

**Eileen Fuchs:**

Guten Morgen. Herzlich Willkommen an alle, die an den Bildschirmen dabei sind, aber auch hier im Raum zum Marktdialog fürs Dateninstitut für Deutschland. Wir sind vereint hier mit einem ressortübergreifenden Team BMI und BMWK. Einige von Ihnen kennen wir schon aus verschiedenen anderen Zusammenhängen, ich stelle aber trotzdem noch mal vor, wer hier heute sitzt. Mein Name ist Eileen Fuchs. Ich bin Referatsleiterin des Referats für Datenpolitik, Datenstrategie, Open Data in der Bundesverwaltung und auch mit dem Dateninstitut betraut und neben mir sitzt Christina Schmidt-Holtmann, Referatsleiterin in BMWK, unter anderem auch zuständig für das Dateninstitut, aber auch für die SPRIND und weitere innovative Datenprojekte in BMWK. Sie kann gleich selbst noch ein paar Worte sagen. Mit hier bei uns aus meinem Team ist Lena Müller, mit ihr hatten Sie vielleicht auch schon Kontakt. Und von Seiten BMWK ist dabei Nabi Siefken, Katrin Rosendahl und dazu noch, das freut uns sehr, unsere beiden Fellows von Work4Germany, David Reinhard und Jane Schek, die uns beim Projekt Dateninstitut unterstützen. Christina, möchtest du noch ein, zwei Worte sagen?

**Christina Schmidt-Holtmann**

Auch von unserer Seite, BMWK-seitig einen herzlichen Guten Morgen und gleichzeitig, das können wir vielleicht auch noch kurz sagen, wir sind ja heute absolut ressortübergreifend unterwegs und begrüßen auch unsere Kolleginnen und Kollegen, die wir in Teilen schon zugeschaltet sehen aus BMBF und BMG. Wir haben mit dem Thema Long-COVID eines, das wir heute gerne mit Ihnen im Kontext des Dateninstitutes besprechen wollen, das eine ganze Vielzahl von Ressorts innerhalb der Bundesregierung umtreibt.

**Eileen Fuchs**

Genau, vielen Dank. Das geht schon über worum geht es heute für die, die jetzt vielleicht nicht die ganzen Emails, die wir verschickt haben, im Hinterkopf haben. Es geht in der Tat um einen Use Cases Anwendungsfall für den Aufbau des zu gründenden Dateninstituts, mit dem wir betraut sind. Diejenigen, die sich schon vorbereitet haben und den Bericht der Gründungskommission für das Dateninstitut kennen, wissen auch, dass dort ein Anwendungsfall aus dem Bereich Post-/Long-COVID vorgeschlagen wurde. Wir nennen es jetzt hier Post-COVID, weil sich in der Diskussion verstärkt herauskristallisiert hat, dass das der treffendere Begriff ist. Und wir wissen auch, dass gerade, es ging durch die Presse, viel dazu diskutiert wird. Es gibt jetzt gerade viele Initiativen, die starten, und wir werden diesen Marktdialog wirklich auch dialogisch mit Ihnen nutzen, um zu ermitteln, wie der Raum für einen solchen Use Case ist, ob sich das eignet für die Gründung des Dateninstituts als exemplarisch für den weiteren Aufbau desselben. Im Vorfeld hatte Frau Müller Ihnen auch schon viele Infos zukommen lassen, ich fasse nochmal den Ablauf der heutigen Veranstaltung zusammen. Also wir werden heute erst mal, das macht Frau Schmidt Holtmann, den Sachstand zum Dateninstitut vorstellen, noch mal ganz kurz, damit Sie reinkommen. Dann werden wir in zwei Fragenkomplexen Sie um Beteiligung bitten, einmal Fragenkomplex 1 Datenlage, Datenquellen, Roadblocker und nicht erfasste Themengebiete, da wird es nach einer vorab festgelegten Rednerliste von Ihnen Beiträge geben. Wir werden insgesamt die ganze Veranstaltung wortprotokollieren und haben deswegen eine Aufzeichnung der WebEx. Also insofern wird das alles wettbewerbsgleich dokumentiert, da müssen Sie sich keine Sorgen machen. Die einzelnen Beiträge und das Ganze wird auf unseren Webseiten auch veröffentlicht, insofern für alle zu gleichen Bedingungen, sollten Sie sich später um die Umsetzung eines dieser Use Cases bewerben, ist für alle einsehbar was wer gesagt hat. Dann werden wir Fragenkomplex 2 zu interessanten Use Cases machen, das

haben wir im Vorfeld nicht zugeschickt. Da sind wir interessiert an Ihren Vorschlägen, was es aus Ihrer Sicht gäbe. Dann sind wir schon am Schluss der Veranstaltung angekommen, die gegen 12:30 Uhr spätestens enden wird. Das ist der kurze Ablauf, noch ein paar Worte zum House Keeping, zum Organisatorischen: Wie gesagt, es ist eine Aufzeichnung, insofern würde ich Sie bitten darauf zu achten, dass Sie, wenn Sie nicht sprechen gemutet sind. Wir haben den Chat im Blick, falls es Probleme technischer oder sonstiger Natur gibt, schreiben Sie in den Chat. Meine Bitte wäre, inhaltliche Fragen oder Anmerkungen eher in die Redebeiträge zu packen, auch aufgrund der Wettbewerbsgleichheit, und nicht im Chat umfangreich zu kommentieren oder Fragen zu stellen. Aber auch der Chat wird aufgezeichnet, insofern sind wir was die Transparenz geht auf der sicheren Seite. Ansonsten würde ich sagen, erteile ich das Wort an Christina Schmidt-Holtmann und sie kann einen kurzen Überblick zum Sachstand geben.

### **Christina Schmidt-Holtmann**

Genau das mache ich sehr gerne. Das Dateninstitut ist ja schon etwas länger in der Pipeline und geht jetzt in die konkrete Umsetzung worüber wir uns sehr freuen. Wir haben heute im Kreis einige Kolleginnen und Kollegen oder Interessierte dabei, die sich noch nicht vertieft damit beschäftigt haben. Deswegen ganz kurz, wo kommen wir her und vor allen Dingen, wo möchten wir gemeinsam mit Ihnen hin? Das Dateninstitut ist ein Projekt, das sich die Koalitionsparteien im Koalitionsvertrag gegeben haben, und zwar soll das Dateninstitut Datenverfügbarkeit und Datenstandardisierung vorantreiben, Datentreuhand-Modelle und Lizenzen etablieren. Das ist quasi der Ausgangspunkt. Da ich gerade die Frage im Chat gesehen hab: Ja, die Folien werden geteilt im Nachgang und auch veröffentlicht. Das Projekt läuft unter gemeinsamer Federführung von BMI und BMWK, aber auch in enger Kooperation mit allen Ressorts der Bundesregierung. Ziele, wie das Dateninstitut aufgebaut sein soll: Als schlagkräftiger nationaler Akteur, der eigenständig, markt- und anwenderbasiert ist und vor allem sehr stark gemeinwohlorientiert ausgerichtet ist. Genau wo kommen wir her, wie sind wir jetzt zu dem Stand gekommen, bei dem wir sind. Um insbesondere die Gemeinwohlorientierung und Bedürfnisorientierung umzusetzen wurde im Sommer letzten Jahres eine Gründungskommission beauftragt mit Mitgliedern aus den unterschiedlichen Stakeholdergruppen, Verwaltung, Wirtschaft, Wissenschaft und Zivilgesellschaft, die sich auf Basis von unterschiedlichen Ideen, die zu einem Dateninstitut schon kursierten und in Vorbereitung waren, intensiv Gedanken gemacht hat, unabhängige Gedanken, das möchte ich unterstreichen, wie ein Dateninstitut aufgebaut werden könnte. Das mündete in einer Empfehlung, die zum Digital Gipfel Ende letzten Jahres vorgestellt wurde und an die beiden Minister und Ministerinnen aus unseren Häusern, Minister Habeck und Ministerin Faeser, übergeben wurde und seitdem machen wir uns darüber Gedanken, wie wir das A finanziell aber B auch ganz praktisch umsetzen können. Die Gründungskommission hat einen starken bedarfs- oder bedürfnisorientierten und iterativen Ansatz zum Aufbau des Dateninstitutes empfohlen. Das heißt die Idee ist, dass man nicht auf einem weißen Papier anfängt und sich ein Organigramm ausdenkt, wie ein Dateninstitut aussehen könnte, sondern, dass man anhand von konkreten Pilotprojekten startet, die Aufgabenbeschreibung des Dateninstitutes zu definieren. Und aus diesen Pilotprojekten dann Learnings und Wegweisendes für den Aufbau des Dateninstitutes ableitet. Parallel zur Arbeit der Gründungskommission wurde eine umfassende Stakeholderkonsultation durchgeführt und eine Begleitstudie, die insbesondere auch die europäische und internationale Vergleichbarkeit im Fokus hatte, erstellt. Alle Dokumente finden Sie auf den Homepages der beiden Ministerien. Wie ging es dann weiter? Wir hatten den Auftrag, aus den Empfehlungen der Gründungskommission ein

