den diese Institutionen in die Lage versetzt wurden, Daten zu sammeln, Daten zu erheben, Daten zu analysieren und nicht nur Analyseergebnisse zur Verfügung zu stellen, sondern auch die Daten selber zur Nutzung zur Verfügung zu stellen.

Im Vorfeld und im Verlauf der Machbarkeitsstudie wurde der Austausch mit einigen potentiellen Stakeholdern (wie RKI, DIMDI, PKV und klinisches Krebsregister) gesucht. Bei durchweg positiver Resonanz, einen kontrollierten Zugang zur Datennutzung im Sinne einer „Börse“ zu schaffen, wurde das Ziel, Daten passager für eine Datenanalyse zusammenführen zu können, weiterverfolgt.

Als Stakeholder betrachtet werden können auch die Dachverbände der potentiellen Nutzer und deren Wirtschaftsbereiche, da über diese Verbände, Weiterentwicklungen zur Nutzarmachung von Daten und die damit verbundenen rechtlichen und technischen Voraussetzungen der das NGD nutzenden Institutionen betrieben werden können.

Mögliche Stakeholder können damit sein:

##### *Robert Koch Institut (RKI)*

Das Robert Koch-Institut ist das Public-Health-Institut für Deutschland. Ziel ist es, die Bevölkerung vor Krankheiten zu schützen und ihren Gesundheitszustand zu verbessern. Es wurde 1891 gegründet, ist eine der ältesten biomedizinischen Forschungseinrichtungen der Welt.

Es ist ein Merkmal des Instituts, neue Herausforderungen zu erkennen und anzunehmen, die eigenen Ansätze immer wieder kritisch zu hinterfragen und neue Wege zu wagen. Es versteht sich als Institut für die Gesundheit der Bevölkerung.

Wissenschaftlerinnen und Wissenschaftler des RKI erheben Daten zu nicht-übertragbaren Krankheiten wie Diabetes und Krebs, zu Infektionskrankheiten und (neuen) biologischen Gefahren. Darauf basierend werden konkrete Empfehlungen und Präventionskonzepte entwickelt. Die Arbeit steht auf zwei Säulen: Forschung, die Daten für Entscheidungen bereitstellt, und unabhängige, verlässliche Beratung der Fachöffentlichkeit, insbesondere des öffentlichen Gesundheitsdienstes und der Politik in Deutschland.

Die wissenschaftlichen Fragen ändern sich – und mit ihnen die Methoden, mit denen Daten zu Gesundheit und Krankheit erhoben und interpretiert werden können. Zusätzlich zu bewährten Ansätzen werden daher kontinuierlich neue Herangehensweisen entwickelt, etwa in den Bereichen digitale Epidemiologie, künstliche Intelligenz und zielgruppenspezifische Kommunikation. Weil nur hochwertige Daten zu verlässlichen Ergebnissen und Empfehlungen führen, wird die Qualität der wissenschaftlichen Arbeit und Infrastruktur regelmäßig intern und extern bewertet.

Das RKI treibt die Vernetzung in einem größeren Rahmen voran, mit dem öffentlichen Gesundheitsdienst, verschiedenen Public-Health-Akteuren, Tier-, Human- und Umweltmedizinerinnen und internationalen Institutionen.

Das RKI veröffentlicht seine Forschungsergebnisse in anerkannten Fachzeitschriften, stellt sie auf Tagungen vor und nutzt sie für die Gesundheitsberichterstattung des Bundes. Die Empfehlungen des RKI und der beim RKI angesiedelten Kommissionen werden frei verfügbar im Internet bereitgestellt. Auch die Öffentlichkeit wird regelmäßig über Aktivitäten des RKI informiert. Das RKI fühlt sich der Förderung des weltweiten Zugangs zu Wissen verpflichtet. Darüber hinaus werden eigene Methoden und Daten zur Verfügung gestellt.

Das Robert Koch-Institut erfüllt außerdem Aufgaben für die Weltgesundheitsorganisation (WHO) und das Europäische Zentrum für die Prävention und Kontrolle von Krankheiten (ECDC) und koordiniert eine Reihe internationaler Forschungsprojekte.

Gesetzliche Vorgaben, zum Beispiel zum Datenschutz, zur Biosicherheit, zur Sicherheit in der Gentechnik und zum Tierschutz, werden effektiv umgesetzt. Mit Forschungsrisiken wird verantwortungsvoll und Wert darauf gelegt, die Regeln guter wissenschaftlicher Praxis einzuhalten. Als Obere Bundesbehörde im Geschäftsbereich des Bundesgesundheitsministeriums ist das RKI nicht auf Finanzierung durch die Industrie angewiesen. Interessenskonflikte und Forschungsprojekte, die nicht öffentlich finanziert werden, werden als solche kenntlich gemacht.

**Bewertung als Stakeholder NGD:** Das RKI erfüllt die notwendigen Konditionen für eine NGD-Nutzung in der institutseigenen Aufgabenstellung. Kenntnisse zur Umsetzung der technischen Konditionen sind vorhanden, da der Umgang mit Datenspeicherung, institutsinternem Datenaustausch und der Daten-Auswertung geläufig ist, die technischen Voraussetzungen zur externen Datenzusammenführung müssten den Bedingungen angepasst werden.

#### *Klinische Krebsregister, Arbeitsgemeinschaft Deutscher Tumorzentren (ADT)*

Ein klinisches Krebsregister strebt die Erfassung aller Krebspatientinnen und Krebspatienten an, die an einer bestimmten Einrichtung bzw. einem Einrichtungsverbund oder (idealerweise) in einem definierten Einzugsgebiet behandelt wurden. Seit dem Krebsfrüherkennungs- und registergesetz (KFRG, 2013) werden deutschlandweit klinische Krebsregister aufgebaut. Hauptziele dieser Erfassung sind die Qualitätssicherung und die Darstellung der Ergebnisqualität der gesamten Behandlung von Krebspatienten. Zusätzlich werden die zuständigen epidemiologischen Landeskrebsregister unterstützt. Einige klinische Register, sogenannte Nachsorgerleitetstellen, organisieren auch die Krebsnachsorge.

Durch die einrichtungsübergreifende Zusammenführung und Auswertung von standardisierten Informationen zu Diagnostik, Therapie und Verlauf von Tumorerkrankungen bearbeitet ein klinisches Krebsregister folgende Aspekte der Behandlung:

- Ergebnisqualität: Überlebenszeit, tumorfreie Zeit, Lebensqualität
- Prozessqualität, z.B. Einhaltung von Therapiestandards, Einhaltung von interdisziplinär beschlossenen Maßnahmen, etc.
- Strukturqualität, z.B. Therapieerfahrung (Häufigkeit bestimmter Operationen)

Die Nachverfolgung der entlassenen Tumorpatienten und die Rückmeldung aktueller Überlebenszeit- und Verlaufsdaten an die primären Behandler bilden einen Arbeitsschwerpunkt klinischer Krebsregister.

Regelmäßige Auswertungen und Analysen der Daten können Verbesserungspotentiale aufdecken und die Grundlagen für geeignete Verbesserungsmaßnahmen bereitstellen.

**Bewertung als Stakeholder NGD:** Klinische Krebsregister erfüllen die notwendigen Konditionen für eine NGD-Nutzung in der institutseigenen Aufgabenstellung bis auf die rechtliche Freigabe der Daten

zur Nutzung zusätzlich im NGD und nicht nur beim Krankenversicherer. Kenntnisse zur Umsetzung der technischen Konditionen sind vorhanden, da der Umgang mit Datenspeicherung, institutsinternem Datenaustausch und der Daten-Auswertung geläufig ist und Analyseergebnisse an die primären Behandler zurückübermittelt werden. Die technischen Voraussetzungen zur externen Datenzusammenführung müssten den Bedingungen angepasst werden.

*Epidemiologische Krebsregister, Gesellschaft der epidemiologischen Krebsregister in Deutschland (GEKID)*

Im Januar 1996 wurde die „Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland“ gegründet, in der alle epidemiologischen Krebsregister Deutschlands sowie die im Robert-Koch-Institut angesiedelte Dachdokumentation Krebs zusammenarbeiten. Diese Arbeitsgemeinschaft war auch dem damaligen Gesamtprogramm zur Krebsbekämpfung der Bundesregierung assoziiert.

Im April 2004 wurde die Arbeitsgemeinschaft in die „Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID)“ überführt.

Vorrangige Aufgabe dieser Gesellschaft ist es, trotz unterschiedlicher landesgesetzlicher Regelungen bundesweit eine weitgehende methodische Einheitlichkeit durch inhaltliche Standards zu erlangen. Nur durch eine deutschlandweite Zusammenarbeit kann die Vergleichbarkeit der Ergebnisse der Krebsregister gewährleistet werden. Darüber hinaus ist GEKID ein gemeinsamer Ansprechpartner der epidemiologischen Krebsregister bei länderübergreifenden Fragestellungen.

Die klinischen Krebsregister erfassen die Krankheitsverläufe der Patienten. Registriert werden Daten von der Diagnose über einzelne Behandlungsschritte und die Nachsorge bis hin zu Rückfällen (Rezidiven); ebenso gesammelt werden Informationen zum Überleben der Patienten. Durch die erfassten Daten wird darüber hinaus sichtbar, ob medizinische Leitlinien beachtet werden oder ob Unterschiede in der Qualität der Behandlung bestehen. Mit der regelmäßigen Rückmeldung der Auswertungsergebnisse an die behandelnden Ärzte und Kliniken können Verbesserungsprozesse in der Krebsbehandlung angestoßen werden.

Weitere Ziele sind:

- Ansprechpartner sowohl für nationale und internationale Kooperationspartner als auch für die interessierte Öffentlichkeit
- Aufbereitung und Publikation der Ergebnisse der Krebsregistrierung in Deutschland
- Vertretung der epidemiologischen Krebsregister im Rahmen des Nationalen Krebsplans
- Information über den Stand der Krebsregistrierung in Deutschland und Vermittlung der Ziele epidemiologischer Krebsregistrierung
- Beiträge zur Sicherstellung der Vollzähligkeit der einzelnen Krebsregister über gemeinsame Informationsaktivitäten
- Definition inhaltlicher Standards als Grundlage der Vergleichbarkeit epidemiologischer Krebsregister
- Koordination registerübergreifender Aufgaben sowie Pflege des Kontakts mit der klinischen Tumordokumentation
- Initiierung gemeinsamer Forschungsaktivitäten
- Förderung der wissenschaftlichen Nutzung der bevölkerungsbezogenen Krebsregister
- Nutzung der Daten zur Qualitätssicherung in der onkologischen Versorgung

Der vor mehr als drei Jahren begonnene Auf- und Umbau entsprechender regionaler Strukturen kam jedoch nur schleppend voran (Prognos-Gutachten i.A. des GKV-Spitzenverbandes). Neben fehlenden Landesgesetzen und noch nicht arbeitsfähigen Einrichtungen in einigen Bundesländern listete das Gutachten vor allem Probleme bei der Datenerfassung, Datenverarbeitung sowie bei der länderübergreifenden Datenvernetzung der Register auf.

Bewertung als Stakeholder NGD: Epidemiologische Krebsregister erfüllen die notwendigen Konditionen für eine NGD-Nutzung in der institutseigenen Aufgabenstellung. Kenntnisse zur Umsetzung der technischen Konditionen sind vorhanden, da der Umgang mit Datenspeicherung, Datenzusammenführung verschiedener Krebsregister und der Daten-Auswertung geläufig ist. Daten

werden neben einer bundeslandübergreifenden Analyse zur Qualitätssicherung in der onkologischen Versorgung genutzt. Die juristischen und technischen Voraussetzungen zur externen Datenzusammenführung müssten den Bedingungen angepasst werden.

#### *Arbeitsgemeinschaft Deutscher Tumorzentren (ADT)/ Tumorzentren*

Aufgabenbereiche sind einmal das Produzieren von Wissen durch die Durchführung von diagnostischen und klinischen Studien, Therapieoptimierungen, Versorgungsforschung und die Förderung und Begleitung der Grundlagenforschung.

Ein weiterer Bereich umfasst das Propagieren von Wissen durch die Aus-, Fort- und Weiterbildung, Beratung von Ärzten, Mitarbeiter(innen) im Gesundheitswesen, Patienten und Angehörigen, interdisziplinäre Fallbesprechungen, Akzeptanz, Implementierung und Rückmeldung von Leitlinien.

Der dritte Bereich widmet sich der Prüfung der Umsetzung des aktuellen Wissens: Verlaufsbegleitende Dokumentation der Diagnose-, Therapie- und Follow-up Behandlung, Prüfung der vorhandenen Qualitätsindikatoren, regionale Rückmeldung und Diskussion der Struktur-, Prozess- und Ergebnisqualität. Gezielte Intervention zur flächendeckenden Verbesserung.

Charakteristika von Tumorzentren sind:

- Definiertes Einzugsgebiet
- Sektorenübergreifende Dokumentation aller Krankheitsverläufe (Klinisches Krebsregister)
- Mitwirkung aller Versorgungsträger sektorenübergreifend
- Akzeptanz wissenschaftlich begründeter Leitlinien
- Interdisziplinär kooperierende themenbezogene Arbeitsgruppen
- Interdisziplinäre Strukturen für Diagnostik und Behandlung
- Regelmäßige interdisziplinäre Konferenzen mit Fallvorstellungen
- Angebote für überregionale onkologische Konsiliardienste
- Fortbildung für alle mit Krebs befassten Berufsgruppen
- Vereinigung von optimaler Patientenversorgung und klinischer Forschung
- Bewertung der Versorgungsqualität im Einzugsgebiet

Bewertung als Stakeholder NGD: Tumorzentren (stellvertretend für andere klinische Zentren) erfüllen die notwendigen Konditionen für eine NGD-Nutzung in der institutseigenen Aufgabenstellung. Therapiebezogene Daten werden Diagnose- und Therapie-spezifisch kodiert (ICD-, ICPM-/ OPS-Kodes), Klinik-intern gespeichert, an die KVen zu Abrechnungszecken weitergeleitet und im Rahmen von Studien genutzt. Abhängig von der Diagnosestellung werden die Daten auf Basis der Gesetzgebung (z.B. KFRG) an Register weitergeleitet. Kenntnisse zur Umsetzung der technischen Konditionen zu Datenerhebung, Datenspeicherung, Datenzusammenführung und Datenweiterleitung sind entsprechend vorhanden.

Die juristischen und technischen Voraussetzungen zur externen Datenzusammenführung müssten den Bedingungen für eine NGD-Nutzung angepasst werden.

#### *Deutsches Institut für Medizinische Dokumentation und Information (DIMDI)*

Das DIMDI ist ein Institut im Geschäftsbereich des Bundesministeriums für Gesundheit, gibt in seinem Auftrag amtliche medizinische Klassifikationen heraus und stellt weitere Terminologien und Standards für das Gesundheitswesen bereit. Das DIMDI gibt Klassifikationen zur Kodierung von Diagnosen und Operationen heraus und pflegt weitere medizinische Bezeichnungssysteme. Außerdem betreibt es Informationssysteme für Arzneimittel, Medizinprodukte und Versorgungsdaten sowie zur Bewertung gesundheitsrelevanter Verfahren (Health Technology Assessment, HTA). Es erstellt Klassifikationen, Terminologien und Standards im Gesundheitswesen. Die Datenbanken beim DIMDI enthalten keine statistischen Daten. Recherche-Möglichkeiten bestehen zu Informationen zu Arzneimitteln, Medizinprodukten und Health Technology Assessment. In den Datenbanken des DIMDI kann mit oder ohne Vertrag recherchiert werden. Es kann die Auswertung von Versorgungsdaten der gesetzlichen Krankenkassen beantragt werden.

Welche Institutionen Daten für welchen Zweck erhalten dürfen, regelt das Gesetz (§§ 303a bis 303e Sozialgesetzbuch V). Nur dort genannte Institutionen können zukünftig Versorgungsdaten über das DIMDI nutzen. Dazu gehören unter anderem bestimmte Einrichtungen der Krankenkassen, der Gemeinsame Bundesausschuss, Interessenvertretungen der Patienten und der Leistungserbringer auf Bundesebene sowie Institutionen der Forschung und Gesundheitsberichterstattung. Ihnen sollen die Daten z.B. helfen, die Qualität der Versorgung zu verbessern oder Leistungsressourcen zu planen.

Finanziert wird das Informationssystem durch die gesetzlichen Krankenkassen. Für die Aufbereitung der Daten durch das DIMDI fallen Nutzungsgebühren an. Näheres regelt die Gebührenverordnung. Die DaTraV-Daten aus dem Informationssystem Versorgungsdaten (Datentransparenz) beinhalten u.a. Angaben über ambulante und stationäre Diagnosen sowie zur ambulanten Arzneimittelversorgung der gesetzlich Versicherten. Diese können insbesondere für Analysen des Versorgungsgeschehens im Rahmen der Versorgungsforschung und für Steuerungsaufgaben in der gesetzlichen Krankenversicherung genutzt werden.

Daten im Rahmen der Datenfernverarbeitung unter Verwendung von SQL (Oracle 11g R2) und SAS (9.4 64bit) zu analysieren. Dabei werden die DaTraV-Daten mit von den Nutzungsberechtigten bereitgestellten Programmen ausgewertet, die wir vorab geprüft und genehmigt haben. DaTraV-Daten entsprechend einer vom Antragsteller formulierten Fragestellung ergänzt um Angaben zur Methodik und zur Struktur der gewünschten Ergebnismenge auszuwerten. D.h. wir entwickeln in diesem Fall die erforderlichen Auswertungsskripte in SQL. In der dritten Ausbaustufe wird es möglich sein eine On-Site-Nutzung zu beantragen bei der Nutzungsberechtigte die pseudonymisierten Einzeldatensätze in den Räumlichkeiten des DIMDI auswerten. Voraussetzung ist, dass eine Analyse auf Basis von anonymisierten Daten nicht möglich ist.

Der Datenbestand zur Versorgungsforschung stammt vom Bundesversicherungsamt (BVA). Das BVA wiederum erhält die Daten von den Krankenkassen für den morbiditätsorientierten Risikostrukturausgleich (Morbi-RSA). Die ans DIMDI übermittelten Morbi-RSA-Daten werden über mehrere Jahre hinweg zusammengeführt. Derartig vollständig waren diese Angaben bisher nicht verfügbar.

Datenschutz: Es wird sichergestellt, dass einzelne Versicherte nicht identifizierbar sind. Dazu sind strenge Datenschutzvorkehrungen in Abstimmung mit der Bundesbeauftragten für den Datenschutz und die Informationsfreiheit getroffen worden. So werden beispielsweise alle Versicherungsnummern durch Pseudonyme ersetzt.

Technische Voraussetzungen für die Recherche beim DIMDI: Für die via Internet zugängliche Datenbankrecherche beim DIMDI wird ein aktueller Webbrowser (z.B. Mozilla Firefox, Microsoft Internet Explorer), der den HTML 4.0-Standard, Cascading Style Sheets und JavaScript unterstützt, benötigt.

Zur Unterstützung nachhaltiger Entscheidungen sollten vorhandene Informationen aus möglichst allen betroffenen Bereichen umfassend analysiert werden. Modelle und Szenarien werden in eine strukturierte und systematische Bewertung einbezogen.

Bewertung als Stakeholder NGD: Das DIMDI erfüllt die notwendigen Konditionen für eine NGD-Nutzung in der institutseigenen Aufgabenstellung. Kenntnisse zur Umsetzung der technischen Konditionen sind vorhanden, da der Umgang mit Datenspeicherung, Datenzusammenführung von BVA-Daten und der Daten-Auswertung geläufig ist. Daten werden für eine weiterführende Analyse zur Verfügung gestellt. Die juristischen und technischen Voraussetzungen zur externen Datenzusammenführung müssten den Bedingungen angepasst werden

*Spitzenverband Bund der Krankenkassen (gkv-spitzenverband)*

In Deutschland sind rund 90 Prozent der Bevölkerung (ca. 72 Mio. Versicherte) in der gesetzlichen Krankenversicherung (GKV) versichert und haben Anspruch auf Leistungen, um ihre Gesundheit zu erhalten und wiederherzustellen oder den Gesundheitszustand zu verbessern.

Die GKV wird getragen von gesetzlichen Krankenkassen, die als Körperschaften des öffentlichen Rechts finanziell und organisatorisch unabhängig sind. Sie führen die ihnen staatlich zugewiesenen

Aufgaben eigenverantwortlich durch. Die gesetzlichen Krankenkassen haben als Träger der GKV das Recht zur Selbstverwaltung. Weite Teile der gesundheitlichen Versorgung werden durch die Selbstverwaltung über Verträge mit den Leistungserbringern gestaltet.

Heute gibt es nur noch 110 Kassen (Stand: 01. Januar 2018).

Eine weitere wichtige Grundlage ist das Selbstverwaltungsprinzip. Dabei wird der Staat durch Delegation von Aufgaben und Verantwortungsbereichen an die Träger entlastet (Subsidiaritätsprinzip). Das heißt, dass die Träger der Sozialversicherung als öffentlich-rechtliche Körperschaft alle Steuerungsaufgaben in Eigenverantwortung unter Rechtsaufsicht des Staates erfüllen. Selbstverwaltung wird durch gewählte Vertreter der Versicherten und der Arbeitgeber ehrenamtlich ausgeübt. Eine Ausnahme besteht bei den Ersatzkassen, hier sind nur die Versicherten in der Selbstverwaltung vertreten.

Gemeinsam mit der Kassenärztlichen Bundesvereinigung (KBV), der Kassenzahnärztlichen Bundesvereinigung (KZBV) und der Deutschen Krankenhausgesellschaft (DKG) entscheidet der gkv-Spitzenverband im Gemeinsamen Bundesausschuss (G-BA) über die konkrete Ausgestaltung des Leistungskataloges der GKV.

Prüfauftrag für die Krankenkassen: Die gesetzlichen Krankenkassen sollten bis Ende 2017 prüfen, ob die klinischen Krebsregister voll arbeitsfähig wären. Nur dann dürften die Krankenkassen sie auf Dauer finanzieren. Neben den Ergebnissen überprüften die Krankenkassen vor allem Strukturen und Prozesse der Datenerhebung der Register entsprechend den 2013 gemeinsam mit Ländern und Registern erarbeiteten Förderkriterien. Um die Mitglieder bei diesem Prüfauftrag zu unterstützen, entwickelt der GKV-Spitzenverband zusammen mit den Krankenkassen einen bundesweit einheitlichen Bewertungsmaßstab.

Bewertung als Stakeholder NGD: Der gkv-Spitzenverband Bund als Dachverband der gesetzlichen Krankenversicherungen hat eine Expertise mit den notwendigen Regularien zum Umgang mit der Bereitstellung von Patientendaten (z.B. DIMDI, klinische Krebsregister). Kenntnisse zur Umsetzung zur Kriterien-Überprüfung bzgl. der Erfüllung von geforderten Regularien und Kriterien im Umgang mit Datenspeicherung, Datenzusammenführung und der Daten-Auswertung sind vorhanden. Dies kann auch bundeslandübergreifend erfolgen. Die juristischen Voraussetzungen zur externen Datenzusammenführung müssten ggf. analog zu den klinischen Krebsregistern über das SGB V den Bedingungen angepasst werden.

#### *Deutsche Krankenhausgesellschaft (DKG)/Kliniken*

Als Bundesverband steht die Deutsche Krankenhausgesellschaft für 28 Mitgliedsverbände von Krankenhausträgern: 16 Landesverbände, 12 Spitzenverbände. Mit dieser Trägervielfalt repräsentiert die Deutsche Krankenhausgesellschaft die gesamte Breite der Krankenhausinteressen. Neben ihren politischen Aufgaben hat sich die Deutsche Krankenhausgesellschaft den öffentlichen Dialog, den Erfahrungsaustausch und die Unterstützung der wissenschaftlichen Forschung zum Ziel gesetzt. Zusammen mit den anderen Partnern der Selbstverwaltung – Krankenversicherungen, Ärzteverbände und Pflegeverbände – entscheidet sie über alle wichtigen krankenhauserrelevanten Themen.

Sie unterrichtet die Öffentlichkeit und unterstützt staatliche Körperschaften und Behörden bei der Vorbereitung und der Durchführung von Gesetzen. Die Deutsche Krankenhausgesellschaft vertritt die Interessen des deutschen Krankenhauswesens auf der europäischen und internationalen Verbandsebene.

Eines der wichtigsten Ziele für die Krankenhäuser ist die Steigerung der Wirtschaftlichkeit, damit auch in Zukunft eine patientennahe und flächendeckende Versorgung sowie eine hohe Behandlungsqualität garantiert sind.

Aktuell 1.951 Krankenhäuser in Deutschland versorgen Jahr für Jahr 19,5 Millionen stationäre Patienten und rund 18 Millionen ambulante Behandlungsfälle. Die Krankenhäuser leisten damit einen entscheidenden Beitrag zur Behandlung von Patienten in unserem Gesundheitswesen.

Die Deutsche Krankenhausgesellschaft verfolgt ausschließlich und unmittelbar gemeinnützige Zwecke.

Bewertung als Stakeholder NGD: Die DKG hat eine Expertise mit den notwendigen Regularien in der Zusammenarbeit der deutschen Kliniken. U.a. gehört die Unterstützung der Forschung zu ihren Aufgaben und die Ausweitung auf eine Nutzung des NGD wäre den Aufgaben entsprechend. Kenntnisse zum Umgang mit Datenspeicherung, Datenzusammenführung und der Daten-Auswertung sind in den Kliniken, insbesondere in den forschenden Institutionen vorhanden. Die juristischen Voraussetzungen zur externen Datenzusammenführung müssten ggf. analog zu den Tumorzentren und klinischen Krebsregistern über das SGB V den Bedingungen angepasst werden.

#### *Bundesärztekammer/Medizinischer Bereich*

Die Bundesärztekammer (Arbeitsgemeinschaft der Deutschen Ärztekammern) ist die Spitzenorganisation der ärztlichen Selbstverwaltung; sie vertritt die berufspolitischen Interessen der 506.014 Ärztinnen und Ärzte (Stand: 31.12.2017) in der Bundesrepublik Deutschland. Als Arbeitsgemeinschaft der 17 deutschen Ärztekammern wirkt die Bundesärztekammer (BÄK) aktiv am gesundheitspolitischen Meinungsbildungsprozess der Gesellschaft mit und entwickelt Perspektiven für eine bürgernahe und verantwortungsbewusste Gesundheits- und Sozialpolitik.

- Sicherung einer guten medizinischen Versorgung der Bevölkerung
- Herbeiführung einer möglichst einheitlichen Regelung der ärztlichen Berufspflichten und Grundsätze für die ärztliche Tätigkeit auf allen Gebieten
- Vermittlung der Position der Ärzteschaft zu gesundheitspolitischen und medizinischen Fragen
- Förderung der ärztlichen Fortbildung
- Förderung der Qualitätssicherung
- Herstellung von Beziehungen zur medizinischen Wissenschaft und zu ärztlichen Vereinigungen des Auslandes
- Regelung insbesondere von Berufsordnung: Regelt ethische und berufsrechtliche Pflichten der Ärzte untereinander und gegenüber den Patienten. Darunter fällt z.B. die Schweigepflicht des Arztes

Bewertung als Stakeholder NGD: Die BÄK ist das Bindeglied zwischen der Ärzteschaft und den Belangen der medizinischen Versorgung. Damit hat sie eine Expertise mit den notwendigen Regularien in der Zusammenarbeit der deutschen Ärzteschaft. U.a. gehören die Unterstützung der Wissensvermittlung, Wissensentwicklung und der Zugang zur Forschung mit ihren Ergebnissen zu ihren Aufgaben. Die Ausweitung auf eine Nutzung des NGD für die Ärzteschaft wäre aufgabenentsprechend, aber mit einigem Aufwand verbunden. Kenntnisse zum Umgang mit Datenspeicherung, Datenzusammenführung und der Daten-Auswertung sind nicht durchgängig vorhanden. Die inhaltlichen, technischen und juristischen Voraussetzungen zur externen Datenzusammenführung müssten angepasst werden.

#### *vfa (Verband der forschenden Pharma-Unternehmen)/Pharmaindustrie*

Deutschland ist einer der führenden Standorte für Pharmaforschung: 23 von 45 Mitgliedsfirmen des vfa - mit Hauptsitz im In- und Ausland – betreiben Labors für Wirkstoff- oder galenische Forschung. 35 Mitgliedsfirmen koordinieren von Deutschland aus klinische Studien. Sie wenden pro Jahr mehr als 5 Milliarden Euro für Forschung und Entwicklung allein in Deutschland auf. Mit 594 Studien liegt Deutschland vor UK, nur überboten von den USA. In den meisten Studien werden Medikamente gegen Krebs- oder Entzündungskrankheiten erprobt. Der vfa setzt sich auch politisch für eine Stärkung von Wissenschaft und Forschung ein.

Vom vfa werden vorgehalten:

- Datenbanken zu Arzneimitteln (personalisierte Medizin, zugelassene gentechnische Arzneimittel in Deutschland, Arzneimittelzulassungen für Kinder)
- Register nicht-interventioneller Studien

Die vfa-Mitgliedsunternehmen haben sich verpflichtet, alle nach dem 1. Mai 2007 begonnenen Anwendungsbeobachtungen und anderen nicht-interventionellen Studien öffentlich zu registrieren. Der

vfa bietet hierzu ein Online-Register an; doch können die Unternehmen dafür auch andere öffentliche Studienregister (wie [clinicaltrials.gov](http://clinicaltrials.gov)) nutzen. Grundlage der Selbstverpflichtung der Unternehmen zur Registrierung nicht-interventioneller Studien sind die 2007 publizierten vfa-Empfehlungen zur Verbesserung der Qualität und Transparenz von nicht-interventionellen Studien.

Bewertung als Stakeholder NGD: Der vfa führt die Interessen der deutschen forschenden Pharmaunternehmen zusammen. Er setzt sich explizit für die Stärkung von Wissenschaft und Forschung ein. Datenbanken und Studienregister werden zur Nutzung vorgehalten. Die Ausweitung auf eine Nutzung des NGD wäre aufgabenentsprechend. Kenntnisse zum Umgang mit Datenspeicherung, Datenzusammenführung und der Daten-Auswertung und Datenbereitstellung sind in den forschenden Pharmaunternehmen vorhanden. Die technischen und juristischen Voraussetzungen im Sinne des Patientenschutzes zur externen Datenzusammenführung müssten angepasst werden. Der wirtschaftliche Benefit ist kalkulierbar.

#### *BVMed (Bundesverband Medizintechnologie)/Medizin-Technische Firmen*

Der BVMed vertritt als Wirtschaftsverband rund 220 Industrie- und Handelsunternehmen der Medizintechnologie-Branche. Im BVMed sind unter anderem die 20 weltweit größten Medizinproduktehersteller im Verbrauchsgüterbereich organisiert.

Der BVMed fördert und vertritt als Wirtschaftsverband die gemeinsamen Interessen der Industrie- und Handelsunternehmen der Medizintechnologie. Der BVMed vertritt nach außen – gegenüber Politik und Öffentlichkeit - die Interessen der Mitgliedsunternehmen. Er gestaltet Gesetze, Richtlinien und Standards mit.

Bewertung als Stakeholder NGD: Der BVMed führt die Interessen der Industrie- und Handelsunternehmen der Medizintechnologie-Branche zusammen. Dazu gehören z.B. auch die Firmen, die digitale Technologien, bildgebende Verfahren oder Instrumente zur Überwachung biometrischer Daten entwickeln und im Gesundheitsmarkt ver- und betreiben. Die Auseinandersetzung mit Regularien über den BVMed ist gängig, der Umgang mit betriebsinternen Daten zur Datensammlung und –auswertung ist geläufig. Die Ausweitung auf eine Nutzung des NGD wäre zu formulieren und müsste über den BVMed kommuniziert werden. Die technischen und juristischen Voraussetzungen im Sinne des Patientenschutzes zur externen Datenzusammenführung müssten formuliert und implementiert werden. Der wirtschaftliche Benefit müsste von den einzelnen Unternehmen kalkuliert und abgewogen werden.

#### *Statistisches Bundesamt*

Das Statistische Bundesamt (StBA / Destatis) stellt Daten zu allen relevanten gesellschaftlichen und wirtschaftlichen Themen bereit. Diese Informationen sind objektiv, unparteilich, qualitativ hochwertig und für die Allgemeinheit verfügbar. Die im Geschäftsbereich des BMI verankerte Bundesoberbehörde hat die Aufgabe, ca. 400 Bundesstatistiken methodisch und technisch vorzubereiten, die Bundesergebnisse aus den Daten der Länder zusammenzustellen und zu veröffentlichen sowie das Programm der Bundesstatistik weiterzuentwickeln.

Zusammen mit den statistischen Ämtern der Länder und weiteren externen Datenproduzenten sorgt das Statistische Bundesamt dafür, dass die Statistiken nach einheitlichen wissenschaftlichen Methoden durchgeführt werden. Das Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BstatG) definiert dabei die Aufgaben des Statistischen Bundesamtes. Im BstatG sind wichtige Grundsätze verankert, unter anderem die Verpflichtung zu Neutralität, Objektivität und fachlicher Unabhängigkeit.

In einer zukunftsorientierten, wissensbasierten Gesellschaft erfolgen Planungen, Entscheidungen und deren Erfolgswertung aufgrund von Fakten. Um die Qualität der Daten zu gewährleisten, wurden verbindliche Qualitätsleitlinien eingeführt.

Bewertung als Stakeholder NGD: Das Statistische Bundesamt erfüllt die notwendigen Konditionen für eine NGD-Nutzung innerhalb der institutseigenen Aufgabenstellung. Kenntnisse zur Umsetzung der technischen Konditionen sind vorhanden, da der Umgang mit bundeslandübergreifender Datenzusammenführung, Datenspeicherung und Datenanalyse geläufig ist. Die technischen

Voraussetzungen zur externen Datenzusammenführung und die juristischen Rahmenbedingungen für eine Datenübertragung zur übergreifenden Nutzung im NGD müssten den Bedingungen angepasst werden.

## E. Differenzanalyse

Da es zwischen einzelnen Institutionen, wie Universitäten, Krankenversicherungen, Krebsregistern, RKI, DIMDI (im Verfahren nach §§303a ff SGB V werden hier z.B. in Forschungsprojekten auf der Basis über die Zeit stabiler und einheitlicher Pseudonyme die Daten der Versicherten aller Krankenkassen aus verschiedenen Berichtsjahren zusammengeführt) bereits Datenaustausch, möglicherweise auch Datenzusammenführung und Datenanalyse gibt, ist das dafür notwendige Wissen vielen der genannten, als NGD-Nutzer in Frage kommenden im Gesundheitssystem Beteiligten bereits geläufig. Eine technische Einrichtung (Hardware- und Softwarelösungen), um den Datenaustausch zu praktizieren, ist grundsätzlich verfügbar, das Know-how zum Datenumgang ist beim Personal vorhanden, auch das Know-how für eine statistische Datenauswertung liegt vor. Jedoch wird das Zusammenführen von Daten aus unterschiedlichen Quellen durch folgende Probleme behindert oder im Einzelfall unmöglich:

- Anonymisierte Daten zusammenzuführen, nachdem die Daten nach Art und Struktur vereinheitlicht wurden, bietet einen nur begrenzten Zusatznutzen in Form einer Vergrößerung der Grundgesamtheit und damit der Power im Sinne einer sog. Megaanalyse. Die Hürden dabei sind offensichtlich so groß, dass sich die Wissenschaft derzeit überwiegend alternativ mit der Methodik der Metaanalyse behilft.
- Pseudonymisierte Daten zusammenzuführen, bietet demgegenüber den Vorteil, zu denselben Personen Daten unterschiedlicher Art zusammenzuführen (als Beispiel zur Illustration: Zensuren aus Schule und Ausbildung – wenn es die denn als Datenbanken gäbe – einerseits und ICD-Kodes andererseits). Sofern Daten in den verschiedenen Datenbanken der Organisationen der Gesundheitsversorgung (Datenhalter) im Bundesgebiet pseudonymisiert sind, so unterscheiden sich aber die Pseudonyme, so dass eine Datenzusammenführung unmöglich ist, solange jeder Datenhalter die Daten nicht repersonalisiert hat. Dem können datenschutzrechtliche Bedenken entgegenstehen.

Bei noch vorhandener Uneinheitlichkeit der Art und Struktur von Daten in den verschiedenen Datenbanken der Organisationen der Gesundheitsversorgung (Datenhalter) im Bundesgebiet, bietet NGD die Möglichkeit, einheitliche Basisdatensätze und vereinheitlichte additive Datensätze zu schaffen und sie durch miteinander kompatiblen digitalen Speichersystemen austauschbar zu machen. In den verschiedenen Institutionen der Gesundheitsversorgung und -forschung im Bundesgebiet herrscht teilweise Unkenntnis zu den vorhandenen Datensätzen an den jeweils anderen Instituten. Die Aufgabenbereiche der großen Institutionen wie z.B. des RKI oder DIMI sind zwar meist bekannt, welche Daten konkret und in welchem Umfang bei Institutionen (wie z.B. Kliniken oder der pharmazeutischen Industrie) vorhanden sind, aufgrund mangelnder Transparenz oder auch fehlender Möglichkeit, dies abzufragen, sind aber oft nicht bekannt. Entsprechend ist die Aufgabe, die Informationen zu den vorhandenen Datensätzen nicht nur für alle verschiedenen Institutionen der Gesundheitsversorgung und -forschung zugänglich zu machen, sondern auch innerhalb der Institutionen selbst zu bündeln, zu beschreiben und Informationen zu den Datensätzen zugänglich zu machen.

Selbst wenn eine grundsätzliche Möglichkeit bestünde, Zugang zu Daten aus verschiedenen Institutionen zur weiteren Analyse zu erhalten (z.B. Universitätsklinik/ Studienzentren, Krebsregister) ist ein Zugang erschwert, da die logistischen Abläufe für einen Datenaustausch nicht standardisiert gegeben sind und je nach Anfrage in jedem Einzelprojekt erneut spezifisch geplant und vorgenommen werden müssen. Notwendig ist, eine Standardisierung der Datenstrukturen und logistischen Abläufe zu entwickeln, sowohl was die Datenauswahl- und als auch Datentransfer-Mechanismen betrifft und mit

technisch einheitlichen bzw. kompatiblen Lösungen, inklusive allgemein kompatibler Schnittstellenlösungen umzusetzen.

Bei für eine Analyse nicht ausreichend großen pseudonymisierten Datensätzen oder fehlender Möglichkeit, die Datensätze mit denen anderer Institutionen zu matchen, bleibt keine Alternative als die Rückführung auf personalisierte Daten, um diese dann mit anderen Datensätzen mit Personenbezug zusammenführen zu können und nach erneuter Pseudonymisierung auszuwerten.

Die Aufbewahrungspflichten für Daten sind in den verschiedenen Rechtsbereichen und für die verschiedenen Institutionen uneinheitlich oder nicht geregelt. Dies gilt nicht nur für die zur Aufbewahrung zu nutzenden Medien und Formate, sondern auch für die Aufbewahrungsfristen. So unterscheiden sich sogar für Arztpraxis und Krankenhaus z.B. die Aufbewahrungsfristen für Patientenunterlagen (10 Jahre seit Abschluss der Behandlung) einerseits und die Aufbewahrungsfristen gemäß Strahlenschutzverordnung und Röntgenverordnung andererseits: Hier müssen Aufzeichnungen und Röntgenbilder nach der letzten Untersuchung 10 Jahre, über die Behandlung mit ionisierten Strahlen 30 Jahre lang nach der letzten Untersuchung oder Behandlung aufbewahrt werden.

Die Aufbewahrungsfrist für Röntgenbilder und die Aufzeichnungen von Röntgenuntersuchungen beginnt für minderjährige Patienten erst ab dem 18. Lebensjahr, so dass alle Röntgenbilder von Kindern und Jugendlichen mindestens bis zur Vollendung des 28. Lebensjahres aufbewahrt werden müssen. An letzter Regelung ist erkennbar, dass sich der Beginn der Frist, solange er als „letzte Behandlung“ definiert ist, nicht scharf bestimmen lässt. Abweichend legt § 304 SGB V für die Krankenkassen fest, die Daten nach § 295 SGB V (Abrechnung ärztlicher Leistungen) seien „spätestens“ nach 10 Jahren zu löschen; sie können also früher gelöscht werden: „Krankenkassen können die rechtmäßig gespeicherten ärztlichen Abrechnungsdaten für Zwecke der Weiterentwicklung und Durchführung des Risikostrukturausgleichs länger aufbewahren; sie sind nach spätestens vier Jahren zu sperren und spätestens nach den in der Rechtsverordnung genannten Fristen zu löschen. Die Krankenkassen können für Zwecke der Krankenversicherung Leistungsdaten länger aufbewahren, wenn sichergestellt ist, dass ein Bezug zum Arzt und Versicherten nicht mehr herstellbar ist“. Die Daten werden bislang nicht für einen ausreichend langen Zeitraum aufbewahrt. Manche Fragestellungen (wie auch in dem hier aufgeführten Beispiel) sind damit nicht beantwortbar. Daher müssen Datensätze langfristig (jedenfalls länger als 10 Jahre) aufbewahrt werden. Die Aufbewahrungsfristen sind vermutlich historisch gewachsen als Kompromiss zwischen der Notwendigkeit ggf. später notwendiger Beweisführung einerseits und den aus der Aufbewahrungspflicht andererseits resultierenden Aufwänden und damit Kosten. Allerdings spielten vermutlich auch Gesichtspunkte des Datenschutzes eine Rolle. Jedoch stellt sich die Frage, ob in der Praxis die Dauer der Aufbewahrung in irgendeinem logischen kausalen Zusammenhang mit dem erfolgreichen Schutz des Individuums vor Datenmissbrauch steht. Nachdem heute Daten grundsätzlich elektronisch aufbewahrt werden (dies nicht zwangsläufig elektronisch auswertbar), dürften die Kosten der Aufbewahrung inzwischen von geringerer Bedeutung sein. Für die Wissenschaft sind diese Fristen jedenfalls hinderlich. Selbst eine Pflicht oder Möglichkeit zur Datenlöschung nach dem Tode des Individuums kann für die Wissenschaft hinderlich sein, indem sie intergenerationelle Forschungen einschränkt.