konkretisierendes Konzeptpapier zu erarbeiten, dass ich Ihnen im Folgenden gerne vorstellen möchte und in dem der Prozess und auch die Ziele konkretisiert und skizziert wurden. Wichtig ist zu betonen, dass das Konzeptpapier innerhalb der Bundesregierung abgestimmt ist und dementsprechend auch die Verfahren und die Gemeinwohlorientierung innerhalb der Bundesregierung konsentiert ist. Danach kommen die Ziele und Ausrichtung des Dateninstitutes, einerseits als zentrale Anlaufstelle eben für die genannten Ziele im Koalitionsvertrag, um gesamtgesellschaftlich Daten besser verfügbar und nutzbar zu machen. Wir haben eine diverse Stakeholder Landschaft im Datenbereich und wollen, durchaus auch inspiriert von anderen europäischen Ländern, UK ist da in der Vorreiterrolle mit dem Open Data Institute, das dort aufgebaut wurde, evidenzbasiert und praxisnah den intersektoralen Austausch von Daten zu befördern. Wichtig ist, dass das Dateninstitut keine bestehenden Initiativen ersetzt, sondern einen Mehrwert schaffen soll, die vorhandenen Initiativen verknüpft und dabei in einen konstruktiven Austausch auch mit den Datenschutzbehörden tritt, ohne sie in irgendeiner Weise ersetzen oder spiegeln zu wollen. Hier sehen Sie im Endeffekt noch einmal, wie das Dateninstitut arbeiten soll und wo inhaltliche Schwerpunkte am Anfang gesetzt werden können. Was unseren Hausleitungen insgesamt sehr wichtig war, ist, dass es nicht das nächste Gremium wird, das abstrakte Beratung leistet, sondern es soll konkret ins Tun kommen, deswegen dieser Begriff des Do-Tank. Und deswegen dieser Use Case basierte Ansatz, den wir sowohl beim Aufbau verfolgen, der sich dann aber auch später in der konkreten Umsetzung oder Arbeit des Dateninstitutes wiederfinden soll. Wichtige Punkte Stakeholder Kommunikation, diese haben wir schon im Aufbauprozess sehr stark betrieben. Das wird auch ein wesentliches Kriterium bei der Arbeit des Dateninstituts sein, aber auch eine Art Mapping und Früh-Warn-Systemen auch in Richtung der Politik um zu zeigen: Wo gibt es Änderungsbedarfe? Wo sind Hürden beim Datenteilen, die adressiert werden müssen? Dass man da Expertise bündeln und sammeln kann, um Themen wie Datenstandardisierung, Datentaxonomie, Datentreuhand-Modelle, da kommt auch gerade eine ganze Menge an Regulierung auf uns zu, mit Data Governance Act und Data Act, befördern kann. Was als negative Abgrenzung oder Definition wichtig ist, da ist vielleicht der Begriff Dateninstitut etwas irreleitend, wir wollen keine wissenschaftliche Forschungseinrichtungen aufbauen. Man hätte das Dateninstitut vielleicht auch Data Agency nennen können. Uns ist wichtig, dass ein interdisziplinärer Ansatz gefahren wird. Auch wichtig zu wissen: Das Dateninstitut wird keine regulatorischen und hoheitlichen Aufgaben bekommen. Es ist als Unterstützung, aber auch als proaktiver Akteur geplant. Was eine spannende Frage in diesem Kontext ist und wo wir einen besonderen Fokus auf diesen Use Case legen werden, ist die Frage, ob das Dateninstitut gegebenenfalls selbst als Datentreuhänder oder Datenvermittlungsdienst auftreten soll. Ich hatte schon gesagt, da kommt eine Menge Regulatorik auf uns zu und gerade was Datentreuhand-Modelle angeht ist die Frage: Wer kann das eigentlich sinnvoll übernehmen? Das ist etwas, was wir uns aus den Pilot Use Cases erhoffen, da Erfahrungen beziehungsweise Ergebnisse abzuleiten.

Was stehen uns für Mittel zur Verfügung? Wir sind aktuell in den Verhandlungen zum Haushalt für 2024. Wir haben aufgrund des Konzeptpapiers jährlich 10.000.000 € zum Aufbau des Dateninstitutes als Anschubfinanzierung vom Haushaltsgesetzgeber gebilligt bekommen. Die Mittel sind formell im Haushaltsplan des BMWK verortet aber quasi mit Zugriffsberechtigung aus beiden Ressorts, um die Umsetzung voranzutreiben.

Was bedeutet bedarfsorientierter Ansatz, wie stellen wir uns das vor? Im Konzeptpapier ist vorgesehen, dass wir zwei Pilot Use Cases umsetzen und ausschreiben. Der eine, Modul 2 im Bereich Energie, da läuft aktuell bereits die Beauftragung und jetzt befinden wir uns im

Marktdialog zu Modul 1 mit Ihnen, worüber wir uns sehr freuen. Wir möchten mit Ihnen sondieren, ob sich in dem Thema ein guter Use Case im Bereich von Long oder Post-COVID finden lässt. Und dann sehen Sie Modul 3. Das ist weniger inhaltlich getrieben, sondern da geht es darum, tatsächlich den konkreten Aufbau des Dateninstitutes voranzutreiben, und zwar auf Basis der Erkenntnisse aus Modul 1 und 2. Zum Modul 3 hatten wir vor der Sommerpause, es war Anfang Juli, bereits einen Marktdialog durchgeführt. Dort haben wir auch die Wichtigkeit der Verknüpfung mit den inhaltlichen Erkenntnissen aus Use Cases 1 und 2 betont. Der ist auch auf der Homepage verfügbar, falls Sie Interesse haben sich das anzuschauen. Was wir uns erhoffen ist wirklich mit diesem Konzept einen anderen Weg zu gehen als klassisch Organisationsaufbau in der Verwaltung in vergangenen Jahren stattgefunden hat. Das man im Kleinen anfängt, pilotiert und dann die Bedarfe skaliert im Aufbau des Dateninstitutes.

Wie kam man bei Modul 1 darauf das Thema Post-COVID anzugehen? Es ist so, dass in der Empfehlung der Gründungskommission das Thema aufgegriffen wurde, weil man an diesem Use Case sehr schön die Gemeinwohlorientierung sehen kann. Also was kann das Datenteilen wirklich für den Einzelnen bringen? Wo ist der Mehrwert daran? Denn wir haben aus den Jahren der Pandemie alle leidvoll erfahren, was es heißt, wenn Datenverfügbarkeit nicht bestmöglich umgesetzt ist. Und deswegen ist die Eingängigkeit dieses Themas sehr gut, also hohe gesellschaftliche Relevanz, aber auch mit Komplexität in der Akteurslandschaft, das wissen Sie alle, und die Kolleginnen aus BMBF und BMG auch noch mal viel besser als wir. Aber genau das ist der Anspruch des Dateninstitutes, diesen Multi-Stakeholder-Charakter beim Datenteilen, dort die Hemmnisse aufzubrechen, die uns im Wege stehen. Insofern ist uns sehr daran gelegen, dass wir eine enge Zusammenarbeit mit BMG und BMBF haben. Sie werden es wissen, weil Sie in Teilen bereits daran beteiligt sind beziehungsweise es aktuell in der Presse lesen können: Es gibt eine Reihe Post-COVID-Initiativen in den beiden Ressorts. Wenn Sie fragen, wo sind die Abgrenzung oder die Synergie? Der Fokus und der Output sind bei uns ein Stück weit ein anderer, weil wir damit exemplarisch eine Arbeit des Dateninstitutes belegen wollen. Das ist für uns der primäre Outcome dieses Pilot Use Cases, damit das gut in das Aufbaukonzept hineinpasst. Das heißt nicht, dass wir nicht auch interessiert daran sind, dass das Thema insgesamt vorangetrieben wird. Das unterscheidet uns eben von den Förderinitiativen, die es noch mal gesondert aus BMBF und BMG gibt. Und als letztes: Wie wollen wir das Ganze umsetzen? Es gibt natürlich klassische Forschungsförderung, die man sich in so einem Kontext anschauen oder umsetzen kann. Das möchten wir hier ein Stückchen weit anders machen, nämlich im Rahmen einer sogenannten Challenge. Da geht es um die vorkommerzielle Auftragsvergabe. Dieses Instrument wird zum Beispiel erfolgreich von der SPRIND, der Agentur für Sprunginnovation, genutzt. Das heißt, Sie würden am Ende dieses Marktdialoges, wenn wir den Use Case eingegrenzt und ressortseitig abgestimmt haben, eine Ausschreibung oder einen Aufruf veröffentlicht sehen, auf die sich dann unterschiedliche Teams bewerben können. Es gibt einen Unterschied zur klassischen Projektförderung: Es gehen mehrere Teams durchaus vielleicht mit völlig unterschiedlichen Lösungsideen und Ansätzen ans Rennen, werden parallel gefördert und dann gibt es nach zuvor festgelegten Meilensteinen Jurysitzungen, wo die Zwischenergebnisse reviewt werden. Dann geht man mit den vielversprechendsten Lösungsansätzen in die nächste Etappe, diese werden dann weiter gefördert. Und gegebenenfalls werden bereits erste Teams ausscheiden, es wird weitere Iterationsstufen geben und am Ende hat man Gewinner-Lösungsansätze. Das Schöne an diesem Prozess, und das zeigt die Erfahrung mit der SPRIND ist, das ist ein lernendes System. Auch Teams, die am Anfang ausscheiden, da kann man, und das ist auch der Grund, warum wir uns für dieses

Verfahren entschieden haben, genau sehen, woran es denn gehapert hat. Warum war das nicht der beste Ansatz? Und genau das erhoffen wir uns als Learning für ein Dateninstitut. Zu schauen, wo wäre der Hebel gewesen, wo ein Dateninstitut unterstützen kann, damit solche Projekte eben nicht scheitern. Das kurz zum Instrument und damit bin ich durch und übergebe wieder an Eileen.

### **Eileen Fuchs**

Wir würden jetzt direkt in die Fragen starten, die wir an Sie vorab versandt hatten. Uns interessieren Ihre jeweiligen Antworten auf diese Fragen. Wir haben vorab eine Rednerliste erstellt, die würde ich durchgehen und Sie nacheinander aufrufen. Ich möchte Sie bitten, im Interesse der Wettbewerbsgleichheit, auch mit Blick auf eine eventuell stattfindende Vergabe in Form einer Challenge, diese Zeit einzuhalten - es sind zwei Minuten pro Person und pro Organisation - und sich da entsprechend kurz zu fassen. Wir werden die Zeit stoppen. Es geht uns darum, von Ihnen ein erstes Feedback auf diese Fragen zu bekommen. Sie brauchen keine vollumfängliche, geschliffene Antwort zu geben. Wir wollen Ihre Ideen wissen und erfahren, wo es Raum gibt für einen Use Case im Bereich der Post-COVID Forschung und zum Austausch bei der Datenlage. Ich würde nun beginnen, die Kollegin Frau Lena Müller stoppt die Zeit und wird mir ein Zeichen geben, wenn Sie Ihre Zeit überschreiten. Dann würde ich Sie gegebenenfalls unterbrechen beziehungsweise Ihnen ein Zeichen geben, zum Ende zu kommen. Zu Beginn wird Frau Müller Ihnen einen Überblick über die Fragen geben und dann starten wir in die Rednerliste.

### **Lena Müller**

Die Fragen wurden im Vorhinein rumgeschickt, von daher gehe ich davon aus, dass Sie die grobe Struktur kennen. Ich würde jeweils einmal kurz sagen, was wir uns da jeweils bei gedacht haben.

Zur ersten Frage, wo es um die Datenlage geht und welche Fragen aktuell nicht beantwortet werden können. Das ist natürlich der Ausgangspunkt für einen Use Case, dass man offene Fragen damit idealerweise beantworten kann. Es wäre für uns nicht nur die medizinische Forschungssicht relevant, sondern auch gesellschaftliche Dimensionen oder die Patientensicht. Es muss nicht unbedingt, wie sich das bei Post-COVID natürlich anbietet, in die medizinische Richtung gedacht werden. Da würde uns interessieren, was für Sie gerade noch offen ist.

Die zweite Frage dreht sich konkreter um die Datenquellen, also welche Datenquellen können Sie sich vorstellen oder kennen Sie, die noch nicht erschlossen oder aufbereitet sind. Insbesondere hier auch Datenquellen, die noch nicht in diesem Kontext genutzt werden, wie privatwirtschaftliche Datenquellen und Datenquellen der Verwaltung?

Die dritte Frage, da geht es darum, ob Sie schon konkrete strukturelle Roadblocker sehen, die Sie behindern bei der Datennutzung in dem Kontext. Was führt zu Verzögerungen, wo gibt es eine fehlende Vernetzung, zu welchen Akteuren fehlt der Kontakt? Das würde uns dort interessieren.

Und zum Schluss dann die Frage: Gibt es ganze Bereiche, die bisher nicht datenbasiert abgedeckt sind? Also wir gehen davon aus, in den medizinischen Bereichen werden an verschiedenen Stellen Daten erfasst. Aber wie sieht das beispielsweise bei sozialwissenschaftlicher Forschung aus, bei dem Alltag der Patienten, gibt es darüber eine Erkenntnislage? Wie sieht es mit der gesellschaftlichen Aufklärung, Akzeptanz aus? Also die verschiedenen Dimensionen von Post-COVID bedacht, gibt es noch Bereiche, wo eine Datenerhebung sinnvoll wäre?

**Eileen Fuchs**

Dann würden wir mit den Redebeiträgen beginnen.

### **Charité – Koordinierungsstelle Netzwerk Universitätsmedizin**

Zu der Frage, was sind relevante Fragen und was kann mit der aktuellen Datenlage nicht beantwortet werden. Das ist natürlich ein Riesenfeld. Ich konzentriere mich jetzt mal auf meine Perspektive Universitätsmedizin. Da ist ein riesen Thema zu verstehen, warum Menschen nach einer akuten COVID Infektion überhaupt an Post-COVID erkranken. Was sind die auslösenden Faktoren? Dazu braucht man Daten am Übergang von der Akutphase der Erkrankung in die Langzeiterkrankungen. Wir haben dazu in Deutschland in den letzten Jahren intensiv Daten gesammelt, auch gute Datensammlungen angelegt. Wir haben aber jetzt insbesondere bei der Verlagerung der akuten Erkrankungen in das niedergelassene Setting das Problem, dass wir eben diese Patienten nicht in der Frühphase der Infektion finden, sondern die nur sehen, wenn sie im Krankenhaus sind. Das heißt, an dieser Schnittstelle ambulante Versorgung, stationäre Versorgung und Forschung, und da haben wir ein Thema, wo man sicherlich über das, was aktuell schon an Daten da ist, hinaus noch sehr viel mehr machen könnte, weil man da eben auch Bioproben braucht und sehr, sehr fein auflösende Krankheitsverläufe beschreiben muss.

Die zweite Frage, welche Datenquellen gibt es, die noch nicht erschlossen, aufbereitet, verknüpft wurden. Es gibt natürlich eine Unmenge an Datensilos, wo man Daten aus verschiedensten Perspektiven zur Post-COVID findet. Meine These wäre, dass die wenigsten dieser Silos bisher verknüpft sind. Ich kann das am Beispiel machen, Patienten-Behandlungsdaten aus der Klinik versus der Arztpraxis, das sind zwei getrennte Welten. Krankenkassendaten wäre eine weitere getrennte Welt und so weiter. Da gibt es sicherlich einen riesen Handlungsbedarf.