### III. Machbarkeitsbetrachtung und Umsetzungsvorschlag

#### A. Datenqualität und -verfügbarkeit, Qualitätsentwicklung

Mit Inbetriebnahme des NGD sollte ein Qualitätskonzept zu Grunde gelegt sein, das die Entwicklung der Datenqualität und -Verfügbarkeit auf Seiten der Datenbereitsteller und die strukturelle Qualitätsentwicklung des gesamten Netzwerkes (NGD mit NGD Nutzern) beschreibt. Während des Aufbaus und der Weiterentwicklung des NGD wird dieses Qualitätskonzept weiterentwickelt und optimiert. Zu berücksichtigende Bereiche des Qualitätskonzeptes sind die Qualität der Daten selbst, der Austausch der Daten, die Verfügbarkeit der Daten für das NGD und die Datenanalyse. Wichtige Elemente, die bei der Formulierung des Qualitätskonzeptes zu berücksichtigen sind, sind die Entwicklung von Qualitätsindikatoren, die Installation eines Überprüfungsprozesses, der Entwicklung der knowledge-base der NGD Nutzer und die Strukturentwicklung des NGD und NGD-Nutzer.

Die Datenqualität ist abhängig von der Datenerhebung, der Validität der erhobenen Daten, und der Datenvollständigkeit. Ist bereits bei der Datenerhebung qualifiziertes Personal notwendig. Bereits bei der Planung der Datenerhebung ist auf Kompatibilität und Vergleichbarkeit zu achten.

Die Kriterien zur Förderung klinischer Krebsregister des GKV-Spitzenverbandes vom 20.12.2013 gemäß § 65c SGB V (KFRG) verlangen so u.a., dass klinische Krebsregister regelmäßige Vollständigkeitsprüfungen durchführen. Klinische Krebsregister können zeigen, mit welcher Methode die Vollständigkeit der Daten zur Diagnose, Verlauf, Therapie und Nachsorge überprüft werden. Klinische Krebsregister haben vor jeder Meldungserfassung eine inhaltlich-klinische Grundprüfung der Meldung einschließlich formaler Datenplausibilisierungen vorzunehmen. Im Rahmen der klinischen Prüfung von Arzneimitteln haben gemäß Arzneimittelprüfrichtlinien-Verordnung in Verbindung mit Richtlinie 2001/83/EG des Europäischen Parlaments und des Rates vom 06. November 2011 zur Schaffung eines Gemeinschaftskodexes für Humanarzneimittel u.a. „source data verifications“ zu erfolgen. Das NGD müsste ähnliche Voraussetzungen zumindest in Stichproben etablieren, also besondere Qualitätsanforderungen sowohl an DN als auch DB stellen.

Für den Datenaustausch ist die technische Ausstattung bis hin zu einer Schnittstellenkompatibilität (siehe Technische Voraussetzungen) grundlegend. Es sollte ein zeitliches Konzept hinterlegt sein, das festlegt, in welchem Zeitraum nach Datenerhebung Daten übertragen werden müssen, wann Aktualisierungen erfolgen etc. Auch hier ist eine kompetente personelle Besetzung relevant für die Qualität der ausgeführten Prozesse. Strukturen, die Grundlage für den Austausch sind, sollten reibungsfrei zugänglich und nutzbar sein.

Für die Qualität der Analyseergebnisse des NGD ist die Datenverfügbarkeit über die NGD-Nutzer entscheidend. Die Analyseprozesse durch das NGD selbst müssen einer Qualitätskontrolle unterliegen und konstant optimiert werden. Eine systematische Qualitätsentwicklung über das Qualitätskonzept beinhaltet als relevante Faktoren die Entwicklung von Qualitätskriterien, um u.a. die folgenden Bereiche zu adressieren und prüfbar zu machen:

- Zugang zu Basisdatensatz und additiven Datensätze
- Vollständigkeit Basisdatensatz und additiver Datensätze
- Zeitliche Komponente
- Datenvorhaltung
- Verhältnis der Menge der Datenzulieferer pro Datenaufnehmende Stelle
- Rückkopplungsmechanismen zu Datenqualität (an NGD-Nutzer und von NGD-Nutzer an Datenerheber)

Nach Implementierung des Qualitätsprozesses sollte in zeitlich festgelegten Abständen eine Qualitätskriterien-Überprüfung (i.S. einer Selbstdeklaration oder Auditierung) erfolgen. Eine Wissensentwicklung sollte über einen strukturierten Erfahrungsaustausch und die geplante Bildung von Austauschformaten von Beginn an mit konzipiert werden.

Auch eine Strukturentwicklung (Datenerheber - NGD-Nutzer – NGD) sollte konstant realisiert werden. Als Beispiele können benannt werden:

- Ausbau der klinischer Krebsregister (Planung über Nationalen Krebsplan; Gesetzgebung KFRG)
- Zertifizierung von Tumorzentren durch die Deutsche Krebsgesellschaft

Über die Implementierung und Durchführung der Qualitätsentwicklungs-Maßnahmen kann die Datengranularität im Verlauf der Zeit konstant erhöht werden.

Beispiel des Ausbaus der klinischen Krebsregister: Im Rahmen der Diskussionen des Nationalen Krebsplans ([www.bmg.bund.de/praevention/nationaler-krebsplan.html](http://www.bmg.bund.de/praevention/nationaler-krebsplan.html)) wurde in verschiedensten Zieldiskussionen überdeutlich, dass für eine Reihe von Fragestellungen verlässliche, deutschlandweite und vergleichbare Daten zum Versorgungsgeschehen fehlen. Der Spitzenverband Bund der Krankenkassen ist beauftragt, die „notwendigen Festlegungen“ in sogenannten Fördervoraussetzungen zu treffen. Ziel der Festlegungen ist es, eine register- und länderübergreifende Vergleichbarkeit der Datenerhebung und -nutzung zu sichern.

Der einheitliche onkologische Basisdatensatz von Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID) wurde im März 2008 verabschiedet und im Februar 2014 aktualisiert. Er gilt für alle Krebsarten und wird fortlaufend um tumorspezifische Module ergänzt. Mit dem einheitlichen onkologischen Basisdatensatz ist ein Instrument geschaffen worden, das einen einheitlichen onkologischen Standard vorgibt, Mehrfachdokumentationen verhindert und in allen Bundesländern und klinischen Strukturen eine vergleichbare Erfassung und Auswertung von Krebsbehandlungen ermöglicht. Die rechtliche Grundlage dafür geschaffen wurde durch das 2013 in Kraft getretene Krebsfrüherkennungs- und -registriergesetz (KFRG) zur Qualitätsentwicklung in der Krebsversorgung. Nach § 65c Absatz 1 Satz 3 des Fünften Buches Sozialgesetzbuch (SGB V) hat die klinische Krebsregistrierung auf der Grundlage des bundesweit einheitlichen Datensatzes der ADT und GEKID zur Basisdokumentation für Tumorkranke zu erfolgen.

## B. Überlegungen zur Herstellung technischer und semantischer Interoperabilität

Durch die Annahme einer Nutzung von HL7 FHIR als universelles Format zur Übertragung von Daten aus Quellsystemen ist die syntaktische Interoperabilität (siehe II.C.3.a) beim Datenaustausch gewährleistet. Eine auf Hadoop basierende Big Data-Plattform mit einem Data Lake im HDFS Format ist grundsätzlich für die technische Verarbeitung heterogener Daten ausgelegt. Dabei wird Apache Sqoop das Laden der Daten aus externen strukturierten Datenquellen in HDFS oder Hive und HBase genutzt (siehe III.B.1.e)

Um eine Abbildbarkeit medizinischer Informationen in hochgradig heterogenen Kodierungen aus unterschiedlichen Quellsystem und damit eine semantische Interoperabilität (siehe II.C.3.b) herstellen zu können, sind verschiedene Ansätze denkbar:

1. das NGD legt für verschiedene medizinische Themenbereiche fest, welche Standards für die Kodierung von Konzepten aus diesen Themenbereichen zu verwenden sind (z.B. Diagnosen über ICD-10, Sedierungstiefen über RASS, Maßeinheiten über UCUM),
2. es werden bilaterale Mapping-Tabellen aufgestellt, die beschreiben, wie ein Codesystem auf ein anderes abzubilden ist, oder
3. das NGD definiert ein Referenz-Codesystem und legt fest, wie einzelne Codesysteme auf das Referenzsystem abzubilden sind.

Alle drei Ansätze sind – zumindest isoliert betrachtet – nicht wirklich praktikabel:

1. Die Anpassungskosten an den laufenden IT-Systemen zur Umstellung auf die vorgegebenen Standards wären enorm und die Migration würde Jahre dauern. Auch wären positive oder negative Incentives erforderlich, um eine solche Standardisierung auch wirklich durchzusetzen. Hinzu kommt, dass viele – insbesondere vom Umfang her kleine – Codesysteme gerade deshalb entstanden sind, weil internationale Terminologien die hier kodierten Konzepte nicht enthalten

oder weil mit den Konzepten spezielle Nutzungsbeschränkungen oder –vorgaben verknüpft sind.

2. Selbst wenn es zur Kodierung eines Sachverhalts nur 5 verschiedene Codesysteme gibt, wären  $4 + 3 + 2 + 1 = 10$  Mapping-Tabellen erforderlich. Die Zahl der Mapping-Tabellen bei potenziell 10.000 nicht trennscharfen und sich überlappenden Codesystemen wird so groß sein, dass es unrealistisch ist, dass diese alle von dem NGD erstellt und gepflegt werden können.
3. SNOMED CT enthält ca. 300.000 Konzepte. Ein Referenzsystem, auf das sich alle in der Praxis eingesetzten Codesysteme abbilden lassen, wird kaum kleiner sein. Im Gegensatz zum 2. Ansatz wird man hier zwar deutlich weniger Mapping-Tabellen benötigen (genau 1 pro Codesystem), dafür ist deren Erstellung aber deutlich schwieriger, da hierzu tiefes Wissen sowohl über das Referenzsystem als auch über die Fachdomäne des Codesystems erforderlich ist.

Gleichzeitig hat aber auch jede dieser Optionen vielversprechende Ansätze in sich:

1. Länder wie Dänemark und Österreich haben gezeigt, dass eine Konsolidierung der im Einsatz befindlichen Codesysteme durch regulative Maßnahmen und „Pflicht-Anwendungen“ möglich ist. Gelingt es, durch solche Maßnahmen einige der „großen“ Bereiche (Diagnosen, Symptome, Medikation, Vitaldaten, Labordaten, etc.) über jeweils einen Standard sinnvoll und möglichst vollständig kodieren zu können, wird sich die Zahl der noch über andere Maßnahmen zu adressierenden Konzepte signifikant verringert haben. Das Interoperabilitätsverzeichnis „vesta“ der gematik soll derartige Ziele unterstützen, erscheint dazu aber in der jetzigen Form als unverbindliche Sammelstelle wenig geeignet.
2. Mappings zwischen zwei (oder mehr) Terminologien können das Mapping gegen ein Referenz-Codesystem vereinfachen. In einem solchen Fall wird nur eine Terminologie semantisch an das Referenzsystem angebunden, während alle anderen Terminologien nur auf diese angebundene Terminologie abgebildet werden.
3. Eine semantisch mächtige und über Konzeptrelationen Elemente eines Domänen-/Wissens-Modells abbildende Referenzterminologie „vererbt“ die damit verbundenen Möglichkeiten einer semantischen Suche auf die angebundene Terminologien. Auch werden sich hier viele „quick wins“ finden lassen, bei denen sich gerade eine Vielzahl kleinerer, fokussierter Vokabulare mit wenig Aufwand auf das Referenzsystem abbilden oder als Ergänzung an dieses anbinden lässt.

Der vielversprechendste Ansatz ist somit eine geschickte Kombination aller drei Ansätze. Dabei besteht kein Zwang, dass das Referenz-Codesystem über eine einzelne bestehende Terminologie repräsentiert wird. Wie in Abschnitt II.C.3.c-f) ausgeführt bietet sich aber die Nutzung von standardisierten Codesystemen wie SNOMED CT und LOINC als Kern eines Referenz-Codesystems an. Weitere domänenspezifische Systeme können dem Referenzsystem hinzugefügt werden (z.B. ICSN für Gen-Daten oder ICNP für Daten aus der Pflege) und über ein ISO 11179 konformes Data Dictionary um automatisch ausführbare Transformationsregeln zur Umrechnung von Referenzeinheiten und Umrechnungsformeln erweitert werden.

## C. Infrastruktur und Technologie

Gemäß dem Stand der Technik, sowie den fachlichen und rechtlichen Anforderungen (siehe II.C), wird eine technische Umsetzung vorgeschlagen, bestehend aus zwei Hauptkomponenten: Zum einen eine *Hadoop* basierte *Big Data-Plattform*, die für den Zweck der Datenverwaltung, Datenanalyse und Datensicherheit zuständig ist. Zum anderen eine speziell entwickelte Webplattform, die als *Datenkatalog* dient und es den Nutzern erlaubt, Daten einzupflegen, Studien zu registrieren und wo alle Aktivitäten der Neutralen Stelle technisch begleitet und koordiniert werden.

## 1. Konzept und Architektur des technischen Zielmodells

Als Konsequenz der fachlichen, rechtlichen und technischen Anforderungen muss das technische Modell auf zwei der Komponenten des Fachmodells eingehen. (1) Eine interne Big Data-Plattform, die in unserem Modell bei der Neutrale Stelle angesiedelt ist, und deren Zweck es ist, die Daten zu verwalten, zu integrieren, analysieren und zu sichern. (2) Eine externe Webplattform welches die Kontaktplattform bildet, ähnlich wie ein Datenkatalog, die es ermöglicht, den Nutzern Datensätze zur Verfügung zu stellen oder ihre Existenz auch nur einzutragen, die aber auch als Vermittlungsplattform für die Erstellung der Studien dient. Zusätzlich muss die Standardkonformität und Interoperabilität (siehe III.A) über einen Terminologie-Server gewährleistet sein. (3) Als ergänzende Ausbaustufe kann eine technische Unterstützung des „Contracting“ durch eine Wissensmanagementsoftware (inkl. Vertragsmanagement) mit (KI-basierter, lernender) Analyse, Suche, Pflege von Vertragsvorlagen / Verträgen und weiteren Wissen diese Aushandlungsprozesse unterstützen und teilweise auch automatisieren.

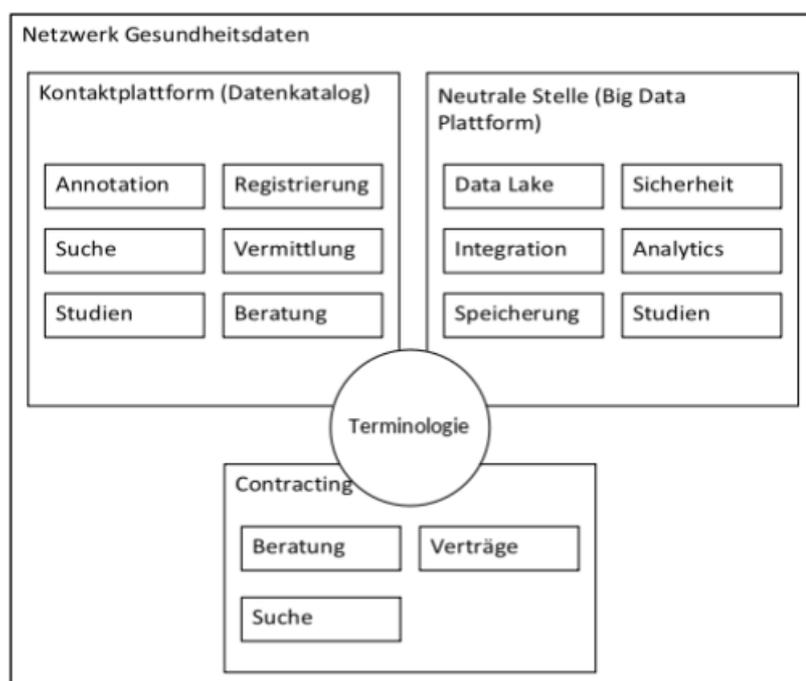


Abbildung 4: Funktionale Architektur NGD

### a) Kontaktplattform

Die folgenden technischen Funktionalitäten sollten gegeben sein um eine effiziente Kontaktplattform anbieten zu können:

**Annotation:** Die registrierten Datensätze müssen gut dokumentiert und mit semantischen Annotationen für die maschinelle Auswertung versehen werden. Die Rolle der Neutralen Stelle ist es, ein Konzept für die Annotation solcher Datensätze zu entwickeln und die teilnehmenden Partner zu beraten wie sie am besten ihre Daten annotieren und formatieren und ein Terminologieserver aufzubauen.

**Suche:** Eine intelligente Suche, die über die Metadaten funktioniert, soll es ermöglichen, schnell die passenden Datensätze für die Erstellung einer Studie zu finden.

**Registrierung:** Die Registrierung soll Organisationen die Daten liefern möchten oder solchen die Studien erstellen möchten sich und ihre Datensätze oder Studien registrieren in einer strukturierten Art und mit begleitender Beratung der Neutralen Stelle.

**Vermittlung:** Die Neutrale Stelle hat die Rolle nach der Anmeldung einer Studie, in ihrer Datenbank nach passenden Datensätzen zu suchen und die Vermittlung zwischen Auftraggeber und Datenlieferant

herzustellen. Im Nachhinein muss sie die Studie durchführen. Eine Teilautomatisierung der Such- und Vermittlungsfunktion ist angestrebt.

**Studien:** Wenn die Studie durchgeführt worden ist, muss die Neutrale Stelle sicherstellen, dass der gesamte Analyseprozess dokumentiert und nachvollziehbar ist und dass die Daten von den Datenlieferanten bei den Datenlieferanten selbst aufbewahrt werden für eine spätere Nachprüfung der Studie.

#### *b) Big Data-Plattform*

Die wichtigsten Komponenten der Big Data-Plattform sind:

**Data Lake:** Hier werden alle Daten im Rohformat abgelegt, zusammen mit den Metadaten für die spätere Analyse.

**Integration:** Die Integration geschieht, wenn möglich durch die mitgelieferten semantischen Metadaten, und den Ontologien und Mappings aus dem Terminologieserver. Wenn diese nicht zu Verfügung stehen, muss ein Experte die Integration vornehmen.

**Speicherung:** Die Daten werden auf einem verschlüsselten HDFS Datei System unter Nutzung von – der maximalen Sicherheitsstufe genügenden – Hardware gespeichert.

**Sicherheit:** Die Sicherheit im Hadoop System wird mit Hilfe von Apache KNOX und Encrypted HDFS verstärkt, zusätzlich zum Sicherheitskonzept.

**Analytics:** Die Analyticsfunktionalitäten werden mit Hilfe von Apache Spark und der ML Library implementiert.

**Studien:** Durch die Vielzahl an Visualisierungsbibliotheken wie D3JS, AngularJS und vielen weiteren, die mit den meisten Hadoop Distributionen in Dashboards für Visual Analytics eingebaut werden, können Visuell Ansprechende Statistiken für die Studien erstellt werden.

#### *c) Contracting*

**Suche:** Eine intelligente Suche soll es den Mitarbeitern des Contracting Bereiches erlauben, effizient passende Vertragsvorlagen zu finden, die der Studie und den benötigten Datensätzen entsprechen.

**Verträge:** Bereits verhandelte Verträge müssen strukturiert gespeichert werden und der Terminologie aus dem Terminologieserver entsprechen.

**Beratung:** Die Beratung ist eine der zentralen Rollen des Contracting Bereiches, hier sollten Juristen in Zusammenarbeit mit medizinischen Domänenexperten, basierend auf existierenden Vertragsvorlagen und einer konkreten Terminologie, die Nutzer beraten, um schnell zu einer Vertragsverhandlung zu führen. Eine Teilautomatisierung ist angestrebt.

#### *d) Terminologieserver*

Eine einheitliche Terminologie auf Basis (semantischer) Standards ist wichtig für die Erstellung von Interoperabilität und eine effiziente Wiederfindung von Informationen. Die Konzepte des Terminologieservers wurden in Sektion 2 dieses Dokuments vorgestellt. Der Terminologieserver dient dazu die Studien und Datensatzbeschreibungen strukturiert und einheitlich zu gestalten, Verträge zu strukturieren und semi-automatisch zu generieren, und am wichtigsten, die verschiedenen Datensätze durch Mappingregeln zu integrieren, damit eine effiziente, nachvollziehbare Datenanalyse stattfinden kann.

Die Verwaltung des Referenzsystems und aller Standards und Mappings erfolgt über den Terminologieserver. Das Referenzsystem ist in dem Sinne offen, als dass es für alle angebundenen Einrichtungen einsehbar ist. Angebundene Einrichtungen werden so dabei unterstützt, neue Datenerfassungen direkt unter Nutzung der Referenzsysteme anzulegen oder aber eigene Mappings auf das Referenzsystem bereitzustellen (Achtung: SNOMED CT Lizenzbedingungen!).