Nach wie vor gibt es strukturelle Roadblocker, mit Sicherheit die Tatsache, dass wir in Deutschland nach wie vor keinen Unique Identifier haben. Das heißt die gleiche Person wird in unterschiedlichen Datensammlungen unterschiedlich bezeichnet, was die Verknüpfung natürlich extrem erschwert. Aber es ist ganz sicher auch ein Ressourcenproblem. Diese Daten müssen von den datenhaltenden Stellen aufbereitet und zusammengestellt werden, um für spezifische Anfragen herausgegeben werden und zur Nutzung zur Verfügung gestellt werden zu können. Gerade dann, wenn gute Daten da sind, kommen auch viele Anfragen und das ist ein hochaufwendiges und komplexes Geschäft. Das heißt Ressource ist da ein Schlüssel und weil diese Ressource oft fehlt, fehlt eben auch die Bereitschaft vieler Akteure, diese Daten herauszugeben, zur Nutzung zur Verfügung zu stellen. Wir haben also ein Ressourcen-, in Teilen aber auch ein Kulturproblem, was das Datenteilen betrifft.

Zum letzten Punkt, wo wirklich Felder noch komplett unerfasst sind, kann ich wenig sagen. Einen Punkt würde ich gerne highlighten, der aus unserer Sicht wichtig ist. Wir haben zwar Daten aus vielen verschiedenen Behandlungssettings, die dann in den jeweiligen Settings vorhanden sind. Wir haben aber im Moment keinerlei Gesamtbild über Behandlungsverläufe dieser Patientinnen im System. Also die Frage, welche Wege sie von der Klinik in die Praxis in die Rehaeinrichtungen und so weiter nehmen. Dazu gibt es im Moment sehr wenig Erkenntnis. Die Hypothese wäre, dass das extrem heterogen verläuft. Wir wissen nicht, wie die Patientinnen behandelt werden und wo sie behandelt werden, da mehr Licht hineinzubringen wäre, glaube ich dringend und wichtig.

### **Simba n<sup>3</sup> Datenlösungen**

Ich bin hier dazu gekommen über die Wirtschaftsförderung vom Erzgebirge. Hintergrund ist, dass wir eventuell einen Beitrag zur Frage 2 geben können. Die Simba n<sup>3</sup> GmbH hat während der COVID Phase eine Software implementiert, um Schnelltests zu erfassen, die an die Corona Warn App etc. weitergeleitet worden sind. Wir haben hier einen gewissen Datenschatz aus unserer Sicht, das sind 22,5 Millionen dokumentierte Schnelltests, die auch damals schon während der Pandemie eine Positivquote ausgespuckt haben. Wo ich nicht nur wusste, wie viele sich mit PCR Test wirklich positiv bestätigt haben, sondern ich habe das Verhältnis von positiv zu negativen Testfällen gehabt. Und gerade im Bereich Sachsen hatten wir acht Landkreise, wo ich in mehreren Landkreisen die Bevölkerung doppelt bis dreifach in Summe getestet hab, dadurch schon sehr gute Indizes vor den möglichen PCR Tests hatten. Das einfach nur als mögliche Datenquelle, die wir hier einbringen können. Das in dem Rahmen mit anzubieten, genau das wäre die Möglichkeit, die Simba n<sup>3</sup> beisteuern kann.

### **Medizininformatik Initiative**

Die Medizininformatik Initiative hat eine Infrastruktur aufgebaut, um die Versorgungsdaten, die in den Universitätskliniken und Kliniken erfasst werden, für Nachnutzungszwecke verfügbar zu machen. Das heißt, dass was wir jetzt bereits an Versorgungsdaten erschlossen haben. Dieser Datensatz wird weiter ausgebaut, was noch fehlt wären eine breite Anbindung von Daten, die aus Arztpraxen zum Beispiel kommen oder aus Pflege- und Rehabilitationseinrichtungen. Hier gibt es bislang nur modellhafte Umsetzungen, die sicherlich weiter ausgebaut werden müssten. Darüber hinaus würden hier Daten fehlen, die zum Beispiel Aufschluss geben über die Arbeitsfähigkeit, über Umwelteinflüsse und die Lebensqualität von Patienten, die an Long oder Post-COVID erkrankt sind. Datenquellen, die entsprechend noch nicht oder noch nicht hinreichend erschlossen sind, sind hausärztliche oder niedergelassene fachärztliche Versorgungseinrichtungen. Aber auch Daten, die zum Beispiel von mobilen Devices kommen und dazu dienen, dass Patienten ihre Symptome selbst tracken können, damit die auch außerhalb dieser ärztlichen Versorgung erfasst werden können. Das wäre aus meiner Sicht wichtig. Die MI arbeitet daran, die Daten, die in der Routinedokumentation in den Krankenhäusern erfasst werden, auch mit Krankenkassendaten und Registern zu verknüpfen. Hier werden gerade die Voraussetzungen entwickelt damit das zukünftig möglich ist, aber noch können wir das nicht. Strukturelle Roadblocker, da kann ich mich dem anschließen, was der Kollege der Koordinierungsstelle Netzwerk Universitätsmedizin eben schon berichtet hat. Ich würde zusätzlich sagen, dass die Dokumentation der Versorgung noch nicht ausreichend ist. Sie ist noch zu abrechnungsorientiert, noch zu wenig standardisiert und dadurch fällt eben viel Ressourcenaufwand an, um diese Daten aufzubereiten und das verringert auch die Datenqualität. Da wäre es wichtig, eben in der Hinsicht etwas zu tun.

### **Universität Leipzig, Medical Data Science**

Universität Leipzig, wir sind involviert in die Medizininformatik Initiative, wie auch in das Netzwerk Universitätsmedizin und NFDI4Health. Ich kann mich dem anschließen, was von der Medizininformatik Initiative und dem Netzwerk Universitätsmedizin schon ausgeführt wurde. Uns interessieren Ursachen für Post-COVID unter Einbeziehung longitudinal Behandlungsdaten, also gesicherte Vorerkrankungen und nicht self-reported Erkrankungen. Aber auch Daten zum Lifestyle: Rauchen die Personen, treiben sie Sport und so weiter. Wichtig wäre auch eine Frage zu Therapieoptionen, welche Therapieoptionen gibt es für

welche Konstellation der Patienten? Wie groß ist der Einfluss auf den Arbeitsalltag und das Privatleben? Lebensqualität wurde auch schon genannt, Ausfallzeiten, mentale Gesundheit spielen hier eine Rolle.

Roadblocker: Natürlich gibt es einzeln Zugriff auf Studiendaten, Behandlungsdaten, Krankenkassendaten, aber die Vernetzung fehlt. Auch hier hat der Kollege vom Netzwerk Universitätsmedizin schon ausgeführt, es fehlt eigentlich der nationale Identifikator, der einheitliche Indikator, der das Recordlinkage ermöglicht. Ich sehe auch den Datenschutz als gesetzliche Hürde, da oftmals eben die Zustimmung der Person eingeholt werden muss. Und wenn das eben nicht passiert, dann kann man die Daten auch nicht zusammenführen. Und das obwohl die technischen Lösungen bereits verfügbar sind. Zumindest für COVID-Plattformen hatten wir da mehrere.

Datenbereiche, die bislang schlecht datenbasiert abgedeckt sind. Die Kollegin der Medizininformatik Initiative hat das auch schon ausgeführt, der Kollege des Netzwerk Universitätsmedizin genauso, ich sehe hier noch dazu Medikation. Oftmals sehen wir nur die Verschreibungen, den Medikationsplan, aber wichtig wäre doch wer hat was wann eingenommen? Niedergelassene Bereich wurde schon genannt und für mich auch relevant, der Heimbereich, der oftmals nur über Studien abgedeckt wird. Also welchen Einschränkungen unterliegen die Post-COVID-Erkrankten? Welche Hilfsmittel gibt es dafür und wiederum noch mal ganz stark heraus genommen der Aspekt der Vernetzung der unterschiedlichen Bereiche.

### **Eileen Fuchs**

Es kam gerade eine Frage auf, wie wir vorgehen. Vorab wurden Ihnen die Rednerliste erstellt und die Nummerierung versandt und daran halten wir uns. Insofern, falls Sie die nicht erhalten haben, würde ich einfach nochmal fragen, falls Sie nicht mitbekommen haben, dass Sie dran sind.

### **DESAM ForNet**

Ich stimme auch völlig dem zu, was Netzwerk Universitätsmedizin, Medizininformatik Initiative und Universität Leipzig gerade gesagt haben. Ich bin für die Deutsche Stiftung für Allgemeinmedizin und das Projekt DESAM ForNet hier aus dem niedergelassenen Bereich. Eine Frage die in den niedergelassenen Bereichen wichtig ist, in welcher Reihenfolge diese diagnostischen Maßnahmen, die man ergreifen soll, überhaupt Sinn machen. Das ist eine Frage die die Allgemeinmedizin umtreibt. Und wie kann man diese Informationen, die wir dafür brauchen, tatsächlich aus den verfügbaren Datenquellen holen. In den allgemeinärztlichen Praxen findet die Datensammlung hier in der Regel relativ konzentriert auf diese Abrechnungsfragestellungen statt und man hat relativ wenig Überblick und wenig Möglichkeiten den Gesamtverlauf zu dokumentieren, der unter Umständen auch speziell ist. Das heißt, wir haben unstrukturierte Daten, Textdaten. Da wäre eine Frage, wie man das entsprechend gestalten kann, dass die Daten und Beschreibungen zu Krankheitsverläufen entsprechend nachvollziehbar und auswertbar sind. Und um an diese Informationen zu kommen, ist im niedergelassenen Bereich tatsächlich diese Strukturierung und Standardisierung der Datenerfassung in den medizinischen Informationssystemen eine Baustelle. Da wäre die Frage, wie kann man da eine entsprechende Standardisierung erreichen und daran anschließend auch wie kann man die Daten aus den verschiedenen Systemen zusammenführen, so dass die auch in einem größeren Kontext genutzt werden können, auf der Bundesebene.



Damit sind wir im Prinzip auch beim vierten Punkt. Da stecken auch die Daten, die bisher nicht erfasst sind, mit drin. Im Grunde alles, was die Vorredner schon gesagt haben, plus der Tatsache, dass man in der Praxis Ausschlussdiagnostik betreibt und ganz viel Forschung macht und hier ein flexibles Modell braucht, verschiedene Daten oder Fragestellungen zu erfassen.

### **Ernst & Young GmbH**

Ich kann mich dem anschließen, was schon gesagt wurde. Aus unserer Sicht ist in der Tat die tatsächliche Prävalenz von Long-COVID inklusive Auslöser und Risikofaktoren sowie deren Auswirkungen auf die direkt Betroffenen und deren Umfeld inklusive natürlich auch Lebensqualität, Arbeit und so weiter von höchster Relevanz. Und dabei ist auch die Untersuchung von Risikofaktoren sehr wichtig und das ist aktuell unzureichend erforscht. Beispielsweise sozialökonomische Deprivation inklusive Lebens-, Wohn- und Arbeitswelt. Hier werden eventuell Daten von der Bundesagentur für Arbeit relevant. Aber auch die Wechselwirkungen zwischen bereits identifizierten Risikofaktoren, also Schwere des Krankheitsverlaufes, Impfstatus oder vorgehende Krankheiten. Und dafür werden große Datensätze benutzt, was nochmal eine Herausforderung ist. Weiterhin bleibt auch eine große Forschungsaufgabe der Systematik-Cluster auszudifferenzieren und längerfristig verschiedene Behandlungsmethoden auf Wirksamkeit und ökonomische Nutzen zu beleuchten. Hinsichtlich relevante Datenquellen sind Daten im Gesundheitsbereich besonders relevant, also beispielsweise von RKI, GKV, DRV und diverse klinische epidemiologische Forschungskohorten, beispielsweise LIFE, aber auch Daten von privaten Unternehmen sind sehr wichtig.

Hinsichtlich kulturelle Roadblocker beinhaltet der Long-COVID Anwendungsfall natürlich die enorme Herausforderung, dass es sich meist um persönliche Gesundheitsdaten handelt, welche den höchsten Schutzbedarf genießen. Und damit sind Themen wie beispielsweise Anonymisierung von Daten wichtig anzugehen.

Darüber hinaus gibt es allgemeine Roadblocker, die man auch bei anderen Use Cases sehen kann, die man auch berücksichtigen muss. Beispielsweise besteht unserer Erfahrung nach in der Praxis leider häufig nur eine geringe Bereitschaft bei Organisationen Daten zu teilen, was in der Konsequenz zum Scheitern von Anwendungsfällen führen kann. Daher ist es wichtig, von Anfang an mitzudenken, wie man die Daten bereitstellen und Organisationen motivieren kann, also Anreizsysteme schafft um die Daten zu teilen.

### **NAKO Gesundheitsstudie**

Ich nehme zusammen mit meinem Kollegen für die NAKO Gesundheitsstudie teil. Die NAKO ist zurzeit die größte deutsche bevölkerungsbezogene Kohortenstudie, die in Deutschland läuft. Und ich kann mich unseren Vorrednern anschließen, kann dazu aber sagen, dass die NAKO Informationen zu diesen ganzen Risikofaktoren sammelt. Wir haben Informationen gerade was die Umweltbelastung anbelangt, aber auch die suszeptiblen Subgruppen, also Omics-Daten von unseren Kohortenteilnehmern. Lifestylefragen oder Lifestyledaten werden erhoben, wie zum Beispiel, ob die Leute physisch aktiv sind, sozioökonomische Fragen etc., das wird bei uns alles mit erhoben.

Wenn wir zu dem zweiten Punkt kommen, welche Datenquellen nicht erschlossen sind oder noch nicht verknüpft wurden, das sind die klinischen Daten mit unserer Kohortendatenbank. Dann natürlich die Sekundärdaten von den Krankenkassen oder auch von anderen Quellen wie RKI, die halt nicht so leicht mit den epidemiologischen Daten verknüpft werden können. Und das sind Roadblocker für uns. Wir versuchen zwar mit den Krankenkassendaten eine

Verknüpfung herzustellen, aber nicht alle Krankenkassen in Deutschland sind bereit, ihre Daten mit der NAKO zu verknüpfen.

Zu der vierten Frage, welche Themenbereiche sind noch nicht erfasst. Das ist meines Erachtens auch der Themenbereich, ob Post-COVID nach der Impfung oder nach der Erkrankung gekommen ist. Wir haben mittlerweile sehr viele Leute, gerade in Deutschland, die beides hatten, die sowohl geimpft sind, als auch schon mehrere Erkrankungen hinter sich haben. Und da ist es ziemlich schwierig, das auseinanderzuhalten. Was war eigentlich der Auslöser für die Post-COVID Erkrankung?

### **Materna Information & Communication SE**

Zu Frage 1 wollte ich keine Antwort geben, weil wir hier so viele Experten hatten und auch schon sehr viele Fragen beantwortet haben. Zu Datenquellen, wir haben beim Paul-Ehrlich-Institut die App SafeVac 2.0 gebaut, die Freiwillige befragt. Das wollte ich hiermit reingeben. Bei der GIZ hat auch eine große Studie auf Basis von vielen verfügbaren COVID-Daten stattgefunden, wo es darum ging warum einige erkranken und andere nicht, also welche positiven Faktoren wirken. Und weil ich aus der Rolle die Datendemokratie auch immer schaue, ob es sozusagen ungewöhnliche Akteure gibt, wollte ich diese mit erwähnt haben.

### **McKinsey & Company Inc.**

Auf Grund der sehr gehaltvollen Antworten der Vorrednerinnen vor mir werde ich mich relativ kurzhalten, was die ersten Fragen angeht. Ich denke wichtig ist hierbei zu beachten, dass natürlich insbesondere im medizinischen Feld viele Daten, die da sind, nicht verknüpft sind. Im Zuge eines Dateninstitutes kann auch interessant sein, wie man über Verwaltungsbereiche hinweg Daten miteinander verschneiden könnte, um daraus Mehrwerte zu erzielen, die dann auch gesamtwohlfahrtschaftlich dienen. Dazu gibt es relativ viele strukturelle Roadblocker, viele wurden schon genannt. Deswegen gehe ich jetzt auf Kultur, Ressourcen und so weiter nicht mehr ein. Ich denke, was dabei sehr wichtig ist, insbesondere auch in ein Aufgabenfeld eines Dateninstituts reinspielen kann, ist die Frage, wie selbst dann, wenn Akteure Daten teilen wollen, sie das dürfen und können. Das auch interoperabel gestaltet wird, also wie gestaltet man das semantisch, organisatorisch, technisch? Wie bekommt man das rechtlich abgebildet? Da gilt es natürlich entsprechende Standards zu haben oder zu entwickeln und insbesondere bei diesen sehr schätzenswerten Daten auch über eine entsprechende Tokenisierung, Identifizierung, und so weiter nachzudenken. Da gibt es sicherlich viele strukturelle Roadblocks, die man durchdenken müsste. Dabei würde ich es aufgrund der sehr umfangreichen Vorreden auch belassen.