An den Terminologieserver angebunden (bzw. in diesen integriert) wird ein Data Dictionary, das maschinenauswertbare Transformationsregeln zwischen im deutschen Gesundheitswesen genutzten Maßeinheiten oder auch zwischen anderen aufeinander regelbasiert abbildbaren Konzepten (z.B. Bewegungsdauer -> Kalorienverbrauch) vorhält.

e) *Referenzarchitektur*

i. *Neutrale Stelle*

Daraus abgeleitet wird als technische Referenzarchitektur für die Neutrale Stelle folgende Hadoop basierte Lösung vorgeschlagen:

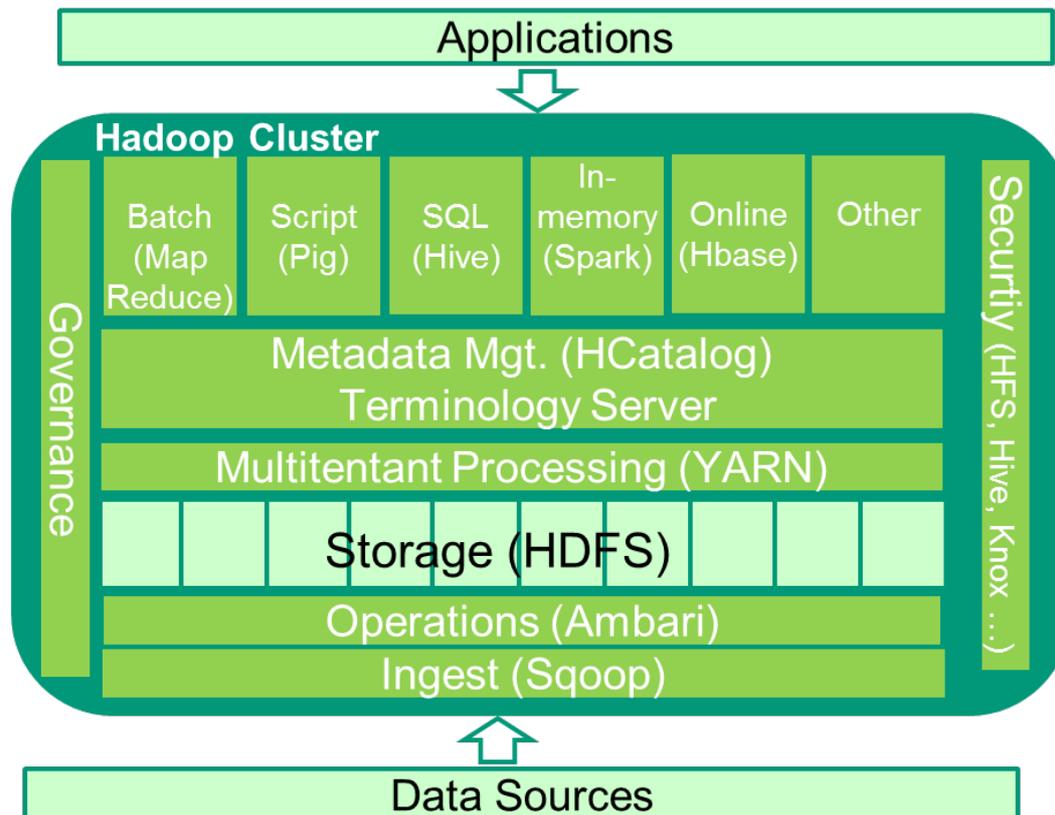


Abbildung 5: Hadoop-basierte technische Referenzarchitektur

#### LADEN

Apache Sqoop wird für das Laden der Daten aus externen strukturierten Datenquellen in HDFS oder Hive und HBase genutzt. Zusätzlich übernimmt der Terminologieserver und Metadaten Katalog zur syntaktischen und semantischen Integration der Daten auf Basis von Standards (wie HL7) in Hadoop.

#### PROZESSIEREN

Abhängig von den Auswertungsaufträgen werden die Daten als Batch (mit Apache Hadoop MapReduce and Apache Pig); interaktiv (mit Apache Hive); online (mit Apache HBase) oder direkt als Datennachrichtenströme (mit Apache Storm) verarbeitet.

#### ANALYSIEREN

Nach der Prozessierung in Hadoop können die Daten im Hadoop Cluster mit den integrierten Analysewerkzeugen (z.B. Spark oder Flink) weiter analysiert werden oder in weitere Analyse-, KI- und Visualisierungsanwendungen exportiert werden, die überwiegend bereits direkt mit den Daten in Hadoop arbeiten können.

Als Querschnittsbelange sind die Steuerung inkl. des Lebenszyklusmanagements (mit Falcon, Atlas) und die Sicherheit mit Administration, Authentifizierung, Autorisierung, und Datenschutz und Auditierung (mit HDFS Encryption, Knox, Atlas, Ranger) umzusetzen.

## ii. Kontaktplattform

Für die Kontaktplattform wird eine Referenzarchitektur vorgeschlagen, welche zwischen dem Datenkatalog und den eigentlichen Datenbeständen trennt, so dass die physische und organisatorische Trennung zwischen Kontaktplattform des NGD und den Datenanbietern/-lieferanten umgesetzt werden kann.

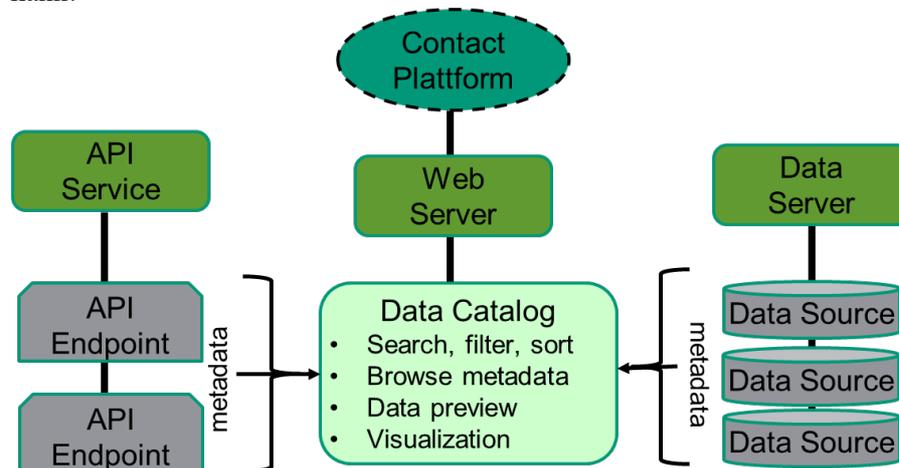


Abbildung 6: Architektur der Kontaktplattform

Für die technische Umsetzung soll eine webbasierte Datenkatalog-Software, wie CKAN (Comprehensive Knowledge Archive Network)<sup>13</sup> oder DKAN<sup>14</sup>, verwendet werden. Für die Umsetzung der wesentlichen Funktionen bietet sich die Nutzung der auf Drupal basierenden DKAN Open Data Plattform an, da die Katalogfunktionen dann bereits in das Drupal Web Content Management System (CMS) für den Webauftritt der Kontaktplattform integriert sind. Der Datenkatalog ist durch die Verwendung von W3C Data Catalog Vocabulary (DCAT) Standard-konform und erleichtert die semantische Dateninteroperabilität zwischen Katalogen und die Erweiterung um semantische Metadatenvokabulare und Domänenterminologien. Typische Metadaten sind z.B. „*catalog record, dataset, description, homepage, language, license, publisher, release date, rights, spatial, themes, title and update date*“, welche mit weiteren spezifischen Metadaten über den Datenanbieter und den angebotenen Datensätzen erweitert werden. Mit HealthData.gov existiert bereits eine Referenzlösung auf Basis von DKAN. Folgendes Bildschirmabbild zeigt ein Beispiel der Metadaten des „Safety Net Clinics“ Datensatzes<sup>15</sup>.

<sup>13</sup> <https://ckan.org/>, Zugriff am 27.7.2018

<sup>14</sup> <https://getdkan.org/>, Zugriff am 27.7.2018

<sup>15</sup> <https://www.healthdata.gov/dataset/safety-net-clinics>

**Dataset Info**

 These fields are compatible with **DCAT**, an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

Field	Value
Publisher	State of Oregon
Modified Date	2016-03-03
Release Date	2016-02-10
Identifier	https://data.oregon.gov/api/views/4q3t-6bdy
Public Access Level	Public

**Additional Info:**

Field	Value
harvest_object_id:	4q3t-6bdy
harvest_source_id:	data.oregon.gov
harvest_last_updated:	
source:	data.oregon.gov

Abbildung 7: Beispiel für Metadaten-Datensatz

Drupal selbst bietet diverse weitere Plug-ins aus der Drupal Community, die für erweiterte Funktionen der Kontaktplattform genutzt werden können. Vorgeschlagen wird die Erweiterung um Apache Stanbol<sup>16</sup>, welches bereits einen Adapter zu Drupal anbietet:

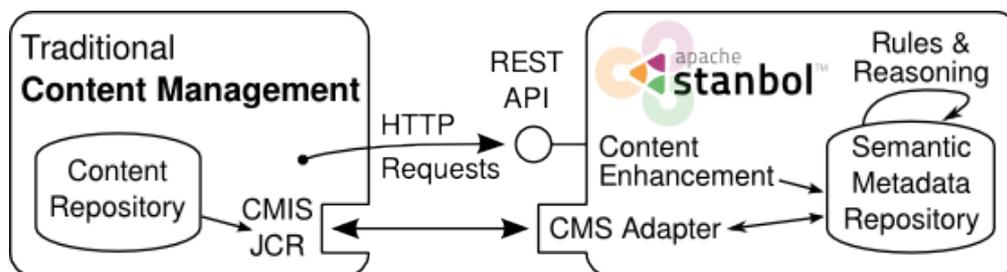


Abbildung 8: apache stanbol als CMS-Erweiterung

Damit lassen sich über den reinen Datenkatalog hinaus zusätzliche semantische KI Funktionen in der Kontaktplattform umsetzen, wie z.B. die Text-/Inhaltsanalyse in Beschreibungen über die Anbieter und deren Datenangeboten und Lizenzen oder semantische Schlussfolgerungen für eine IT-gestützte Vermittlung (z.B. Empfehlungssysteme mit semantischen „Matchmaking“) und die regel-basierte Bearbeitung von Abläufen (z.B. Workflowautomatisierung).

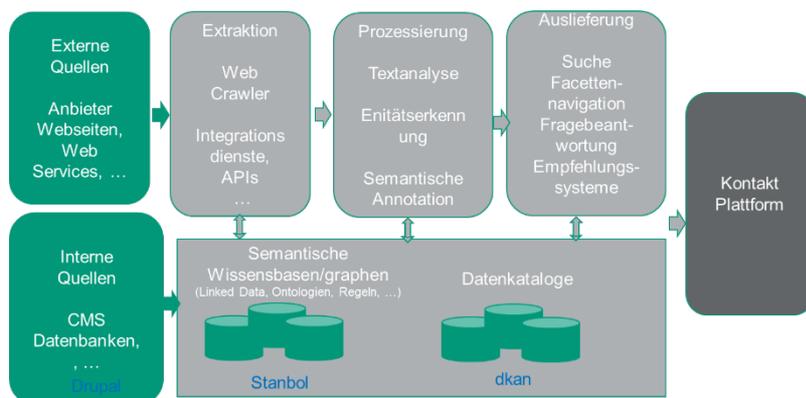


Abbildung 9: Blockbild Funktionalitäten Kontaktplattform

iii. Contracting

Für die technische Umsetzung der wesentlichen Unterstützungsfunktionen des Contracting, der Suche und Bereitstellung von passenden Vertragsvorlagen, wird für das Wissens- und Vertragsmanagment

<sup>16</sup> <https://stanbol.apache.org/>, Zugriff am 27.7.2018

eine feingranulare (semantische) Repräsentation auf Basis existierender Standards und die Verwaltung als „Vertragskomponenten“ in einem sogenannten Component Content Management System (CCMS) vorgeschlagen. Im Gegensatz zu einem klassischen CMS, werden Inhalte nicht auf ganzer Dokumentenebene sondern auf feingranularer Komponentenebene beschrieben und verwaltet. Jede Komponente wird als einzelnes Thema (Topic), Konzept oder Inhalt (Asset) repräsentiert und verwaltet. Damit können Vertragsentwurfsbausteine flexible zusammengestellt und durchsucht werden und mit Themen und Konzepten aus der juristischen Gesetzgebung verknüpft/annotiert werden.

Das internationale Standardisierungsgremium OASIS<sup>17</sup> stellt mit der Darwin Information Typing Architecture (DITA<sup>18</sup>) einen offenen Standard für die Inhaltserfassung in CCMS und mit OASIS LegalDocumentML und LegalRuleML zwei Standards für die semantische Auszeichnung von juristischen Wissen bereit, auf deren Basis die semantische Suche und teil-automatisierte Generierung von passenden Vertragstemplates implementiert werden kann.

## 2. Anonymisierungs- und Pseudonymisierungskonzept

Der Betreiber der Plattform muss über geeignete Maßnahmen gemäß dem Stand der Technik die Sicherheit der verarbeiteten Daten und insbesondere auch den Schutz der informationellen Selbstbestimmung der Betroffenen sicherstellen. Immer wenn eine auf der Plattform durchgeführte Studie eine Zusammenführung von Daten einer Person aus verschiedenen Quellen erfordert, ist ein alleiniges Erheben, Einspielen und Verarbeiten anonymer Daten nicht möglich. In diesem Fall verlangt die Datenschutz-Grundverordnung die Nutzung von Verfahren einer Pseudonymisierung (EG 78 DSGVO)<sup>19</sup>.

Die Neutrale Stelle ist darauf ausgelegt, personenbezogene, pseudonymisierte oder anonymisierte Daten zu verarbeiten. Falls die Neutrale Stelle feststellt, dass die Analyse mit anonymisierten oder pseudonymisierten Daten gemacht werden kann, übermittelt sie den Datenbereitstellern den genauen Prozess der Anonymisierung/ Pseudonymisierung, und die Analyse wird nur mit pseudonymisierten Daten durchgeführt. Die damit verbundene Herausforderung einer Pseudonymisierung über verschiedene Systeme hinweg ist nicht neu, sondern tritt vielmehr in ähnlicher Form auch in der „klassischen“ Forschung auf. Die TMF hat entsprechend hierzu bereits vor einigen Jahren ein Pseudonymisierungskonzept für die medizinische Forschung ausgearbeitet, das auf einer technisch-organisatorischen Trennung von identifizierenden Daten (IDAT) und medizinischen Daten (MDAT) beruht. Im IDAT werden dabei personenidentifizierende Daten verwaltet und mit einem oder mehreren Pseudonymen verknüpft. Die medizinischen Daten selbst sind immer nur mit dem Pseudonym indexiert. Mittlerweile liegen mit der sog. Mainzliste der Uniklinik Mainz und MOSAIC der Uniklinik Greifswald auch zwei Implementierungen vor, die u.a. im Kontext der Projekte der Medizininformatikinitiative des BMBF zum Einsatz kommen. Alternativ ist auch eine Umsetzung über die IHE Profile PIX (Patient Identifier Cross-Referencing) und PDQ (Patient Demographics Query) möglich, was einen Einsatz von Standard-Komponenten der Industrie erlauben würde.

Alle genannten Lösungen erfordern die Installation einer unabhängigen vertrauenswürdigen Stelle (trusted third party) zur Verwaltung des IDAT. Dies muss nicht zwingend eine vollständig eigenständige Organisation sein, vielmehr reicht hier die Umsetzung eines Chinese-Wall-Policy nach dem Vorbild des Finanzsektors (strikte gegenseitige Abschottung der Geschäftsbereiche zur Kreditvergabe und zum Investment-Banking) aus. Dieses ist insbesondere über geeignete organisatorische Maßnahmen herzustellen.

Für das NGD sollte für jede auf der Plattform mit pseudonymisiert einzuliefernden Daten durchzuführende Studie einen eigenen IDAT aufsetzen, der technisch von den anderen IDAT-Instanzen anderer Studien getrennt ist. Vor die eigentliche Durchführung der Studie auf der Big-Data-Plattform muss nun eine Pseudonymisierung aller eingeschlossenen Patienten erfolgen. Hierzu führen alle

<sup>17</sup> <https://www.oasis-open.org/>, Zugriff am 27. Juli 2018

<sup>18</sup> [https://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=dita](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=dita), Zugriff am 27. Juli 2018

<sup>19</sup> Ergänzend hierzu wird in Artikel 89 Absatz 1 DSGVO auch zumindest nahegelegt, bei der Datenverarbeitung zu wissenschaftlichen Zwecken immer eine Pseudonymisierung vorzunehmen sofern dies möglich ist.

datenbereitstellenden Einrichtungen nacheinander für jeden eingeschlossenen Patienten die folgenden Schritte durch:

- Übermitteln identifizierender demographischer Daten (Name, Geburtsdatum, KVNR, etc.) an den IDAT
- Der IDAT prüft, ob zu den übermittelten Daten bereits eine Person registriert ist. Ist dies der Fall, wird das für diese Person vergebene Pseudonym zurückgegeben. Ist dies nicht der Fall wird die Person im IDAT registriert und es wird ein Pseudonym für diese Person generiert. Das Pseudonym wird im IDAT gespeichert und an die anfragende Einrichtung zurückgegeben.

Die Big Data-Plattform selbst nimmt nun die Rolle des MDAT ein, d. h. alle angebotenen Einrichtungen verknüpfen die für die Studie eingebrachten Patientendaten ausschließlich mit der vom IDAT abgefragten pseudonymen ID. Es wird empfohlen, dass die Neutrale Stelle des NGD auf den für eine konkrete Studie eingestellten Daten in jedem Fall eine (ggf. weitere) Pseudonymisierung durchführt, die auch durch den IDAT nicht auflösbar ist. Dies ist technisch einfach umsetzbar und kann beispielsweise durch Nutzung einer Einwegfunktion auf den an die Daten gebundenen Identifizierern erfolgen. Ausnahmen sollten nur bestehen, wenn identifizierende Daten (z. B. als Kombination mehrerer demographischer Daten) für den Gegenstand der Studie relevant ist und daher im Rahmen der Studie in granularer Form verarbeitet werden müssen.

Nach dem Abschluss der Studie wird der IDAT vollständig gelöscht.

### 3. Sicherheitskonzept

Eine Big Data-Plattform bringt eine Menge von verschiedenen Daten zusammen in einen Data Lake und gibt Datenwissenschaftlern die Möglichkeit die besten Ergebnisse unter Verwendung von fortgeschrittenen Analysewerkzeugen und entsprechenden Infrastrukturen zu erzielen. Dies bringt aber auch Risiken mit sich: Datenverlust, Datenschutzverletzungen, oder unerlaubte Systemzugriffe sind möglich.

Art. 32 der DSGVO verlangt Risiko-angemessene technisch-organisatorische Maßnahmen nach dem Stand der Technik, die in einem Sicherheitskonzept beschrieben werden müssen. In der Rechnersicherheit spricht man üblicherweise von zwei Kategorien von Aspekten der Datensicherheit:

- (1) Technischer Aspekt, die sogenannten Ringe der Verteidigung. Dies umfasst die Sicherung der physischen Umgebung, die Authentifizierung und Autorisierung auf Ebene jeder Anwendung, die Sicherung des Betriebssystems und die Sicherung und Kontrolle des Zuganges zu den Daten.
- (2) Konzeptuelle Aspekte: fünf Säulen der Sicherheit, diese umfassen Administration, Authentifizierung, Autorisierung, Auditing und die Sicherung und Kontrolle des Zuganges zu den Daten.

Da es sich hierbei um sensible und personenbezogene Daten handelt, müssen hierzu entsprechende Vorkehrungen hinsichtlich der IT-Sicherheit z.B. von Kommunikation, Zugriff auf sowie Speicherung der Daten getroffen werden. Die IT Sicherheit ist dabei eine notwendige, aber nicht hinreichende Voraussetzung für den Datenschutz.

Als De-Facto-Standard für IT-Sicherheit folgt das vorgeschlagene Sicherheitskonzept der IT-Grundschutz-Vorgehensweise mit den IT-Grundschutz-Katalogen und den Empfehlungen der Standard-Sicherheitsmaßnahmen. Die Umsetzung der Vorgehensweise ist im BSI-Standard 100-2<sup>20</sup>: IT-Grundschutz-Vorgehensweise zu finden, sowie dem BSI-Standards 100-1<sup>21</sup>: Managementsysteme für Informationssicherheit und 100-3: Risikoanalyse auf der Basis von IT-Grundschutz. Das NDG betreibt dazu ein „State of-the-Art“ Informationssicherheits-Managementsystem (ISMS) nach der international

<sup>20</sup> [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/BSI-Standard\\_1002.pdf?blob=publicationFile](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/BSI-Standard_1002.pdf?blob=publicationFile) Zugriff am 26. Juli 2018

<sup>21</sup> [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/BSI-Standard\\_1001.pdf?blob=publicationFile](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/BSI-Standard_1001.pdf?blob=publicationFile) Zugriff am 26. Juli 2018

anerkannten Norm ISO/IEC 27001<sup>22</sup> (Kernanforderungen der Kapitel 4 bis 10 sowie der Anforderungen aus den Kapiteln A.5 bis A.18). Dies erlaubt eine ISO 27001- Zertifizierung des NDG als vertrauenswürdiger IT Dienstleister auf der Basis des IT-Grundschutzes durch einen BSI zertifizierten Auditor.

Zusätzlich wird ein IT-Risikomanagement gemäß IEC 80001-1<sup>23</sup> bzw. der deutschen Fassung DIN EN 80001-1 eingeführt. Auch wenn das NDG keinen Schwerpunkt auf die Integration von Medizinprodukten hat, gibt diese Norm einen akzeptierten Standard für die Risikoanalyse und den Unterhalt eines IT-Netzwerke mit sicherer und störungsfrei Kommunikation. Dies umfasst auch die Umsetzung organisatorischen Maßnahmen wie Notfall-Fahrpläne und die Benennung von klaren Verantwortlichkeiten, einen Risikomanager und einen Risikomanagement-Prozess.