### **Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin**

Auch ich schließe mich vielen Vorrednerinnen vollumfänglich an. Ich glaube eines der wesentlichen Probleme ist, dass die Post-COVID Erkrankung kein einheitliches Krankheitsbild ist und dass eine sehr schwierige Definition hat. Uns fehlen absolut biologische Marker der Erkrankung und wir werden auch keine gezielten Ansätze für die Prävention und Behandlung entwickeln können, solange jeder Mensch, der in der Folge einer COVID Erkrankung irgendwie eine Veränderung seines Gesundheitszustandes bemerkt, unter diesem Begriff geführt wird. Wir müssen da noch sehr viel definitorische und auch biologische Arbeit leisten. Dafür brauchen wir zum einen wirklich gute systembiologische Ansätze, wo wir versuchen verschiedenste Hochdurchsatzverfahren der Multi-Omics zum Ansatz zu bringen und dann bioinformatisch uns nochmal den gesamten Problemkomplex anzunähern. Wo wir das Proteingenom, das gesamte Genom, das Transkriptom dieser Patientinnen genau

beleuchten und versuchen, bessere Cluster zu bilden von reproduzierbaren Krankheitssyndromen, für die wir eine biologische Ursache finden, die wir in Zukunft testen können. Aus der nationalen Pandemiekohorte heraus haben wir einen der besten charakterisierten Datensätze, den wir sehr frei zur Verfügung stellen. Wir haben im Netzwerk Universitätsmedizin entsprechende Plattformen aufgebaut, um datentechnisch solche und andere Datensätze mit der Community zu teilen. Da sind wir für Kooperationen ausgesprochen offen. Es sind sehr viel mehr Daten, als wir alleine in einer vernünftigen Zeit auswerten können. Ich glaube, dass der aktuelle Fokus auf Post-COVID sehr, sehr gefährlich ist, wenn wir nur in diese Ambulanzen und in Register reinschauen. Ich würde dringend empfehlen sich auch Datensammlungen zu erschließen, die nach 2019 und davor zurückreichen. Wirklich Gesundheit im Verlauf beobachten können, um aufzudröseln, was hat die Pandemie allgemein gemacht, was haben Impfungen gemacht, was hat tatsächlich die Virusinfektion gemacht? Ich glaube da liegen noch ungeborgene Schätze wie die Nationale Kohorte oder eben die Routinedaten von Versicherern von den Datenintegrationszentren, Medizininformatik. Das wir noch mal vor der Pandemie ansetzen und das Ganze von vorne aufrollen.

### **Charité – Institut für Medizinische Soziologie und Rehabilitationswissenschaften**

Wir beschäftigen uns mit sozialwissenschaftlicher Forschung, uns interessieren Versorgungsverläufe, sozialstrukturelle Merkmale bei unterschiedlichen Klientelen der Patienten und Patientinnen. Für uns ist es also interessant, Krankenkassendaten, aber auch Daten anderer Versicherungsbereitsteller zu verknüpfen mit Personendaten. Fehlende Datenquelle ist in erster Linie, wer hatte überhaupt Corona? Wir haben keine einheitliche Datenbank darüber, wer eine Infektion durchlebt hat, was auch der Erkrankung selber geschuldet ist. Wir haben keinen Überblick, wer hat Post-COVID? Die Definition wurde schon angesprochen, aber auch die unterschiedlichen anderen Aspekte, die diese Erkrankung mit sich bringt. Wir haben kein Impfregister, wir wissen nicht, wer überhaupt alles geimpft ist in einer einheitlichen Datenbank. Und wir haben ein nicht sehr gut verknüpft System aus Krankenpflege und Rentenversicherungen mit deren Daten. Der Identifier wurde bereits genannt. Das alles sehe ich als Schwierigkeiten, eine gute, aussagekräftige Studie zu machen. Die besten Studien kommen aus solchen Systemen mit großer Integration, zum Beispiel dem NHS. Wir sind gezwungen, uns mit einzelnen Krankenkassen zusammenzutun und die sozialstrukturellen Merkmale selber zu erheben. Sowas wie Geschlecht, Herkunft, wie auch immer und patient reported outcome measures, also was bedeutet das für die Leute, die betroffen sind von den Erkrankungen oder von diesem Komplex? Da sehe ich die größten Herausforderungen und ich hoffe und bin gespannt, ob dieses Vorhaben hier diese Fragen beantworten kann.

### **PwC Strategy (Germany) GmbH**

Ich schließe mich allen Vorrednern an, es wurde schon sehr viel gesagt. Ich bin heute auch viel mehr in der Rolle da als Schnittstelle zwischen Modul 3 und 1. Um auf die Fragen einzugehen müsste man genau beleuchten, wo man auch den größten Mehrwert stiften kann als Dateninstitut. Vielleicht auch Fragen aufgreifen, was ist eigentlich Post-COVID, warum ist es aufgetreten, welche Folgen hat es und welche Therapien sind für wen erfolgreich? Der letzte Punkt wäre wahrscheinlich mit der interessantesten, ganz vereinfacht ausgedrückt. Zu den Fragen 2 und 3, zu den Datenquellen und strukturellen Herausforderungen, wäre eine Frage, ob sich das Dateninstitut primär mit Daten beschäftigt, die schon vorliegen oder darauf hinwirkt, dass mehr und bessere Daten vorliegen können.

Ich denke da an die Hausarztpraxen, die aktuell nicht wirklich incentiviert sind, an Forschung teilzunehmen und ihre Daten und das Wissen, dass sie haben, strukturiert aufzubereiten und bereitzustellen. Eine Möglichkeit wäre da die elektronische Patientenakte. Ob das Dateninstitut mit darauf hinwirken könnte, dass Anreize geschaffen werden. Da sehe ich schon noch Werte, die genutzt werden können, um die Fragen, die ich zu Anfang genannt habe beantworten zu können. Ansonsten, das ist jetzt nicht zwingend ein strukturelles Hindernis, aber schon eine gewisse Herausforderung, sich mit den diversen Aktivitäten und Gesetzgebungsinitiativen, die aktuell in diesem Kontext laufen, abzustimmen und einzubetten. Der europäische Raum für Gesundheitsdaten ist auch in der Mache, Registergesetz im BMG. Dass wir uns da gut einbetten und all das auf dem Schirm haben und parallel berücksichtigen, wie das unsere Arbeit beeinflusst.

### **Philipps Universität Marburg**

Es ist schon einiges gesagt worden, deswegen möchte ich es recht kurz halten. Ich glaube, dass wir wenig Daten haben über die Versorgungslücken, das heißt die Wartezeiten der Patienten, die Infrastruktur in den unterschiedlichen Landkreisen. Das ist zumindest mein Alltag, dass die Patienten dort auf sehr unterschiedliche Strukturen zurückgreifen können. Der Alltag ist zudem erschwert und weil wir keine Ahnung haben, wie viele Arztbesuche, wie viele unterschiedliche Therapien im Vorfeld stattgefunden haben. Nicht immer liegen die Briefe vor und die Patienten führen das alles nicht ganz so strukturiert. Es ist so, dass dieser Bereich der Phänotypen, der schon angesprochen wurde, für mich ein ganz wichtiger ist. Wir müssen versuchen einheitlich festzulegen, wie wir auf diese Phänotypen hin versuchen, Diagnostik und Therapie maßzuschneidern. In diesem Zusammenhang spielen die Komorbiditäten für mich eine große Rolle, die nicht so wirklich gut erfasst sind, auf die wir bislang an vielen Stellen leider nicht zurückgreifen können und das spiegelt sich auch in der Qualität der Versorgung wieder. Wir brauchen hier Qualitätsstrukturen, eine Möglichkeit zur Qualitätsüberprüfung. Für den Bereich, den ich überblicke, für den rehabilitativen Bereich, ist es so, dass sich viele Leute tatsächlich die Reha draufschreiben, aber es gibt keine einheitlichen Strukturen, an denen man Erfolg messen kann. Die Vernetzung ist dort sicherlich wichtig und da ist der Datenschutz ein großes Hindernis. Das ist schon mehrfach genannt worden. Welche Quellen wir bislang noch nicht angezapft haben, das sind die digitalen Quellen. Das, was an Trackingdaten auf dem Smartphone ist, ist sicherlich hilfreich, um zum Beispiel Bewegungsprofile erstellen zu können. Einen letzten Punkt, den ich machen möchte, sind die unterschiedlichen Strukturen hinsichtlich Akutmedizin und Rehabilitation. Die sind sehr stark getrennt und müssten meines Erachtens fließend sein, auch im Bereich Somatik und Psychosomatik. An vielen Stellen bräuchten wir da eigentlich Hybridkonzepte.