Damit folgt das NDG den im Juli 2017 durch den Branchenarbeitskreis Medizinische Versorgung (BAK MV) im Rahmen der UP KRITIS (Umsetzungsplan Kritische Infrastruktur)<sup>24</sup> veröffentlichten Handlungsempfehlungen zur Verbesserung der Informationssicherheit, welche auch für nichtkritische Infrastruktur (nach ITSIG) u.a. ein Information Security Management System (ISMS) gemäß ISO 27001 und ein IT-Risikomanagement vorsehen, wie es beispielsweise die IEC 80001-1 empfehlen.

Mit der Erweiterung des BSI-Gesetzes (BSIG)<sup>25</sup> sind Betreiber „Kritischer Infrastrukturen“ (KRITIS)<sup>26</sup> verpflichtet, sich nach dem Stand der Technik abzusichern und dies gegenüber dem Bundesamt für Sicherheit in der Informationstechnik (BSI) nachzuweisen. Das NDG kann optional am UP KRITIS teilnehmen und die gemäß § 8a BSIG angemessenen organisatorischen und technische Vorkehrungen gemäß dem Stand der Technik treffen, welcher durch den baldig zu erwartenden branchenspezifische Sicherheitsstandard (B3S<sup>27</sup>) des BAK MV konkretisiert werden soll. Dies impliziert auch den Aufbau von geeigneten Detektionsmechanismen für außergewöhnliche IT-Störungen, die dann freiwillig über die Meldestelle der Allianz für Cyber-Sicherheit dem BSI gemeldet werden.

Bei der Umsetzung der besonders sicherheitskritischen Neutralen Stelle geht das von uns vorgeschlagene technische Sicherheitskonzept über die traditionelle Sicherung einer Big Data-Plattform durch die Implementierung eines Sandkastenmodells (Sand Box) hinaus. In diesem Konzept gibt es eine sehr starke Sicherung der Umgebung, wo alle Maschinen, die für die Analyse und Speicherung verwendet werden, keinen direkten Zugang zum Internet haben.

Um die maximale Sicherheit gewährleisten zu können, kommt bei der Plattform der „Air Gap“ genannte Prozess zum Einsatz. Hierbei hat die Plattform keinerlei Zugang zum Internet und die Kommunikation findet über verschlüsselte Medien statt. Dabei kommt das vom BSI vorgeschlagene Verfahren RSA zum Einsatz, bei dem der privaten Schlüssel innerhalb der Plattform verbleibt und nur der öffentliche

<sup>22</sup> DIN EN ISO/IEC 27001:2005 „Informationstechnik - IT-Sicherheitsverfahren - Informationssicherheits-Managementsysteme - Anforderungen“

<sup>23</sup> DIN EN 80001-1:2011 Anwendung des Risikomanagements für IT-Netzwerke, die Medizinprodukte beinhalten

<sup>24</sup>

[https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Kritis/Fortschreibungsdokument.pdf?\\_\\_blob=publicationFile&v=2](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Kritis/Fortschreibungsdokument.pdf?__blob=publicationFile&v=2)  
Zugriff am 26. Juli 2018

<sup>25</sup>

[https://www.bgbl.de/xaver/bgbl/start\\_xav?startbk=Bundesanzeiger\\_BGBl#\\_bgbl\\_%2F%2F\\*%5B%40attr\\_id%3D%27bgbl\\_115s1324.pdf%27%5D\\_1532683940106](https://www.bgbl.de/xaver/bgbl/start_xav?startbk=Bundesanzeiger_BGBl#_bgbl_%2F%2F*%5B%40attr_id%3D%27bgbl_115s1324.pdf%27%5D_1532683940106) Zugriff 27. Juli 2018

<sup>26</sup> [https://www.kritis.bund.de/SharedDocs/Downloads/Kritis/DE/BSI\\_Kritisverordnung\\_Final.pdf?\\_\\_blob=publicationFile](https://www.kritis.bund.de/SharedDocs/Downloads/Kritis/DE/BSI_Kritisverordnung_Final.pdf?__blob=publicationFile)  
Zugriff 27. Juli 2018

<sup>27</sup> [https://www.bsi.bund.de/DE/Themen/Industrie\\_KRITIS/KRITIS/IT-SiG/Was\\_tun/Stand\\_der\\_Technik/B3S\\_BAKs/B3S\\_BAKs\\_node.html](https://www.bsi.bund.de/DE/Themen/Industrie_KRITIS/KRITIS/IT-SiG/Was_tun/Stand_der_Technik/B3S_BAKs/B3S_BAKs_node.html) Zugriff am 27. Juli 2018

Schlüssel nach außen hin bekannt ist. Selbst bei Diebstahl des Datenträgers, ist kein Zugriff auf diese möglich.

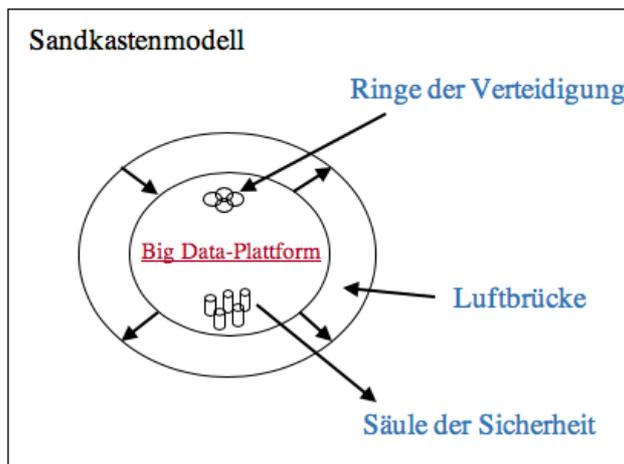


Abbildung 10: Sicherheitskonzept

Neben dieser nach außen hin gerichteten Technik, muss das System auch in seinem internen IT-Infrastrukturbetrieb gegenüber unbefugtem Zugriff geschützt werden. Die im Weiteren vorgestellten und dem aktuellen Stand der Technik entsprechenden Methoden folgen dabei der FHIR Sicherheitsempfehlung<sup>28</sup> der HL7.

Damit die maximale Sicherheit der Datenübertragen gewährleistet werden kann, wird dies wie in der Communications Security Policy der FHIR Security mittels eines verschlüsselten Kanals per HTTPS unter Verwendung von TLS/SSL realisiert. Hierbei kommt eine RSA sowie eine AES Verschlüsselung zum Einsatz, welche den aktuellen Stand der Technik darstellt. Durch eine Private Key Infrastructure, welche per RSA realisiert wird, ist ein sicherer Austausch des Schlüssels für die nachfolgend eingesetzte AES Verschlüsselung gewährleistet. Die Neutrale Stelle erstellt ein Public/Private Key Paar und übergibt dann den Public Key dem Datenbereitsteller zur Verschlüsselung der Daten zusammen mit einer Beschreibung des Verschlüsselungsprozesses. Der Private Key bleibt hingegen bei der Neutralen Stelle, damit die Daten nur in der gesicherten Neutralen Stelle entschlüsselt werden können. Beide Verfahren entsprechen der Technischen Richtlinie BSI TR-02102-1<sup>29</sup> des Bundesamts für Sicherheit in der Informationstechnik (BSI) für Kryptografischen Verfahren.

Neben der Verschlüsselten Kommunikation muss auch sichergestellt werden, dass nur autorisierte Benutzer Zugriff zum System und auf die für sie zugelassenen Ressourcen erhalten dürfen. Die FHIR Authentication und Authorization Policy sehen hierzu eine Authentifizierung mittels OAuth vor. Dies deckt sich mit dem IT-Grundschutzkatalog M 4.456<sup>30</sup> „Authentisierung bei Web-Services“ des BSI, welches hierbei diese per OAuth oder per SAMLv2 vorsieht. Beide Verfahren basieren auf der Verwendung eines Zugriffs-Tokens, der vom Client als Repräsentation der Autorisierung übermittelt wird und diesem dann den Zugriff auf geschützte Ressourcen des Servers ermöglicht. Der Token ermöglicht eine feingranulare Festlegung der Berechtigungen.

Diese Aspekte und wie sie technisch umgesetzt werden, steht z.B. detailliert in den Anleitungen von Big Data-Plattform Systemanbietern wie Hortonworks oder Cloudera. Von besonderem Interesse ist hierbei die Integration von Apache Knox. Apache Knox ist ein REST API Gateway, der eine Interaktion mit Hadoop-Clustern ermöglicht und wie ein Reverse Proxy arbeitet. Es erweitert das existierende Hadoop Security-Modell um Authentifizierungsrollen für angemeldete Benutzer. Darüber hinaus stellt es weitere Funktionen zur Cluster-Verwaltung, Monitoring- und Automatisierungsfunktionen zur Verfügung wie AD- und LDAP-Authentication-Provider, sowie

<sup>28</sup> <https://www.hl7.org/fhir/security.html> Zugriff am 26. Juli 2018

<sup>29</sup> [https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr02102/index\\_hm.html](https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr02102/index_hm.html) Zugriff am 26. Juli 2018

<sup>30</sup> <https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/content/m/m04/m04456.html> Zugriff am 26. Juli 2018

Funktionen für Auditing und Service-Level-Autorisierung. Knox kann sogar mit via Kerberos abgesicherten Hadoop-Clustern zusammenarbeiten. Ferner ist eine Anbindung zu SAMLv2 vorhanden. Zwar empfiehlt die FHIR Security eine Authentifikation mittels OAuth, SAMLv2 bietet eine vergleichbare Sicherheit und ist ebenfalls vom BSI empfohlen. Die nachfolgende Abbildung verdeutlicht, wo Knox in der Hadoop-Welt angesiedelt ist.

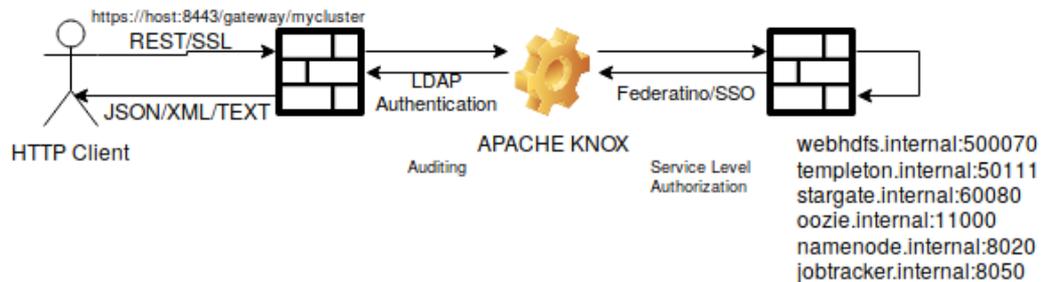


Abbildung 11: Einsatz von Knox

Ferner ermöglicht Apache Knox auch eine Anbindung an Apache Ranger. Ranger bietet ein zentralisiertes Sicherheitsmanagement für Apache Hadoop, wodurch ein feingranularer Zugriff auf die einzelnen Komponenten sowie Daten geregelt wird. Administratoren können hierbei Policies für einzelne Benutzer oder Gruppen festlegen, welche dann über den vollständigen HDP Stack vollzogen wird. Ferner stellt Ranger KMS einen skalierbaren Kryptographie Management Service bereit, welcher für die HDFS „data at rest“ Verschlüsselung notwendig ist.

HDFS stellt die fortschrittlichste Verschlüsselungsform dar, die in Hadoop möglich ist. Im Gegensatz zur einfachen Verschlüsselung auf Datenträger-Ebene, werden „encryption zones“ geschaffen, die die Verschlüsselung auf Datei und Ordner-Ebene mit unterschiedlichen Schlüsseln und Rechten realisiert. Dieser bei Erstellung der Zone erzeugte Schlüssel gilt dann nur für diese einzelne Zone. Daten liegen physisch stets verschlüsselt vor und werden nur bei Lese oder Schreibzugriff transparent entschlüsselt. Dieses Vorgehen wird auch vom BSI in der Empfehlung M4.433 „Einsatz von Datenträger-verschlüsselung“<sup>31</sup> vorgeschlagen.

Neben der Sicherheit gegenüber Diebstahl sensibler Daten ermöglicht eine Verschlüsselung sensibler Daten auch ihre vereinfachte Löschung. Für die Löschung der Informationen werden zuerst ihre zugeordneten Schlüssel gelöscht und danach werden die Daten mehrfach gemäß der Empfehlung M 2.515<sup>32</sup> vom BSI zur „Datenschutzgerechten Löschung und Vernichtung von Datenträgern“ physikalisch überschrieben.

Für eine maximale Sicherheit der Löschung sieht das Konzept noch eine weitere Funktion bei besonders sensiblen Daten vor. Die Big-Data-Plattform wird auf isolierten Servern kopiert, die Analyse geschieht dann nur auf für diesen Zweck bereit gestellten und vom Netzwerk entkoppelten Systemen, und danach werden alle Datenträger und Arbeitsspeicher mit dafür speziell vorgesehenen Maschinen der Empfehlung M 2.167<sup>33</sup> des BSI zur „Auswahl geeigneter Verfahren zur Löschung oder Vernichtung von Daten“ physisch zerstört. Dies ist jedoch nur für höchst sensible Daten sinnvoll, da damit entsprechend große Mehrkosten für Material und Arbeitszeit verbunden sind.

<sup>31</sup>

[https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/\\_content/m/m04/m04433.html?nn=6610622](https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m04/m04433.html?nn=6610622) Zugriff am 26. Juli 2018

<sup>32</sup> [https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/\\_content/m/m02/m02515.html](https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m02/m02515.html) Zugriff am 26. Juli 2018

<sup>33</sup> [https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/\\_content/m/m02/m02167.html](https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m02/m02167.html) Zugriff am 26. Juli 2018

#### 4. Bewertung

##### a) *Machbarkeitsbetrachtung der technischen Plattform*

Als bestehende Referenzlösung zur Bewertung der technischen Machbarkeit und der technischen Risiken wird das Informationssystem Versorgungsdaten betrachtet. Das Informationssystem Versorgungsdaten am DIMDI<sup>34</sup> stellt auf Basis der Datentransparenzverordnung seit Anfang 2014 Routinedaten der gesetzlichen Krankenkassen (DaTraV Daten) für gesetzlich festgelegte Nutzergruppen zur Verfügung. Das Antragsverfahren bietet in der aktuellen 2. Ausbaustufe zwei Möglichkeiten: Der Antragsteller kann sich auf die Formulierung der Forschungsfrage sowie die Beschreibung der Methodik beschränken und die Skripterstellung den Analysten des Informationssystems Versorgungsdaten im DIMDI überantworten. Alternativ besteht die Möglichkeit, bei Antragstellung ein lauffähiges Skript (aktuell in SQL (Oracle 11g R2) oder SAS (9.4 64bit)) einzureichen. Dabei werden die DaTraV-Daten mit von den Nutzungsberechtigten bereitgestellten Programmen ausgewertet. Dieses Skript wird nach inhaltlicher Prüfung auf dem DaTraV-Volldatensatz ausgeführt. Bei syntaktischen oder inhaltlichen Auffälligkeiten kann das Skript in Abstimmung mit dem DIMDI nachträglich angepasst werden. Bei beiden Möglichkeiten erhält der Antragsteller im Rahmen eines Bescheids eine so genannte Ergebnismenge. Zur Wahrung des Identifikationsschutzes der Versicherten darf diese nur aggregierte Daten enthalten und muss faktisch anonymisiert sein. In der geplanten dritten Ausbaustufe wird es möglich sein, eine On-Site-Nutzung zu beantragen bei der Nutzungsberechtigte die pseudonymisierten Einzeldatensätze in den Räumlichkeiten des DIMDI selbst auswerten. Voraussetzung ist, dass eine Analyse auf Basis von anonymisierten Daten nicht möglich ist.

Grundsätzlich ist im, für das NDG vorgeschlagenen, technischen Lösungsmodell einer Big Data-Plattform in Verbindung mit einem Datenkatalog ein ähnlicher drei-stufiger Ausbau, wie für die Datentransparenzstelle des DIMDI möglich. Die zentralen Eigenschaften einer semantisch erweiterten Big Data Lösung erlauben dabei mit der Heterogenität der Datenquellen umzugehen und die notwendigen Informationen über die angebotenen Datensätze über den Datenkatalog bereit zu stellen. Im Evaluationsbericht für das Informationssystem Versorgungsdaten wurde eine Reihe von Problemen identifiziert. Zu den primären technischen zählen:

1. nur für die Lauffähigkeitsprüfung von Skripten verwendbarer Beispieldatensatz, keine inhaltliche Prüfung möglich
2. fehlendes Datenmanual und fehlende Basisauswertungen / Standardauswertungen / Datencharakterisierungen
3. Inadäquates Analyseverfahren: Bei dem aktuellen Antragsmodus ist lediglich die Übermittlung einer einzigen Ergebnismenge per Bescheid am Antragsende vorgesehen, (mehrmalige) Zwischenergebnisse dürfen nicht übermittelt werden. Notwendige iterative Plausibilisierungsschritte sind daher nicht durchführbar und Auswertungsskripts können nicht auf Basis von Zwischenergebnissen angepasst und optimiert werden. Hieraus folgt, dass zuweilen komplexe und unterschiedliche Modelle parallel programmiert werden, die ihrerseits wieder einen erheblichen Prüfaufwand nach sich ziehen.
4. Fehlender Zugriffe der Nutzer über gesicherte VPN-Verbindungen um die Skripte über den Remotezugang zu entwickeln, zu testen, auf Performance zu optimieren und interaktiv zu modifizieren

Im Folgenden wird kurz dargestellt, wie diese technisch adressiert werden können

1. Erzeugung eines Beispiel- und Testdatensatzes zur Entwicklung und inhaltlichen Prüfung des Auswertungsskripts durch eine Zufallsstichprobe oder Generierung eines modellbasierten Beispieldatensatzes aus dem Originaldatensatz
2. Datenmanual mit semantischer und inhaltlicher Beschreibung des Datenkörpers im Datenkatalog und Beispielskripte für Basisauswertungen

<sup>34</sup> <https://www.dimdi.de/dynamic/de/weitere-fachdienste/versorgungsdaten/> Zugriff am 26.Juli 2018

3. Übermittlung iterativer Zwischenergebnisse über die Big Data-Plattform
4. Einsatz eines personalisierten Token-Generators für einen VPN-gesicherten Remotezugriff kann sichergestellt werden, dass der Zugriff auf die Daten oder Untermengen davon nur durch vertraglich benannte und autorisierte Personen erfolgt.

b) *Grenzen der semantischen Interoperabilität*

Semantische Interoperabilität kann nie vollständig hergestellt werden. Die wesentlichen Gründe hierfür liegen in bereits in der ursprünglichen Kodierung einer Information potenziell enthaltenen Unschärfen, die potenziell durch das Mapping weiter verstärkt werden. Hierzu ein Beispiel aus der oben zitierten Fallstudie aus der Schweiz:

Textauszüge aus Fallbeispiel	Datemodul: Probleme			
	ePA-AC	NANDA-I	ICNP	SNOMED CT
Frau Muster kann sich die Kompressionsstrümpfe trotz Anleitung nicht selbst anziehen...	E3_I_0118 stark beeinträchtigte Fähigkeit den Unterkörper zu kleiden	00109 Selbstversorgungsfähigkeit Sich Kleiden (Dressing self-care deficit)	10027578 Beeinträchtigte Fähigkeit sich zu kleiden (Impaired Ability To Dress) 10006586 Elastischer Strumpf (Elastic Stocking)	425353005 Difficulty with dressing lower body 334863006 Elastic hosiery

Abbildung 12: Beispiel semantische Interoperabilität

Wie das Beispiel zeigt, entstehen Unschärfen bereits durch die per se begrenzte Ausdrucksmächtigkeit des gewählten Codesystems. ICNP<sup>35</sup> und SNOMED CT<sup>36</sup> können den zu kodierenden Sachverhalt durch Kombination von zwei Konzepten recht gut wiedergeben während bei ePA-AC<sup>37</sup> und NANDA-I<sup>38</sup> bestimmte Details in unterschiedlichem Maße wegfallen. Eine in der Quelle über NANDA-I kodierte Information ist damit semantisch unschärfer als eine bereits von vorne herein in SNOMED CT kodierte Information. Hier ist mit Blick auf die durchzuführende Datenanalyse zu entscheiden, welcher Präzisionsgrad erforderlich ist, um zu den gewünschten Ergebnissen zu gelangen.

Will man die im obigen Beispiel dargestellten Codes auf einen Referenz-Code abbilden, so wird man üblicherweise beim grobgranularsten Konzept (hier 00109 aus NANDA-I) landen, was jedoch bedeutet, dass bei dem Mapping aus ePA-AC, ICNP und SNOMED CT eine Vielzahl an Informationen verloren geht. Betrachtet man das Ziel des Mappings jedoch nicht als Konzept, sondern als System, so lassen sich die einzelnen Konzepte mit deutlich geringerem Informationsverlust auf SNOMED CT als Referenz-Codesystem abbilden:

ePA-AC : E3\_I\_0118 --> SNOMED CT : 425353005

NANDA-I : 00109 --> SNOMED CT : 284972002 (Difficulty performing dressing activity)

ICNP : 10027578 --> SNOMED CT : 284972002 (Difficulty performing dressing activity)

ICNP : 10006586 --> SNOMED CT : 334863006

Wie dargestellt, lassen sich die aus verschiedenen Codesystemen stammenden Konzepte zwar nicht auf ein Konzept aber in ein System (hier: SNOMED CT) abbilden, in dem anschließend die weitere semantische Verarbeitung stattfinden kann. Aber auch hierbei verbleibt eine Rest-Unschärfe, da die in ePA-AC festgehaltene „stark beeinträchtigte Fähigkeit“ auf eine „Schwierigkeit (Difficulty)“ abgebildet wurde, d. h. die Qualifizierung „stark“ beim Mapping weggefallen ist.

Auch für das Mapping in ein Referenzsystem hinein gilt somit das oben gesagte, dass Unschärfen nie vollständig vermieden werden können und es einer Bewertung durch einen Menschen bedarf, ob eine Unschärfe für eine durchzuführende Auswertung der Daten akzeptabel ist oder nicht.

<sup>35</sup> <http://www.icn.ch/what-we-do/ICNP-Browser/> Zugriff 27. Juli 2018

<sup>36</sup> <https://www.snomed.org/snomed-ct> Zugriff am 27. Juli 2018

<sup>37</sup> <https://www.epa-cc.de/> Zugriff am 27. Juli 2018

<sup>38</sup> <http://www.nanda.org/> Zugriff am 27. Juli 2018

Hierbei ist jedoch festzuhalten, dass insbesondere Referenzsysteme mit der Möglichkeit einer Post-Koordination Mechanismen bieten, über die sich Unschärfen potenziell durch Bildung neuer Konzepte über eine semantische Erweiterung bestehender Konzepte verringern oder gar beheben lassen. Von den beiden oben aufgeführten Codesystemen, die sich durch ihre Prä-Koordination gut als Referenzsysteme eignen, bietet nur SNOMED CT auch die Möglichkeit einer Post-Koordination von Konzepten.

#### c) *Bewertung des technischen Sicherheitskonzepts*

Das vorgeschlagene technische Sicherheitskonzept folgt den Vorschlägen der FHIR Sicherheitsempfehlung. Diese ist nicht als ein Security-Protokoll zu verstehen, es beschreibt vielmehr den jeweiligen, für verschiedene Bereiche in Frage kommenden Technologien. Diese einzelnen technischen Lösungen entsprechen oder übersteigen dabei den Stand der Technik für sichere Systeme und insbesondere die Empfehlungen für den IT Grundschutz des BSI.

Durch den „Sandkastenansatz“ kann aus rein technischer Sicht den höchsten Ansprüchen bezüglich Datenschutzes und Sicherheit genügt werden. Jedoch sind Probleme aufgrund personeller Fehler nicht auszuschließen, so dass das die Plattform bedienende Personal auch entsprechend geschult werden muss. Mit der optionalen Teilnahme am UP KRITIS würde das NGD entsprechende organisatorische Maßnahmen umsetzen und wäre bereits jetzt zukunftsorientiert aufgestellt.

Die Integration und Auswertung weniger sensibler, nicht-personeller und offener Webdaten (Open Data, Linked Data<sup>39</sup>), als auch das (in einer weiteren Ausbaustufe angedachte) „Cognitive Computing“ erfordern aus Kosten- und Praktikabilitätsgründen allerdings eine gewisse Öffnung der Kommunikation, zumindest für die Analysten der Neutralen Stelle, um auf im Web vorhandene Wissensquellen für die Analysen zuzugreifen zu können.

Denkbar ist auch die Nutzung der Telematikinfrastruktur (TI), welche es perspektivisch erlauben würde über die TI sicher online zu kommunizieren.

## D. *Governance, Organisation und rechtliche Ausgestaltungsoptionen*

Nachfolgend werden die Governancestrukturen, das Organisationsmodell sowie die rechtlichen Ausgestaltungsoptionen des „Netzwerk Gesundheitsdaten“ beschrieben. Darunter werden verstanden:

- die Rollen in der Governance des NGD,
- die in der Governance abzubildenden Kernprozesse,
- die möglichen Rechtsformen des NGD und Ausgestaltung der Rechtsbeziehungen,
- die rechtlich-organisatorischen Anforderungen der Umsetzung,
- die Ausgestaltung des IT-Sicherheits- und Datenschutzmanagements und
- die Ausgestaltung der Stakeholder-, Gewährleistungs- und Aufsichtsfunktionen

Die Governance- und Organisationsstrukturen sowie die rechtliche Ausgestaltung des „Netzwerk Gesundheitsdaten“ orientiert sich an nachstehenden Prämissen und Kriterien:

#### *Steuerung des „Netzwerk Gesundheitsdaten“*

Das „Netzwerk Gesundheitsdaten“ ist eine langfristig angelegte, im Geschäftsbereich des Bundesministeriums für Gesundheit institutionalisierte Einrichtung, die von keinem Datenbereinsteller oder Datennutzer beherrscht wird. Die Governance sieht besondere Kontroll- und Steuerungsmöglichkeiten durch die öffentliche Hand vor, um die Integrität, Vertrauenswürdigkeit und Gemeinwohlorientierung des NGD sicherzustellen. Um die langfristige Bedarfsgerechtigkeit des NGD sicherzustellen, wird ein Konzept zur Rückkopplung der Stakeholder (Datennutzer und Datenbereinsteller) an das NGD entwickelt und etabliert. Die Letztverantwortung für die **Steuerung** soll bei der öffentlichen Hand liegen.

<sup>39</sup> <http://linkeddata.org/> Zugriff am 27. Juli 2018

### *Unabhängigkeit der Neutralen Stelle*

Die für die eigentlichen Datenauswertungen verantwortliche Neutrale Stelle besitzt bedingt durch die Sicherungsanforderungen notwendige eine eigene organisatorische Stellung innerhalb des NGD. Durch die Organisationsstruktur des NGD wird die Einflussnahme von Stakeholdern/NGD-Nutzern verhindert. Sie wird als Vertrauenskern des NGD ausgestaltet und unterliegt zusätzlich der Aufsicht der beaufsichtigenden Behörden.

### *Skalierbarkeit*

Der Aufbau des NGD wird sich schrittweise vollziehen. Die Organisation und Steuerung des NGD sind daher in Bezug auf die Skalierbarkeit sowie sich verändernde Anforderungen und Rahmenbedingungen anpassungsfähig (eine Teil-Automatisierung einzelner Prozesse ist vorgesehen) und erweiterbar auszugestalten, um eine operative Flexibilität (z.B. bei der Gewinnung von Personal) zu ermöglichen.

### *Agilität*

Die Möglichkeiten zur Durchführung von Big-Data-Analysen werden sich in den kommenden Jahren schrittweise erweitern. Das NGD muss sich hieran anpassen können. Die gewählte Rechts- und Organisationsform des NGD soll daher eine gewisse Agilität ermöglichen.

### *Effizienzpotenzial*

Mit zunehmender Nutzung des „Netzwerks Gesundheitsdaten“ können sich Synergien zwischen Datenbereitstellern und Datennutzern sowie Effizienzpotentiale für die Durchführung von Auswertungen ergeben. Diese Potentiale sollen nutzbar gemacht werden.

### *Gesetzlicher Auftrag*

Zur Sicherstellung von Auftrag, Governance und Grundstruktur des NGD soll eine gesetzliche Regelung geschaffen werden, wie sie auch in anderen Bereichen der Datenauswertung des Gesundheitswesens existiert (vgl. hierzu: DIMDI und DaTraV; BIG und das BIGG; RKI und IfSG).

## 1. Rollen in der Governance des NGD

Die Rollen sind wie folgt definiert:

- **NGD-Organisation**  
Das „Netzwerk Gesundheitsdaten“ ist eine Organisation, die Zusammenführung und Auswertungen von Gesundheitsdaten ermöglicht, indem eine digitale Plattform zur Kontaktaufnahme zwischen Partnern, eine Dienstleistung für Vertragsabschlüsse und eine Dienstleistung für Datenzusammenführung und Datenanalyse (Neutrale Stelle) etabliert wird.
- **Neutrale Stelle**  
Die Neutrale Stelle ist ein innerhalb der NGD-Organisation besonders ausgestalteter Funktionsbereich, der die Leistung der Datenzusammenführung Datenanalyse bereitstellt.
- **NGD-Nutzer**  
NGD-Nutzer sind als Datennutzer (DN) und/oder Datenbereitsteller (DB) Vertragspartner des „Netzwerk Gesundheitsdaten“, die die genannten Leistungen des NGD auf Basis entsprechenden Vereinbarungen in Anspruch nehmen. Zunächst wird nur von NGD-Nutzern in Deutschland ausgegangen, eine Erweiterung auf Nutzer innerhalb der Europäischen Union soll möglich sein.
- **Stakeholder (Institutionalisiert)**  
Stakeholder sind die Vertreter von Nutzergruppen des NGD, die unterschiedliche Interessenträger des Gesundheitswesens abbilden. Hierzu gehören öffentliche und private Leistungserbringer im Gesundheitswesen, gesetzliche und private Krankenversicherungen, Bundes-, Landes- und Kommunalbehörden, Privatunternehmen im Gesundheitssektor

(z.B. Pharmaunternehmen) sowie private und öffentliche Einrichtungen der Gesundheitsforschung.

- **Träger**  
Träger des NGD ist eine gesetzlich bestimmte Einrichtung, die einer öffentlich-rechtlichen Verantwortung unterliegt und im Geschäftsbereich des BMG angesiedelt ist.
- **Aufsichtsbehörden**  
Aufsichtsbehörden sind die für die Aufsicht über die Tätigkeit des NGD zuständigen Behörden im Bereich des Gesundheitswesens, des Datenschutzes und der IT-Sicherheit.

## 2. In der Governance abzubildende Kernprozesse

In der Governance des NGD sind einige zentrale Kernprozesse abzubilden. Die Organisation des NGD einschließlich der Ausgestaltung von Rechtsform und Rechtsbeziehungen hängen von der adäquaten Abbildbarkeit der Prozesse und von der zu gewährleistenden Sicherheit ab. Im Einzelnen werden folgende, für die Governance-Entscheidungen zentralen Prozesse betrachtet:

### a) *Wirtschaftliche Steuerung*

Für die dauerhafte Leistungsfähigkeit des NGD ist eine hinreichende materielle Ausstattung erforderlich. Der Prozess der wirtschaftlichen Steuerung stellt sicher, dass NGD über ausreichende materielle Mittel und Personal für die Wahrnehmung der Aufgabe. Zu dem Prozess gehört eine Grundfinanzierung, eine bedarfsgerechte Mittel- und Haushaltsplanung, die Entwicklung, Fortschreibung und Festsetzung von Finanzierungsmodellen (einschließlich Umlage- bzw. Gebührenmodellen), die Personalbedarfsplanung und Personalbemessung, die Bereitstellung von Investitionsmitteln sowie die Budget- und Erfolgskontrolle.

### b) *Technisch-organisatorische Steuerung*

Dieser Prozess stellt die dauerhafte Leistungsfähigkeit im Hinblick auf die Serviceerbringung und Servicequalität der drei Kerndienstleistungen des NGD sicher. Hierzu gehören die Definition von Services und Servicequalität und ihre organisatorische und technische Abbildung. Es wird davon ausgegangen, dass das NGD ein internes IT-Servicemanagement einführt, das auch eine Dienstleistersteuerung umfasst.

### c) *Untervertragnahme von Nutzern*

Die Entscheidung über die Untervertragnahme eines Nutzers durch das NGD erfolgt auf Grundlage gesetzlich festgelegter Kriterien. Eine Untervertragnahme eines Nutzers unterbleibt für den Fall der Nichterfüllung der festgelegten Kriterien oder den Verstoß gegen diese Kriterien.

### d) *Entscheidung über Auswertungen*

Dieser Kernprozess umfasst die fachliche und rechtliche Prüfung der Durchführbarkeit und Zulässigkeit der von Nutzern angefragten Daten-Analysen über das NGD. Dazu gehört auch die Prüfung der Vereinbarkeit einer Daten-Analyse mit dem Zweck und den Nutzungsbedingungen des NGD sowie die datenschutzrechtliche Zulässigkeit der Auswertung.

### e) *Fachliche Qualitätssicherung*

Alle Prozesse der verschiedenen Funktionsbereiche unterliegen einem Qualitätsentwicklungskonzept. Dieses schließt über das NGD hinaus auch die Nutzer des NGD mit ein, damit im Verlauf des Betriebes des NGD eine systematische Qualitätsentwicklung der vorhandenen und potentiell nutzbaren Daten angestrebt werden kann.

### f) *Gewährleistung der Kernziele Vertrauenswürdigkeit und Integrität*

Kernziele des NGD neben der fachlichen Qualität sind die belastbare Gewährleistung von Vertrauenswürdigkeit und Integrität der Prozesse des Umgangs mit persönlichen Daten. Hierfür muss das NGD ein internes Kontrollsystem errichten, dass die Einhaltung dieser Anforderungen ständig prüft

und ggf. nachsteuert. Hierzu gehört auch die Einrichtung eines Systems der Korruptionsprävention, das sich an den Vorgaben für die Bundesverwaltung orientieren kann.

#### g) *Zusammenarbeit mit Aufsichtsbehörden*

NGD muss im Zusammenwirken mit den Gesundheits-, Datenschutz-, IT-Sicherheits- und sonstigen Aufsichten eine hohe Belastbarkeit und große Transparenz an den Tag legen. Hierfür sind in der Organisation geeignete Vorbereitungen zu treffen.

### 3. *Rechtsform des NGD und Ausgestaltung der Rechtsbeziehungen*

Neben der Berücksichtigung der Kriterien der effektiven Organisation und Steuerung, der abgebildeten Kernprozesse der Governance sowie unter Berücksichtigung der zuvor aufgeführten Prämissen hat sich die Ausgestaltung des NGD an grundlegenden rechtlichen Vorgaben zu orientieren. Hierbei besteht ein Spielraum zur Auswahl einer Rechtsform.

Die Rechtsformen sind danach zu unterscheiden, ob sie dem privaten oder dem öffentlichen Recht zuordnen sind. Als Rechtsform des öffentlichen Rechts, in denen Forschungseinrichtungen betrieben werden, stehen dem NGD auf Bundesebene - neben der unmittelbaren Bundesverwaltung - die folgenden Rechtsformen offen:

- die Anstalt des öffentlichen Rechts,
- die Körperschaft des öffentlichen Rechts, und
- die öffentlich-rechtliche Stiftung;

Als Rechtsformen des Privatrechts, in denen Forschungseinrichtungen betrieben werden, stehen dem NGD die folgenden Rechtsformen offen:

- der eingetragene Verein,
- die privatrechtliche Stiftung,
- die (g)GmbH, und
- die AG.

Wesentliche Anforderungen an die Auswahl der geeigneten Rechtsform sind

- die optimale Funktionalität und maximale Sicherungsmöglichkeit
- die öffentliche Steuerbarkeit,
- die Berücksichtigungsfähigkeit mehrerer Stakeholder in Nutzbarmachung des NGD,
- der Aufwand der Gründung,
- die Möglichkeit der Ausgestaltung spezifischer Anforderungen von Teilbereichen,
- die Flexibilität im Hinblick auf Personalaufbau und technisch-organisatorische Leistungsfähigkeit sowie
- die Skalierbarkeit.

Aufgrund dieser Anforderungen scheiden bestimmte Rechtsformen aus (z.B. eine Behörde der unmittelbaren Bundesverwaltung im Hinblick auf die Einbindung von Stakeholdern, die Bundes- oder – im Falle der Beteiligung von Ländern an der AÖR – Landesbehörden sind) oder erweisen sich als ungünstig (z.B. eine privatrechtliche Stiftung im Hinblick auf die Flexibilität, ein Verein im Hinblick auf eine Haftungsbeschränkung).

Für die Machbarkeitsbetrachtung werden im Folgenden die Modelle „Anstalt des öffentlichen Rechts“ und die „gGmbH im Besitz der öffentlichen Hand“ auf ihre Eignung untersucht.

#### a) *Rechtsfähige Bundesanstalt*

##### i. *Allgemeine Merkmale:*

Die rechtsfähige Bundesanstalt ist eine öffentlich-rechtliche Verwaltungseinrichtung, die mit einer öffentlichen Aufgabe betraut ist. Die Aufgabe wird gesetzlich zugewiesen und das Verhältnis zwischen der jeweiligen Anstalt und ihren Nutzerinnen und Nutzern durch eine entsprechende Satzung geregelt. Oberstes Ziel ist die Sicherung einer nachhaltigen Daseinsvorsorge bei wirtschaftlicher Unternehmensführung und unter Beachtung gesellschaftspolitischer Rahmenbedingungen

(Sozialauftrag). Die Anstalt des öffentlichen Rechts (AöR) wird nicht mitgliederschaftlich organisiert, sondern bietet eine Benutzungsmöglichkeit. Sie untersteht der Aufsicht des Anstaltsherren (Träger). Für den Jahresabschluss und den Lagebericht gelten die allgemeinen Bestimmungen für große Kapitalgesellschaften. Organe einer AöR sind typischerweise ein Verwaltungsrat und ein Vorstand. Eine Mitwirkung von Stakeholdern kann im Verwaltungsrat erfolgen. In den Verwaltungsrat sollten nur Bundes- oder – im Falle der Beteiligung von Ländern an der AöR – Landesbehörden aufgenommen werden.

Beispiele für Anstalten des öffentlichen Rechts im Bereich der Digitalisierung sind die Bundesanstalt Digitalfunk für Behörden und Organisationen mit Sicherheitsaufgaben (BDBOS), errichtet durch ein Bundesgesetz, oder die in Gründung befindliche „Föderale IT-Kooperation“ (FITKO), die durch einen Staatsvertrag errichtet wird. In beiden Fällen sind die Länder als weitere Stakeholder im Verwaltungsrat vertreten. Ein weiteres Beispiel ist die geplante Umwandlung des Informationstechnikzentrums Bund (ITZ Bund) in eine AöR; dort soll neben dem federführenden BMF andere Bundesressorts als Stakeholder im Verwaltungsrat vertreten sein.

## ii. Vor- und Nachteile

### Anstalt des öffentlichen Rechts (AöR)

Vorteile	Nachteile
Möglichkeit, öffentlich-rechtlich zu handeln, insbesondere durch Verwaltungsakte	Geringere Flexibilität bei der Rekrutierung und Reduzierung von Personal, da sie dem öffentlichen Dienstrecht unterliegt
Möglichkeit der Beschäftigung von Beamten	Typischerweise Tarifbindung (TVöD)
Hohe politische Einflussmöglichkeiten und optimale Kontrollfunktion durch entsprechendes Bundesministerium	Zwar rechtliche und organisatorische Verselbständigung wie bei der (g)GmbH möglich, dafür aber erhebliches Organisationsermessen des Bundes bezüglich Kontroll- und Steuerungsfähigkeit
Anschluss- und Benutzungszwang möglich	Bindung an das Vergaberecht
Höhere öffentliche Reputation gegenüber der (g)GmbH	Möglicherweise umständliche Entscheidungsstrukturen aufgrund erforderlicher Genehmigungen durch staatliche Aufsicht
Gesetzgeber hat einen größeren Gestaltungsspielraum bei der Governance	

## b) (g)-GmbH im Besitz der öffentlichen Hand

### i. Allgemeine Merkmale:

Eine GmbH ist eine selbständige juristische Person des Privatrechts in Form einer Kapitalgesellschaft (Eintragung im Handelsregister). Die Gesellschaft schließt auf eigene Rechnung Verträge ab und ist im Geschäftsleben tätig. Ihr Handeln kann auf einen bestimmten Gesellschaftszweck beschränkt werden. Sie hat klare Ergebnisverantwortung und nutzt kaufmännische Buchführung. Die Bilanzierung erfolgt als selbständiges Steuersubjekt. Dies führt zu steuerlichem Aufwand (Umsatzsteuer, Körperschaftsteuer, Gewerbesteuer), bietet aber auch die Möglichkeit steuerlicher Begünstigungen (Investitionszulage, Vorsteuer).

Eine GmbH mit einem Daseinsvorsorge-Auftrag kann als gemeinnützige GmbH ausgestaltet werden, so dass der Gesellschaftszweck gemeinnützig sein muss und eine Ausschüttung von Gewinnen an Gesellschafter ausgeschlossen werden kann. GGmbHs sind durch Änderungen des Gesellschaftsvertrages durch den oder die Gesellschafter leicht umgestaltbar, auch in ihrer Zweckbestimmung. Soll eine längerfristige Verfestigung des Gesellschaftszweckes und eine Bindung an einen bestimmten übergeordneten Auftrag erreicht werden, kann die gGmbH ergänzt werden um eine

privatrechtliche Stiftung, die als Gesellschafter der gGmbH fungiert und durch den (schwer zu ändernden) Stiftungszweck eine längerfristige Stabilität bewirkt.

Organe einer gGmbH sind mindestens die Gesellschafterversammlung und die Geschäftsführung, ggf. noch ein Aufsichtsrat. Die Gesellschafterversammlung hat starken Durchgriff auf die Geschäftsführung. Die Stakeholder (Bundes- oder Landesbehörden) können Mitgesellschafter sein und in die Gesellschafterversammlung eingebunden werden oder – ohne die Funktion eines Teilhabers – in einen Aufsichtsrat eingebunden werden.

Beispiele für gGmbH im Bereich öffentlicher Daseinsvorsorge und Digitalisierung sind die Klinisches Krebsregister Brandenburg gGmbH, deren gesetzlicher Auftrag sich aus dem Krebsregistergesetz (KRG) ergibt, die Gematik GmbH (Auftrag aus § 291b SGB V), die Bundeswehr Informationstechnik GmbH (BWI), die keinen gesetzlichen Auftrag hat, jedoch den Bund als 100%-Gesellschafter, oder die juris GmbH mit einem Gesellschafteranteil des Bundes von 51% und privaten Teilhabern.