### **Deutsche Rentenversicherung Bund – erste Rednerin**

Ich möchte mich konzentrieren auf die Punkte 3 und 4, weil vieles zu 1 und 2 schon gesagt wurde. Mein Kollege wird noch was zu den Datenquellen zur DRV sagen. Aus meiner Sicht ist das Problem, dass wir im Moment viele Daten aus self-reported Studien haben. Wir müssen uns noch stärker fokussieren auf Studien wie die NAKO oder es gab vor zwei Jahren die Gutenberg-Studie aus Mainz, um valide Daten auch in unserem System zu haben. Ein anderer Punkt, der aus unserer Sicht wichtig ist, dass wir eine bessere Einteilung nach Schweregraden der Erkrankung, nach vorherrschenden Symptomkomplexen brauchen. Ein Roadblocker ist die aktuelle Kodierung nach ICD-10. Die U09.9 ist ein Sekundärkode, das heißt es muss im Bereich des SGB V noch ein Primärkode gefunden werden. Nicht immer findet sich eine Diagnose aus anderen Bereichen und Symptome, die vorherrschend sind,

können im Moment nicht gut über die ICD-10 abgedeckt werden. Auch da sind noch Anpassungen möglich, um aus den vorhandenen Daten die Fälle herauszufinden. Und jetzt würde ich gerne an meinen Kollegen übergeben.

### **Deutsche Rentenversicherung Bund – zweiter Redner**

Ich würde mich auf die ersten beiden Dinge konzentrieren. Hintergrund ist, dass ich den Reha-wissenschaftlichen Part vertrete und meine Kollegin den medizinischen Part. Welche relevanten Fragen können wir mit der aktuellen Datenlage nicht beantworten? Es ist bei Post-COVID wie bei allen anderen Erkrankungen auch, wir können Versorgung über Sozialgesetzbücher hinweg nicht abbilden, egal welches Krankheitsbild. Das heißt Menschen, die einen Akutaufenthalt hatten oder bei einem niedergelassenen Arzt waren: Wir wissen nicht, wie viele davon in Rehabilitation der Rentenversicherung gelandet sind, es sind ja alle im erwerbsfähigen Alter, und wie geht es danach weiter im Akutbereich oder eben auch im ambulanten Bereich? Diese Versorgungspfade sind momentan nicht darstellbar, nur wenn man SGB V und VI Daten zusammenspielen könnte. Es sollte gerade bei Post-COVID auch noch einen großen Schwerpunkt auf dem Thema Teilhabe geben, das ist in Bezug auf die gesellschaftliche Relevanz dieses Krankheitsbildes enorm wichtig. Da können auch Daten der Rentenversicherung sehr helfen, weil wir lückenlose Erwerbsverläufe unserer Versicherten haben und sehr viele Daten, so diese Menschen in der Rehabilitation waren. Sowohl zu der medizinischen als auch der beruflichen Rehabilitation und die Reintegration ins Erwerbsleben.

[schlechte Verbindung] Meine Botschaft: Vergesst nicht die Daten des SGB VI, die sehr wertvoll und vollständig sind und Auskunft zur Teilhabe bieten.

### **TMF e.V.**

Wir sind aktiv unter anderem bei GenomDE, NFDI und nicht zuletzt in der Koordination der Medizininformatik Initiative. Ich möchte nicht alles wiederholen, es ist sehr viel Richtiges gesagt, sondern nur akzentuieren. Gerade die Post-COVID Fragestellungen hat einerseits die Herausforderung, die wir im deutschen Gesundheitswesen bezüglich der Erschließung verteilter Daten haben, offengelegt und positiv gesprochen auch erfordert, dass wir sie lösen. Gerade der Bezug und die Vernetzungsmöglichkeit zur Rentenversicherungsdaten, die andere Identifikatoren nutzen und eigene rechtliche Rahmenbedingungen haben, ist ein wichtiger Punkt, den wir nicht nur rechtlich, sondern auch technisch am Ende umsetzen müssen. Wir müssen schauen, dass wir für diese Fragestellung neben prospektiven Kohorten auch den Blick in die transsektoral vernetzbaren Versorgungsdaten gewährleisten müssen. Da sind wir in vielen Punkten noch nicht ideal aufgestellt und insofern ist gerade die spannende, zu diskutierende Frage: An welchen Stellen kann ein Dateninstitut helfen? Weniger vielleicht bei der primären Datenerhebung, sondern genau bei diesen strukturellen Verknüpfungen. Das von meiner Seite, ansonsten ist sehr viel Richtiges gesagt worden.

### **Labvantage - Biomax GmbH**

Wir sind unterwegs im Bereich Versorgung chronischer Atemwegserkrankungen, COPD. In diesem Zusammenhang würde ich gerne unterstreichen, dass eine ganzheitliche, systematische Diagnose wichtig ist, um die Pathophysiologie von Post-COVID wirklich zu verstehen. Insbesondere unter Einbeziehung von Komorbiditäten, die für solche Krankheiten grundsätzlich eine Rolle spielen. Das Erfassen, Beobachten und Verbessern als einen solchen Prozess etablieren und dabei auch den Burden of Disease gerade bei jüngeren betroffenen Patienten mit aufnehmen.

Welche Datenquellen sind noch nicht erschlossen? Nach unserer Einschätzung sind zurzeit die neurologischen Aspekte, also die neurologischen bildgebenden Daten, unterrepräsentiert. Hier gibt es große Daten oder Referenzdatensätzen aus dem Bereich neurodegenerativer Erkrankungen, die man nutzen und verknüpfen kann mit einer systematischen Datenerfassung im Bereich Post-COVID.

Viele strukturellen Roadblocker wurden hier schon erwähnt. Technisch würde ich das gerne auf das Thema Harmonisierung der Daten fokussieren, da wir es mit einem sehr heterogenen Datensatz unterschiedlicher Quellen zu tun haben. Es muss eine technologische Lösung her, die die semantische Integration erfordert. In dem Sinn „Inputs sind die Rohdaten, die kommen von allen und zurück kriegt die Community Harmonised Data als Output“. Im Sinne einer Implementierung eines Fair Data Prinzips.

Gibt es Bereiche, die bisher nicht datenmäßig erfasst sind? Da würde ich zurückkommen auf das Thema neurologische Daten, Kollektomdaten. Aus meiner Sicht ist es sehr wichtig, dass man in die Lage kommt, die Symptomatik, die man erfasst, auf Organebene zu objektivieren. Da gibt es sehr gute Ergebnisse im neurobiologischen Bereich. Und nachdem das in Post-COVID eine große Rolle spielt, würden wir empfehlen das mit aufzunehmen.

### **DGUV**

Meine Vorredner haben schon das meiste vorweggenommen, ich möchte nur zwei Punkte aufgreifen, die aus meiner Sicht besonders wichtig sind. Das betrifft zum einen die Ausführungen der Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin. Wir haben Schwierigkeiten bei der Kausalität von möglichen Post-COVID Symptomatik und das betrifft uns als Unfallversicherung besonders, da wäre Bedarf besser zu werden. McKinsey hatte das Thema Interoperabilität angesprochen, da haben wir auch eine sehr große Baustelle. Das betrifft uns indirekt, wenn das Projekt für eine Verbesserung sorgen kann. In Abstimmung und im Gleichschritt mit dem BMG, dass seinerseits zwei Gesetzgebungsvorhaben auf den Weg gebracht hat, bringt das alle voran.