## ii. Vor- und Nachteile

### (g) GmbH

Vorteile	Nachteile
Möglichkeit der einfacheren und schnelleren Gestaltung von Entscheidungswegen aufgrund weniger Vorgaben für die Ausgestaltung der Hierarchien (im Rahmen der Kompetenzen der Geschäftsführung)	(ohne Beleihung) keine Möglichkeit, öffentlich-rechtlich zu handeln, insbesondere durch Verwaltungsakte
Größere Flexibilität bei der Rekrutierung und Reduzierung von Personal, da sie nicht dem öffentlichen Dienstrecht unterliegt (Tariffreiheit)	Anschluss- und Benutzungszwang nur kompliziert auszugestalten
Größere Flexibilität des Handelns am Markt, z.B. bei Investitionen (Kreditaufnahme)	Einfluss des Hoheitsbereiches nur über Gesellschafterversammlung bzw. Aufsichtsrat
Direkte Entscheidungsstrukturen ohne Genehmigung durch staatliche Aufsicht wie bei einer Anstalt d.ö.R., d.h. GmbH unterliegt nicht direkt der staatlichen Aufsicht	Starker Einfluss des Gesellschafters der GmbH (= Anteilsinhaber der GmbH-Anteile) auf dessen Ausrichtung und Bestand
Gesellschaft unterliegt (bei mehreren Gesellschaftern) nicht dem Vergaberecht.	Geringere Reputation gegenüber der Anstalt d.ö.R.
Möglichkeit der wirtschaftlichen Betätigung (bei gGmbH nur im Rahmen des gemeinnützigen Zwecks)	Funktionalität wird von wirtschaftlichen Interessen geleitet. Gesellschafter sind Entscheidungsträger, das heißt, die Gesellschafter können jederzeit eine Änderung des Zweckes oder des Gesellschaftervertrages herbeiführen oder die gGmbH auflösen.
gGmbH ist berechtigt, Spenden entgegenzunehmen und dem Spender steuerwirksame Spendenbescheinigungen auszustellen und nicht verpflichtet, strategisch auf einen wirtschaftlichen Gewinn hin zu agieren.	

### c) Gegenüberstellung der Rechtsformen

Eine Gegenüberstellung der beiden Rechtsformen unter Berücksichtigung der Kriterien ergibt folgendes Bild:

Kriterium	AöR	(g)GmbH
Steuerbarkeit	++	+
Stakeholder-Beteiligung	+	+
Gründungsaufwand	o	++
Fachaufsichtsfreie Neutrale Stelle	++	-
Flexibilität (Personal, Orga, Technik)	-	++
Skalierbarkeit im Sinne der Wirtschaftlichkeit	o	++
Skalierbarkeit im Sinne der Menge der Datenanalyseprozesse	++	++

Im Ergebnis spricht für die AöR die bessere Abbildbarkeit der besonderen Aufsichts-, Kontroll- und Integritätsvoraussetzungen des NGD, vor allem die Unabhängigkeit der Neutralen Stelle, während für die (g)GmbH die höhere Flexibilität, Agilität und der geringere Gründungsaufwand spricht.

Alle diese Faktoren können jedoch durch die Detail-Ausgestaltung der Einrichtung (Gesetz, Satzung, Gesellschaftervertrag) weiter beeinflusst, negative Faktoren ggf. abgeschwächt werden. Daher erscheinen insgesamt beide Modelle grundsätzlich geeignet. Im Hinblick auf die Akzeptanz in der Öffentlichkeit („Umgang mit kritischen Gesundheitsdaten“) und die Sicherung einer interessenneutralen Datenauswertung ist voraussichtlich eine rein öffentlich-rechtliche Einrichtung politisch eher durchsetzbar, selbst wenn sie Flexibilitäts- und finanzielle Skalierungsnachteile hat.

### d) Rechtsbeziehung zwischen dem NGD und den Nutzern

Für die Nutzung des NGD ist zwingend notwendig, dass das NGD als „Dienstleister“ für die Analyse und Auswertung von Gesundheitsdaten ein Vertrag mit den Nutzern schließt. Da durch die Nutzer Gesundheitsdaten bereitgestellt werden müssen, werden diese ein starkes Interesse daran haben, die Leistungspflichten klar abzustecken sowie Verantwortlichkeiten und den Umgang mit den Daten zu definieren. Denkbar wäre ein Rahmenvertrag für die Nutzung des NGD aufzusetzen, der keine unmittelbaren Pflichten der Nutzer zur Bereitstellung Daten begründet, sondern auf den Abschluss von Einzelverträgen zur Durchführung einer Studie während der Vertragsdauer zur Nutzung des NGD gerichtet ist. Vertragsgegenstand des Rahmenvertrags wären die allgemeinen Bedingungen und Voraussetzungen künftig abzuschließender einzelner Verträge zur Durchführung einer Studie, ohne dass die Nutzer selbst schon rechtlichen Bestimmungen unterliegen. Dabei muss die Nutzung des NGD, insbesondere die Anforderungen an die Nutzung, und nicht die Auswertung selbst im Vordergrund stehen.

Für den Regelungsumfang des Rahmenvertrags zur Nutzung des NGD können beispielhaft die folgenden Punkte genannt werden:

- Vertragsgegenstand
- Abschluss von Einzelaufträgen
- Leistungen des NGD
- Technische Anforderungen
- Rechtliche Anforderungen
- Datenschutz- und Datensicherheit
- Kosten
- Gewährleistung
- Haftung

- Laufzeit und Kündigung

Der auf dem Rahmenvertrag basierende Einzelvertrag zwischen dem Datenbereiter und dem Datennutzer würde im Folgenden die Leistungspflichten der Vertragsparteien klar definieren.

Für den Regelungsumfang des Einzelvertrags über die gemeinsame Studiendurchführung zur temporären Datenzusammenführung und Datenanalyse unter Nutzung des NGD müsste dabei folgendes umfassen:

- Datensatz
- Pflichten des Datenbereitstellers
- Pflichten des Datennutzers
- Spezifische Zusatzleistungen
- Nutzungsrechte der Auswertung

#### 4. Rechtlich-organisatorische Ausgestaltung der Empfehlung

##### a) Gesetzliche Grundlage des NGD

Materiell-rechtliche Kernfrage der Arbeit des NGD ist die Verarbeitung der Gesundheitsdaten der Betroffenen. Angesichts der Sensitivität der Daten, der beabsichtigten großen Reichweite des NGD im deutschen Gesundheitswesen und der Regelungen der DSGVO zu Gesundheitsdaten (einschließlich der Öffnungsklauseln) wird eine spezial-gesetzliche Rechtsgrundlage für den NGD für erforderlich gehalten. Angesichts der bereits bestehenden umfangreichen bereichsspezifischen Regelungen für die öffentlichen Akteure des Gesundheitswesens könnte eine solche Rechtsgrundlage zudem die Voraussetzungen schaffen, dass die betroffenen Einrichtungen sich am NGD beteiligen dürfen oder sogar müssen. Es liegt zudem nahe, die datenschutzrechtliche Ausgestaltung mit der organisationsrechtlichen Ausgestaltung gesetzgeberisch zu verbinden, um eine umfassende Regelung der Errichtung und Nutzung des NGD zu schaffen.

Die gesetzliche Grundlage für die Datenverarbeitung des „Netzwerk Gesundheitsdaten“ muss die wesentlichen Aspekte selbst regeln. Hierfür ist zu prüfen, inwiefern dies durch Öffnungsklauseln des Art. 9 Abs. 2 lit. j der DSGVO gedeckt wird.

Der Öffnungsklausel des Art. 9 Abs. 2 lit. j DSGVO zufolge dürfen Mitgliedstaaten im nationalen Recht die Verarbeitung von Gesundheitsdaten für privilegierte Verarbeitungszwecke erlauben. Zu den privilegierten Verarbeitungszwecken zählen, wie bereits erörtert, auch wissenschaftliche Forschungszwecke und statistische Zwecke. Die Öffnungsklausel setzt voraus, dass nationalen Vorschriften dabei den Verhältnismäßigkeitsgrundsatz und die Wesensgehaltsgarantie wahren. Hierbei handelt es sich um Vorgaben, die den grundrechtlichen Direktiven des Grundgesetzes und der Grundrechtecharta entsprechen. Relevanter ist hingegen die Anforderung, dass eine entsprechende nationale gesetzliche Regelung „angemessene und spezifische Maßnahmen“ vorsehen muss, um jene Rechte zu schützen. Dabei lässt Art. 9 Abs. 2 lit. j DSGVO offen, ob diese Maßnahmen über die allgemeinen Verpflichtungen der DSGVO hinausgehen müssen. Der Hinweis auf den spezifischen Charakter der Maßnahme und auch der Verweis auf Art. 89 DSGVO spricht dabei eher für eine eigenständige Absicherung. Somit wird man davon ausgehen können, dass die Mitgliedstaaten einen weiten Spielraum bezüglich der Festlegung solcher Maßnahmen haben. So könnte beispielsweise als eine solche Maßnahme ein besonderes Kontrollgremium im Bereich der medizinischen Forschung vorgesehen werden.

Denkbar wäre somit, auf Grundlage der Öffnungsklausel des Art. 9 Abs. 2 lit. h DSGVO, die Schaffung eines gesetzlichen Auftrags, welcher der Realisierung des NGD dient und möglicherweise sogar eine gesetzliche Verpflichtung zur Datenbereitstellung durch die Stakeholder darstellt. Als Beispiel für die Schaffung eines gesetzlichen Auftrags zur Auswertung von Gesundheitsdaten kann hier beispielhaft die Datentransparenzverordnung (DaTraV) und das „Informationssystem Versorgungsdaten“ (Datentransparenz) des DIMDI angeführt werden. Die Datentransparenzverordnung regelt das Nähere zur Wahrnehmung, Durchführung und Finanzierung der Aufgaben der Datentransparenz nach den §§ 303a bis 303e SGB V. Das DIMDI nimmt dabei die Aufgaben der Vertrauensstelle nach § 303c SGB V und die Aufgaben der Datenaufbereitungsstelle nach § 303d SGB

V wahr. Hierfür übermittelt das Bundesversicherungsamt (BVA) dem DIMDI die jährlich erhobenen und auf Vollständigkeit und Plausibilität geprüften Daten für die vorgesehenen Zwecke gemäß der Vorgaben des SGB V. Das BVA wiederum erhält die Daten von den Krankenkassen für den morbiditätsorientierten Risikostrukturausgleich (Morbi-RSA). Die DaTraV-Daten beinhalten u.a. Angaben über ambulante und stationäre Diagnosen sowie zur ambulanten Arzneimittelversorgung der gesetzlich Versicherten. Diese können insbesondere für Analysen des Versorgungsgeschehens im Rahmen der Versorgungsforschung und für Steuerungsaufgaben in der gesetzlichen Krankenversicherung genutzt werden. Das Informationssystem Versorgungsdaten ermöglicht es Nutzungsberechtigten, beim DIMDI die Auswertung von Versorgungsdaten der gesetzlichen Krankenkassen zu beantragen und durchführen zu lassen. Welche Institutionen Daten für welchen Zweck erhalten dürfen, regelt das Gesetz (§§ 303a bis 303e SGB V). Nur dort genannte Institutionen können zukünftig Versorgungsdaten über das DIMDI nutzen. Dazu gehören unter anderem bestimmte Einrichtungen der Krankenkassen, der Gemeinsame Bundesausschuss, Interessenvertretungen der Patienten und der Leistungserbringer auf Bundesebene sowie Institutionen der Forschung und Gesundheitsberichterstattung. Ihnen sollen die Daten z.B. helfen, die Qualität der Versorgung zu verbessern oder Leistungsressourcen zu planen. Finanziert wird das Informationssystem durch die gesetzlichen Krankenkassen. Für die Aufbereitung der Daten durch das DIMDI fallen Nutzungsgebühren an. Näheres regelt die Gebührenverordnung. Mit der Datentransparenzverordnung wurde für das DIMDI eine gesetzliche Grundlage geschaffen, das Informationssystem Versorgungsdaten (Datentransparenz) zu realisieren.

Ein vergleichbarer gesetzlicher Auftrag für die Arbeit des NGD ist – unabhängig von der gewählten Rechtsform – mit umfangreichen Gesetzesänderungen auf Bundes- und Landesebene verbunden. Im Folgenden wird skizziert, welche Anpassungen notwendig wären.

#### i. [Änderung im nationalen Recht](#)

##### § 27 Abs. 1 BDSG n.F.

§ 27 Abs. 1 BDSG n.F. konkretisiert u.a. die Ermächtigung des Art. 9 Abs. 2 lit. j DSGVO zur Verarbeitung von Gesundheitsdaten. Er gilt für die öffentliche und private Forschung durch öffentliche und nicht-öffentliche Stellen und schafft eine Regelung im nationalen Recht für die Verarbeitung von Gesundheitsdaten zu wissenschaftlichen Forschungszwecken und zu statistischen Zwecken. Dabei ist zu beachten, dass nationale Vorschriften im bereichsspezifischen Recht als *lex specialis* der allgemeinen Regelung des § 27 Abs. 1 BDSG n.F. vorgehen.

Da ein Großteil der Arbeit des NGD auf Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F. gestützt werden kann, diesem aber bereichsspezifisches Recht vorgeht, liegt der gesetzliche Anpassungsbedarf primär dort.

#### ii. [Spezialgesetzliche Vorschriften](#)

Art. 9 Abs. 4 DSGVO enthält eine weitere Öffnungsklausel, die es den Mitgliedsstaaten erlaubt, zusätzliche Bedingungen, einschließlich Beschränkungen, einführen oder aufrechterhalten, soweit die Verarbeitung von genetischen, biometrischen oder Gesundheitsdaten betroffen ist. Somit sind grundsätzlich spezialgesetzliche Vorschriften im nationalen Recht weiterhin anwendbar und den allgemeineren Regelungen aus DSGVO und BDSG n.F. vorzuziehen. Die verschiedenen spezialgesetzlichen Vorschriften regeln den Umgang mit Gesundheitsdaten, insbesondere im Hinblick auf die Nutzung zu wissenschaftlichen Forschungszwecken und statistischen Zwecken. Diese Vorschriften sehen oftmals strenge Regelungen vor, sodass dies die Rechtsgrundlage des Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F. stark einschränken. Spezialgesetzliche Vorschriften finden sich u.a. in den Sozialgesetzbüchern (SGB I, SGB V und SGB X), in den krankenhausspezifischen Landesgesetzen sowie den Spezialgesetzen des RKI, DIMDI und der Krebsregister.

## iii. Fallbeispiele

Anhand der Beispiele (siehe Kapitel II A) sollen in groben Zügen die Zulässigkeitsvoraussetzungen sowie der gesetzgeberische Anpassungsbedarf aufgezeigt werden.

Beispiel	1a	1b	2
Datenbereinsteller:	<ul style="list-style-type: none"> <li>• Deutsches Institut für Medizinische Dokumentation und Information (DIMDI)</li> <li>• Versicherungen</li> <li>• Klinische Krebsregister</li> <li>• Krankenhäuser</li> </ul>	<ul style="list-style-type: none"> <li>• Robert Koch-Institut (RKI)</li> <li>• Versicherungen</li> <li>• Klinische Krebsregister</li> </ul>	<ul style="list-style-type: none"> <li>• Robert Koch-Institut (RKI)</li> <li>• Krankenhäuser</li> <li>• Statistisches Bundesamt (Destatis)</li> <li>• Deutscher Wetterdienst (DWD)</li> </ul>
Rechtsgrundlage (lex generalis):	Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F.	Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F.	Art. 9 Abs. 2 lit. j DSGVO i.V.m. § 27 Abs. 1 BDSG n.F.
Spezialgesetzliche Vorschriften (lex specialis):	<ul style="list-style-type: none"> <li>• Datentransparenzverordnung (DaTraV)</li> <li>• Sozialgesetzbuch (SGB) V, VII und X</li> <li>• Krebsregistergesetzes (KRG)</li> <li>• Krankenhausspezifische Landesgesetze</li> </ul>	<ul style="list-style-type: none"> <li>• Infektionsschutzgesetz (IfSG)</li> <li>• Sozialgesetzbuch (SGB) V, VII und X</li> <li>• Krebsregistergesetzes (KRG)</li> </ul>	<ul style="list-style-type: none"> <li>• Infektionsschutzgesetz (IfSG)</li> <li>• Krankenhausspezifische Landesgesetze</li> <li>• Bundesstatistikgesetz (BStatG)</li> <li>• DWD-Gesetz</li> </ul>
(Möglicher) Gesetzgeberischer Änderungsbedarf:	<p>§ 5 DaTraV</p> <p>Sozialdaten:</p> <ul style="list-style-type: none"> <li>• § 287, 303e SGB V</li> <li>• § 206 SGB VII</li> <li>• § 75 SGB X</li> <li>• § 98 SGB XI</li> </ul> <p>Krankenhausspezifische Landesgesetze:</p> <ul style="list-style-type: none"> <li>• Baden-Württemberg: § 46 Abs. 1 Nr. 2a LKHG</li> <li>• Bayern: Art. 27 Abs. 4 BayKrG</li> <li>• Berlin: § 25 Abs. 1 S. 2 Nr.1 bis 4 LKG</li> <li>• Brandenburg: § 31 BbgKHE</li> <li>• Bremen: § 7 Abs. 1 und 2 BremKHDS</li> <li>• Hamburg: § 12 HmbKHG</li> <li>• Hessen: § 12 HKHG</li> <li>• Mecklenburg-Vorpommern: § 38 Abs. 1 und 2 LKHG M-V</li> <li>• Niedersachsen: keine Regelung</li> <li>• Nordrhein-Westfalen: § 6 Abs. 2 GDSG NRW</li> <li>• Rheinland-Pfalz: § 37 Abs. 1 und 2 LKG RP</li> </ul>	<p>§ 4 IfSG</p> <p>Sozialdaten:</p> <ul style="list-style-type: none"> <li>• § 287, 303e SGB V</li> <li>• § 206 SGB VII</li> <li>• § 75 SGB X</li> <li>• § 98 SGB XI</li> </ul>	<p>§ 4 IfSG</p> <p>Krankenhausspezifische Landesgesetze:</p> <ul style="list-style-type: none"> <li>• Baden-Württemberg: § 46 Abs. 1 Nr. 2a LKHG</li> <li>• Bayern: Art. 27 Abs. 4 BayKrG</li> <li>• Berlin: § 25 Abs. 1 S. 2 Nr.1 bis 4 LKG</li> <li>• Brandenburg: § 31 BbgKHE</li> <li>• Bremen: § 7 Abs. 1 und 2 BremKHDS</li> <li>• Hamburg: § 12 HmbKHG</li> <li>• Hessen: § 12 HKHG</li> <li>• Mecklenburg-Vorpommern: § 38 Abs. 1 und 2 LKHG M-V</li> <li>• Niedersachsen: keine Regelung</li> <li>• Nordrhein-Westfalen: § 6 Abs. 2 GDSG NRW</li> <li>• Rheinland-Pfalz: § 37 Abs. 1 und 2 LKG RP</li> <li>• Saarland: § 14 Abs. 1 SKHG</li> <li>• Sachsen: § 34 SächsKHG</li> <li>• Sachsen-Anhalt: keine Regelung (14e KHG LSA)</li> </ul>

	<ul style="list-style-type: none"> <li>• Saarland: § 14 Abs. 1 SKHG</li> <li>• Sachsen: § 34 SächsKHG</li> <li>• Sachsen-Anhalt: keine Regelung (14e KHG LSA)</li> <li>• Schleswig-Holstein: keine Regelung</li> <li>• Thüringen: § 27a Abs. 1 und 2 ThürKHG</li> </ul>		<ul style="list-style-type: none"> <li>• Schleswig-Holstein: keine Regelung</li> <li>• Thüringen: § 27a Abs. 1 und 2 ThürKHG</li> </ul> <p>§ 21 BStatG</p> <p>§ 4 DWD-Gesetz</p>
--	---	--	--

#### b) *Regelungen zur Organisation des NGD*

In einem Gesetz zur Errichtung und Nutzung des NGD müsste im Falle der Auswahl der AöR als Rechtsform auch die Errichtung der AöR, ihr Anstaltszweck und die Governance der AöR geregelt werden, insbesondere die Zusammensetzung des Verwaltungsrats und die Einbindung von Stakeholdern. Im AöR-Gesetz wäre auch festzulegen, wer Gewährträger der Anstalt ist (z.B. BMG) und wie die Finanzausstattung geregelt wird (Wirtschaftsplan, Gebührenerhebung etc.). Auch für den Fall der Auswahl der Rechtsform einer GmbH kann die Gründung durch eine gesetzliche Regelung begleitet werden, wie dies beispielsweise bei der Gründung der Infrastrukturgesellschaft für Autobahnen und andere Bundesstraßen durch das InfrGG erfolgt ist. Davon ist aber grundsätzlich abzuraten, weil beim NGD nicht die Überführung bestehender, weitgehend ausgestalteter und mit Personal und Ressourcen unterlegter öffentlicher Aufgaben in private Rechtsform geplant ist, sondern die organisatorische Verankerung einer neuen, öffentlich gesteuerten Aufgabe. Hierzu reicht ein entsprechender Gesellschaftsvertrag aus.

Eine Ermittlung des Ressourcenbedarfs der Organisation ist derzeit noch nicht möglich, weil sie von Rechtsformwahl und Geschäftsmodell abhängt. Unter rein funktionaler Betrachtung kann jedoch abgeschätzt werden, dass eine Minimalausstattung einer selbständigen Organisationseinheit NGD etwa 19 VZÄ bedarf (Kontaktplattform 3, Contracting 3, Neutrale Stelle 5, Support und Qualitätsentwicklung 2, Steuerung/Führung 3, Assistenz/Z-Aufgaben 3).

### 5. *Ausgestaltung des IT-Sicherheits- und Datenschutzmanagements*

#### a) *Datenschutzmanagement*

Das Datenschutzmanagement legt diejenigen Prozesse fest, die notwendig sind, um die Umsetzung der gesetzlichen Anforderungen des Datenschutzes bei der Planung, Einrichtung, dem Betrieb und nach Außerbetriebnahme von Verfahren zur Datenverarbeitung sicher zu stellen (vgl. BSI-Grundschutz-Katalog, M 2.501 Datenschutzmanagement). Das Datenschutzmanagement stellt somit die übergeordnete Umsetzung der gesetzlichen datenschutzrechtlichen Pflichten sowie den Schutz der Betroffenen im Netzwerk Gesundheitsdaten sicher. Unter rechtspolitischen Gesichtspunkten sowie unter Berücksichtigung des Transparenzgrundsatzes (vgl. Art. 5 Abs. 1 lit. a DSGVO) muss für die Erstellung eines Datenschutzmanagements eine transparente Organisation und Arbeitsweise maßgebend sein. Dies gilt vor allem für den Umgang mit Gesundheitsdaten. Das Datenschutzmanagement muss somit Transparenz hinsichtlich aller wesentlichen Aspekte der Datenverarbeitung innerhalb des NGD und insbesondere der Neutralen Stelle schaffen.

Hierzu zählen die Organisation, die Verfahrensabläufe, die Rechtsgrundlagen, die Zweckbindung, die technischen und organisatorischen Maßnahmen und deren Dokumentation sowie das IT-Sicherheits-Management.

Im Rahmen des Datenschutzmanagements wäre es denkbar, auf den Musterprozess des BSI-Grundschutz-Katalogs zurückzugreifen. Der Prozess orientiert sich an den BSI -Standards 100-1 und 100-2 und ist als integrativer Bestandteil des Sicherheitsprozesses nach IT-Grundschutz anzusehen, kann aber auch als eigenständiger Prozess behandelt werden, wenn vorrangig der Datenschutzaspekt

behandelt werden soll. Der Prozess ist für die gesamte Organisation und alle Verfahren innerhalb der Organisation, in denen personenbezogene Daten verarbeitet werden, ausgelegt.<sup>40</sup>

Der Prozess besteht aus drei Phasen:

i. **Gap-Analyse (Soll-Ist-Abgleich)**

Für die Gap-Analyse ist es erforderlich festzulegen und zu dokumentieren, welche Anforderungen des Datenschutzes bei der Verarbeitung personenbezogener Daten eingehalten werden müssen und wie diese Anforderungen umgesetzt werden. Vorrangig ist hierbei die Datenschutz-Grundverordnung zu beachten. Für die Gap-Analyse ist somit ein Datenschutzkonzept erforderlich, welches sich an den Aspekten des BSI-Grundschutzkatalog orientieren sollte.<sup>41</sup> Ziel des Datenschutzkonzeptes ist es, in einer zusammenfassenden Dokumentation die datenschutzrechtlichen Anforderungen darzustellen, um als Grundlage für datenschutzrechtliche Prüfungen verwendet werden zu können.

ii. **Umsetzung fehlender/erforderlicher Maßnahmen**

Im zweiten Prozessschritt sollen die bislang noch nicht umgesetzten, durch das Datenschutzkonzept definierten Anforderungen, umgesetzt werden. Die Umsetzung erfolgt im Rahmen eines klassischen Projektmanagements mit einem Projekt- und Arbeitsplan. Hierfür müssen die entsprechenden Strukturen und Kapazitäten im NGD vorhanden sein.

iii. **Aufrechterhaltung des anvisierten Datenschutzniveaus**

Im dritten Prozessschritt sieht vor, dass das anvisierte Datenschutzniveau regelmäßig kontrolliert wird. Dabei muss, in Bezug auf die Vorgänge in denen personenbezogener Daten verarbeitet werden, auf Änderungen und Störungen im laufenden Betrieb der Verfahren des NGD reagiert werden. Zu den Änderungs- und Störungsquellen gehören insbesondere:

- Änderungen im europäischen Datenschutzrecht
- Änderungen in den (IT-)Verfahren
- Störungen in den operativen Betriebsabläufen, die als Sicherheitsvorfall zu klassifizieren sind
- Technischer Fortschritt und reduzierter Aufwand für bisher nicht realisierte Maßnahmen.

Im Zusammenhang mit der Aufrechterhaltung des anvisierten Datenschutzniveaus sind umfassende Sub-Prozesse notwendig, um die Änderungen und Störungen aus Datenschutzsicht eigenständig bearbeiten bzw. lösen zu können sowie zur Anpassung und Aktualisierung des Datenschutzkonzeptes. Hierfür müssen ebenfalls die entsprechenden Strukturen und Kapazitäten im NGD vorhanden sein.

b) **IT-Sicherheitsmanagement**

Das NGD wird als eine kritische IT-Dienstleistung im Sinne des IT-Sicherheitsgesetzes betrachtet. Vorgeschlagen wird daher die Änderung der BSI-KRITIS-VO und die Ergänzung einer kritischen Dienstleistung „Sektorenübergreifende Auswertungsplattform“ im Sektor Gesundheitswesen., Aus dieser Festlegung ergibt sich, dass NGD über (Mindest-)Anforderungen der Art. 32, 89 DSGVO hinaus auch die Anforderungen nach § 8a Abs. 1 BSIG zu erfüllen hat und IT-Sicherheitsmaßnahmen nach dem Stand der Technik ergreifen muss. Über die im Datenschutzrecht bereits abgebildeten Schutzziele der Vertraulichkeit und Integrität der Daten hinaus wird hierdurch auch das Schutzziel der Verfügbarkeit des Systems adressiert. Orientierungshilfe für die Erfüllung dieser Anforderungen können die Vorgaben von Betreibern Kritischer Infrastrukturen (bspw. Referenzumgebungen wie Unikliniken) oder die branchenspezifischen Sicherheitsstandards sein, welche durch deren Fachverbände erarbeitet werden. Gemäß § 8a Abs. 3 BSI-G müsste sich NGD alle zwei Jahre einem Audit unterziehen und das Ergebnis dem BSI nachweisen. Vorzusehen sind darüber hinaus Prozesse zum Umgang mit IT-Sicherheitsvorfällen (Incident Response), die auch die für kritische Infrastrukturen vorgeschriebene Erfüllung der Meldepflicht umfassen.

<sup>40</sup> vgl. BSI-Grundschutz-Katalog, M 2.501 Datenschutzmanagement.

<sup>41</sup> vgl. BSI-Grundschutz-Katalog, M 2.503 Aspekte eines Datenschutzkonzeptes.

## 6. Ausgestaltung der Stakeholder-, Gewährleistungs- und Aufsichtsfunktionen

Unabhängig von der gewählten Rechtsform sind Steuerungsprozesse für das NGD zu planen und zu etablieren. Hierbei sind folgende Arten von Prozessen zu unterscheiden:

### a) *Stakeholder-Beteiligung*

Für eine langfristige Akzeptanz, Bedarfsdeckung und Qualität des NGD ist die geeignete Einbindung der Stakeholder erforderlich. Damit sollten zwei Bedürfnisse abgebildet werden:

- Stakeholder sollen bei der Umsetzung von fachlichen, technischen und juristischen Maßnahmen unterstützt werden, um das NGD möglichst niederschwellig und reibungsfrei nutzen und damit umfangreich profitieren zu können. Hierzu wären ein strukturierter Austausch und Rückkopplungsprozesse zu etablieren.
- Fach- und auswertungsspezifische Rückmeldungen sollten unmittelbarer und direkter zwischen den Fachleuten der Datenbereiter und Datennutzer und den Experten des NGD ausgetauscht werden. Hierzu wären Facharbeitskreise einzurichten.

### b) *Gewährleistungsverantwortung*

Das NGD sollte dauerhaft unter einer öffentlich-rechtlichen Gewährleistungsverantwortung stehen, die garantiert, dass die zentralen Ziele und Anforderungen, insbesondere des Datenschutzes und der fachlichen Vertrauenswürdigkeit sichergestellt sind. Dies könnte in der Rechtsform einer AöR durch eine entsprechende Besetzung des Verwaltungsrates erreicht werden, bei der gGmbH durch die Beteiligungsstruktur und den Gesellschaftsvertrag. Neben dem Bund wären die Länder wegen ihrer Zuständigkeiten im Gesundheitswesen am Verwaltungsrat bzw. der Gesellschaftsteuerung zu beteiligen. Nicht erforderlich erscheint eine zusätzliche gesetzliche Aufsichtsmöglichkeit (wie in § 291b SGB V für die Gematik), da das NGD vom Bund (und Ländern) als Anstaltsträger oder Gesellschafter direkt gesteuert und beaufsichtigt werden muss.

### c) *Datenschutz- und IT-Sicherheitsaufsicht*

Die Aufsicht über den Datenschutz beim NGD sowie IT-Sicherheit ergibt sich unmittelbar aus den einschlägigen gesetzlichen Regelungen und den obigen Ausführungen zum Datenschutz- und IT-Sicherheitsmanagement. Zuständige Datenschutzaufsichtsbehörde ist im Falle der Ausgestaltung als Anstalt des Bundes die BfDI, in allen anderen Fällen die Datenschutzaufsicht des Sitzlandes.

## IV. Weiteres Vorgehen

Mit der vorliegenden Machbarkeitsstudie ist grundsätzlich dargelegt, dass der Ansatz eines virtuellen Netzwerks Gesundheitsdaten fachlich, rechtlich und technisch machbar ist. Er eröffnet die Option, einen im Anwendungsbereich nahezu unbeschränkten Rahmen für die gesundheitspolitisch erwünschte Ausweitung von Big Data-Analysen in der Gesundheitsforschung zu schaffen.

Eine Umsetzung des Modells kann nur schrittweise erfolgen und erfordert eine Härtung des Konzeptes, insbesondere durch Pilotierung(en), eine tiefere und breitere Debatte mit Stakeholdern, die Erarbeitung des Geschäftsmodells, die Herbeiführung einer politischen Entscheidung einschließlich Gesetzgebung sowie die anschließende praktische Errichtung des Netzwerks.

Im Einzelnen werden die nachfolgenden nächsten Schritte vorgeschlagen.



Abbildung 13: Vorgehensvorschlag

### 1. Stakeholder-Dialog

Von entscheidender Bedeutung für den Erfolg des NGD ist die Einbindung der Stakeholder in die Planung des NGD. Der Stakeholder-Dialog sollte an die bisherige Stakeholder-Diskussion anknüpfen, frühzeitig beginnen und während der gesamten Planung und Errichtung fortgesetzt werden. Es bietet sich an, hierfür eine dauerhafte Organisationsform zu wählen. Vorgeschlagen wird die Einrichtung eines Fachforums NGD, das zunächst mit einer Arbeitsgruppe beginnen, später mit einer Steuerungsgruppe und Facharbeitsgruppen aufwachsen kann.

Mitglieder des Fachforums sollten Vertreter derjenigen Einrichtungen sein, die potentielle Datenbereitsteller oder Datennutzer im NGD werden können. Hierbei sollte eine Mischung aus institutionellen Interessenträgern (Verbänden) und profilierten Einzelpersonen angestrebt werden. Wegen der Notwendigkeit der Etablierung und Gesetzgebung sollten Vertreter der Politik des Bundes und der Länder in das Fachforum eingeladen werden. Das Fachforum sollte unter der Ägide des BMG stattfinden. Im weiteren Verlauf wäre die Geschäftsstelle des Fachforums zunächst beim BMG und nach der Gründung des NGD dortselbst anzusiedeln.

Neben dem vor allem der institutionsübergreifenden Weiterentwicklung und gesellschaftlichen Verankerung des NGD dienenden Fachforum muss wegen der speziellen Betroffenheiten spätestens im Vorfeld von Gesetzgebungsverfahren und Gründung eine Beteiligung von drei speziellen Gruppen erfolgen:

- Die Datenschutzaufsichtsbehörden des Bundes und der Länder, gemeinsam organisiert in der Datenschutzkonferenz, sind frühzeitig anzusprechen, um das datenschutzrechtliche Konzept

zu diskutieren. Hier wird empfohlen, im ersten Schritt einzelne Landesbeauftragte und ergänzend die BfDI anzusprechen. Die Auswahl der Landesbeauftragten könnte sich an den Zuständigkeiten im Kontext der Pilotierung orientieren. Die BfDI ist wegen der Trägerschaft des Projekts im BMG ebenfalls einzubinden, auch sofern BfDI möglicherweise nicht direkt zuständig ist.

- Gerade im Hinblick auf die Beteiligung von Kliniken am NGD sollte die Gesundheitsministerkonferenz (GMK) frühzeitig informiert und eingebunden werden. Die Einbindung könnte mit der Unterrichtung über die Ergebnisse der Studie und Information über eine geplante Pilotierung beginnen. Gerade wenn eine Erleichterung der Errichtung des NGD durch gesetzgeberische Maßnahmen vorgesehen ist, ist eine frühzeitige harmonisierte Vorbereitung mit den Ländern anzuraten.
- Die Einrichtungen der Selbstverwaltung im Gesundheitswesen haben ein vielfältiges Interesse an der Art und Weise der Etablierung des NGD. Sie sind daher ebenfalls frühzeitig einzubinden. Dabei sollte allerdings vermieden werden, das Projekt zu einem Projekt der Selbstverwaltung zu machen, um die im Zuge der Pilotierung und späteren Gesetzgebung nötigen Design-Entscheidungen in rechtlicher, organisatorischer und technischer Hinsicht nicht mit zu viel Abstimmungsnotwendigkeiten zu belasten.

## 2. Pilotierung

Zentral für eine weitere Härtung des Konzeptes ist die Durchführung einer geeigneten Pilotierung. Dabei sollten mindestens zwei bis drei fachlich möglichst weit auseinander liegende Piloten, entlang der beschriebene Beispielfällen I und II (s. Kapitel II A) konzipiert und durchgeführt werden. Neben der Entwicklung des fachlichen Konzeptes und auch der Etablierung von Qualitäts- und Rückkopplungskonzept sollten frühzeitig die für die Pilotierung notwendigen Stakeholder für die Projektformulierung des Piloten hinzugezogen werden.

Im weiteren Schritt wäre zur Vorbereitung der Pilotierung die Entwicklung eines Prototyps der technischen Plattform des NGD erforderlich. Der Prototyp sollte folgende Funktionalitäten anbieten:

- Big-Data-Plattform
- Kontaktplattform
- Terminologieserver

Auf die technische Pilotierung des Funktionsbereichs Contracting kann zunächst verzichtet werden, weil die Funktionalität während der Pilotierung nicht automatisiert benötigt und die technische Komplexität der Realisierung des Contracting nicht sehr hoch ist.

Zur Pilotierung sollte zum einen die für die Bereitstellung des Systems notwendige Hardware erworben werden, zum anderen anschließend eine auf Hadoop basierte Big Data-Plattform bestehend aus den einzelnen Software Bausteinen mit ihren dazugehörigen Lizenzen gekauft werden. Basierend auf den benannten Anforderungen sollte das NGD die Entwicklung der Kontaktplattform in Auftrag geben oder selbst vornehmen. Die dabei zum Einsatz kommende Software muss dem hier vorgeschlagenen Sicherheitskonzept entsprechen, so dass diese von einem unabhängigen externen Auditor ihre Konformität nach ISO 27001 auf Basis von IT-Grundschutz des Bundesamts für Sicherheit in der Informationstechnik (BSI) verifiziert wird. Ferner muss dabei auch das Personal entsprechend der Sicherheitsrichtlinien geschult werden.

Das NGD sollte eine Lizenz für SNOMED CT erwerben und dieses gemeinsam mit den lizenzkostenfrei nutzbaren Systemen ICD-10 GM, OPS und LOINC als Kern eines Referenz-Codesystems nutzen. Die genannten Systeme müssen hierzu nicht untereinander vernetzt werden, da SNOMED CT nur als Referenz für in ICD-10 GM, OPS und LOINC nicht vorhandene Konzepte genutzt werden sollte.

Weitere domänenspezifische Systeme können als Referenzsysteme hinzugefügt werden (z.B. ICSN für Gen-Daten oder ICNP für Daten aus der Pflege).

An das NGD angebundene Einrichtungen können anonymisierte, kodierte Datensätze zusammen mit Beschreibungen der genutzten Codesysteme an das NGD senden. Über NLP-, ML- und KI

Verfahren werden hieraus Vorschläge für ein Mapping auf das Referenzsystem berechnet, die anschließend durch Experten validiert werden. Lassen sich einzelne, in den Quellsystemen genutzte Konzepte nicht auf das Referenzsystem abbilden, werden diese selber zum Teil des Referenzsystems.

Nach Fertigstellung des Prototyps der Kontakt-Plattform kann mit dem Probelauf der Datenanalyse über das NGD begonnen werden. Die Pilotierung könnte zunächst im Rahmen eines Forschungsauftrags bei einer Forschungseinrichtung stattfinden, die entsprechend qualifiziert ist. Entlang der Pilotierung könnte eine Fortschreibung der vorliegenden Machbarkeitsstudie hin zu einem Grobkonzept vorgenommen werden.

### 3. Feinkonzept

Sofern die Pilotierung erfolgreich verläuft, sollte sich eine Phase der Feinkonzeptionierung anschließen. In ihr wären vier für den Aufbau des NGD kritische Leistungsbereiche näher zu spezifizieren:

#### a) *Geschäftsmodell*

In einem Geschäftsmodell sind die in dieser Machbarkeitsstudie nicht betrachteten Kosten der NGD zu kalkulieren und ein, ggf. verschiedene, Modell(e) der Finanzierung zu betrachten. Dabei muss ins Betracht gezogen werden, dass

- NGD wegen des öffentlichen Auftrags keine Refinanzierung durch wettbewerblich gebildete Preise zulässt,
- die Aufwände der Datenbereitsteller kompensiert werden müssen sowie
- die sehr unterschiedlichen (monetären oder nichtmonetären) Vorteile der Datennutzer aus den Auswertungen im NGD Berücksichtigung finden müssen.

#### b) *Fachkonzept*

Basierend auf den Darlegungen der Machbarkeitsstudie wird das fachliche Konzept weiter ausgearbeitet und um das Qualitätskonzept sowie das Konzept zur Rückkopplung von Stakeholdern und NGD ergänzt.

#### c) *Rechtsformentscheidung*

Basierend auf einer politischen Grundentscheidung, den Diskussionen mit den Stakeholdern und den Finanzierungsüberlegungen im Geschäftsmodell ist anschließend die Rechtsform des NGD festzulegen. Die aus der Rechtsformentscheidung folgenden organisatorischen und rechtlichen Anforderungen an den Aufbau des NGD sind zu ermitteln, zu priorisieren und zu beplanen.

#### d) *Datenschutzrechtliches Gutachten*

Wegen der besonderen Kritikalität der datenschutzrechtlichen Fragestellungen wird empfohlen, nach der Rechtsformentscheidung ein vertieftes datenschutzrechtliches Gutachten zu beauftragen, das die datenschutzrechtliche Zulässigkeit für eine größere Zahl von Konstellationen aus Datenbereitstellern und Datennutzern prüft und den Gesetzgebungs- oder sonstigen datenschutzrechtlichen Umsetzungsbedarf für die neue Entität beschreibt. Das Gutachten sollte nach Möglichkeit mit einer Datenschutzaufsichtsbehörde besprochen und deren Sichtweise aufgenommen werden.

#### e) *Organisationskonzept*

Für die Zielorganisation NGD ist ein Organisationskonzept NGD zu erarbeiten, das folgende Aspekte beschreibt:

- Personalbemessung für die einzelnen Funktionsbereiche und Hierarchieebenen
- Aufbau- und Ablaufplanung
- Budgetplanung einschließlich Vorbereitung eines ersten Wirtschaftsplans zwecks haushalterischer Verankerung
- Stufenplan zum Aufbau der Organisation

Die gesamte Feinkonzeptionierung sollte durch Facharbeitsgruppen des Fachforums NGD begleitet werden.

#### 4. Aufbau

An die Feinkonzeptionierung kann sich die Erarbeitung eines Gesetzentwurfes der Bundesregierung und ein entsprechendes Gesetzgebungsverfahren anschließen. Unter Umständen sollten auch Musterentwürfe zur Änderung von Landesrecht (z.B. Landeskrankenhausgesetze) erstellt werden. Mit der Verabschiedung des Entwurfs durch die Bundesregierung sollte eine Veranschlagung der für die Finanzierung des NGD nötigen Anschubmittel, ggf. je nach Geschäftsmodell auch nötigen dauerhaften Mittel, im Bundeshaushalt erfolgen.

An die Verabschiedung des Gesetzes sollte sich die Errichtung eines Aufbustabes NGD im BMG anschließen, der die weiteren Schritte, insbesondere die Gründung der Organisation, die Besetzung von Schlüsselfunktionen und die Etablierung der Steuerungsprozesse anleitet.