### **Deloitte Consulting GmbH**

Ich möchte mich der DGUV anschließen. Sie haben zwar bei den strukturellen Roadblockern extra hingeschrieben „abgesehen von gesetzlichen Hürden“ aber ich finde das diese bei diesem Themenkomplex unwahrscheinlich wichtig sind. Das BMG hat aktuell vier Gesetzesvorhaben laufen, die dieses Thema unmittelbar betreffen. Es wäre notwendig, dass diese Gesetzesvorhaben abgeschlossen werden, eh man diese Struktur überhaupt aufbrechen kann, das ist schlichtweg eine Grundlage, um an diesem Fall zu arbeiten. Der zweite Punkt bei gesetzlichen Hürden ist das Thema Datenschutz, das sollte man sich nochmal ansehen. Die ganzen Dinge, die hier beschrieben wurden, gehen insbesondere aus gesetzlichen Gründen nicht. Die Fragestellung ist dann, wenn man sich für so einen Fall entscheidet beim Dateninstitut, wo ich die entsprechenden gesetzlichen Grundlagen oder Möglichkeiten gar nicht habe, ob das nochmal als strukturellen Roadblocker untersucht wird. Und auch nochmal der Hinweis von der Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin aufzugreifen, was die Definition betrifft. Da muss man sich einig sein, bevor wir mit so einem Use Case da hinein gehen. Das scheint in diesem Kreis noch nicht abschließend geklärt zu sein. Das war unsere Anmerkungen zum Punkt 3, die strukturellen Roadblocker.

### **Charité / BIH**

Ich möchte zum Punkt 1 sagen, dass wir keinerlei Bedarf haben an zusätzlichen Fragestellungen mit hoher gesundheitlicher, gesellschaftlicher oder sozioökonomischer Relevanz. Die wurden vielfältig beleuchtet und da sind die Ideen sehr zahlreich. Die Frage ist, warum werden diese, die alle sehr hoch relevant und spannend sind, nicht angegangen, zumindest im deutschen Datenraum. Das liegt natürlich daran, dass die Datenquellen, die wir dafür brauchen sehr isoliert sind. Sie sind aus gesetzlichen, regulatorischen Gründen anlassfrei oder auch anlassbezogen, nicht oder nur sehr schwer miteinander verknüpfbar, wir brauchen Einverständniserklärungen von den individuellen Personen, deren Daten verknüpft werden sollen. Das geht in der Regel prospektiv, aber rückwärtsgewandt, wir wollen die Krankheitsverläufe in der Vergangenheit mit einbeziehen, ist das schlichtweg nicht möglich. Was brauchen wir dafür, um so etwas möglich zu machen? Wir brauchen erstmal einen nationalen Gesundheits-Identifizierer, damit wir überhaupt Personendaten miteinander verknüpfen können. Und als zweites würde ich gerne auf den Punkt einer zentralen Aufgabe des Dateninstituts zurückkommen. Dass sie hier als Datentreuhänder für Datenspende Aufgaben wahrnehmen kann. Diese sind getrieben durch individuelle Bürger und Bürgerinnen, die sich hier beteiligen. Entsprechende Daten kann eine vertrauensvolle Institution zusammenführen.

### **Bundesdruckerei GmbH**

Die Bundesdruckerei ist nicht in der COVID Forschung aktiv, wir sind jedoch technischer Enabler in vielen Projekten des Bundes. Wir sind beispielsweise während der COVID Pandemie im deutschen Impfstoff Monitoring oder mit den Datentreuhänderthemen, aber auch Datenhosting beauftragt gewesen. Wir sind mit dem RKI in vielen Vorhaben unterwegs, wo wir diese Daten für die weiteren Nutzung bereitstellen können, sehen uns also sowohl in dem Thema Datenteilen, also als technischer Enabler. Wir haben auch ein Produkt im Bereich Datentreuhänder und sind mit unserem Produkt CenTrust gerne bereit mitzuwirken. Wir sind in der Bundesdruckerei mit anderen Kompetenzen unterwegs auf sicheren Infrastrukturen und möchten uns hier gerne involvieren.

### **Open Government Netzwerk**

Aus Sicht des Dateninstituts ist es natürlich ein extrem herausfordernder und anspruchsvoller Use Case, er birgt aber auch viele Chancen. Beispielsweise gründen viele chronische Krankheiten auf Entzündungen, unerkannte Post-COVIDs könnten in Zukunft erkannt werden. Wir reden oft von symptomatischen Post-COVID, aber wer weiß, was langfristig noch in den Leuten schlummert. Auch was die individuellen medizinischen Behandlungen angeht, werden damit Grundlagen geschaffen. Dann haben wir, wie die NAKO Gesundheitsstudie und Philipps Universität Marburg es vorhin angesprochen haben, dieses Delta zwischen Sentinel Stichprobe und klinischer Gesamtschau. Dort geht es um Mustererkennung und zu schauen was der Typisierung dient. Und so komme ich auf das Thema digitaler Datenaltruismus, was ich dem digitalen Ehrenamt zu rechnen was wiederum dem Thema Open Government nahesteht. Dort sind unsere Chancen wo wir sagen können, wir haben jetzt durch die Pandemieerfahrung den Digitalansatz stark verfolgt, den können wir hier weiter gehen. Jetzt bewegen wir uns in der Königsdisziplin des Datasharings was Datenaltruismus angeht. Da gibt es vieles noch zu erarbeiten, zu erforschen, herauszufinden. Es gibt eine Verwandtschaft zu generellen Crowd Source Data, wie vom BMI in der IMAGI Studie auf Geodaten bezogen beauftragt wurde. Citizen Science spielt mit rein, auch das Thema My Data Concept, also selbst Hoheiten über seine Daten zu haben. Auch ständige Zugänge bis hin zum Thema Disposable ID, praktisch persönliche Identifikatoren auch

nachträglich einem Datensatz zu entziehen. Und der gesetzliche Rahmen ist noch nicht optimal. Winfried Veil hat dazu 2021 schon alles aufgeschrieben. Wo können wir anfangen mit transparenten Prozessen? Die Corona Warn App hat ja sehr gut gezeigt, wie ebenso ein Prozess zu einer starken Akzeptanz und zum Vertrauen führen kann. Datenaltruismus ist im Prinzip schon in der nationalen Kohorte vorhanden. Ich freue mich sehr, dass die hier anwesend sind. Man kann dort im Prinzip die Dinge weiterspinnen und gucken, wie wir in digitalen Datenaltruismus reinkommen können. Der objektorientierte Identifier ist natürlich wichtig, um die Kontexte schneller zu bekommen. Da ist das Thema FAIR (FDO) gerade auch dabei, faire digitale Objektivierung oder objektbezogene Infrastrukturen herzustellen und die Daten mit den Pathologen abzugleichen.

#### **NFDI4Health**

Ich vertrete die nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten und die Task Force COVID-19, die sich daraus ergeben hat. Ich möchte darauf eingehen, wie Ressourcen besser genutzt werden könnten. Wir möchten die Auffindbarkeit und Nachnutzung insbesondere von klinischen Studien, bevölkerungsbezogenen Studien erhöhen. Im Laufe von COVID hat es bei den bevölkerungsbezogenen Studien sehr gut funktioniert, zusätzliche Fragebögen über COVID 19 zu erstellen. Wir haben versucht, die zu harmonisieren, das war fast unmöglich. Da fände ich eine Koordinierungsstelle des Dateninstitut gut, um wichtige Fragen abzufassen, die dann im Konsens an die Studien verteilt werden. Es würde enorm helfen, wenn man nicht nachträglich harmonisieren müsste und würde Aufschluss geben über Lebensqualität und soziale Auswirkungen, wenn die abgefragt würden. Wir haben in der NFDI die Forschungsdateninfrastrukturen von verschiedenen Bereichen, zum Beispiel auch den Sozialwissenschaften. Das könnte man nutzen, um weitere Koordinationen durchzuführen. Wir arbeiten an einer zentralen Auffindbarkeit der Daten, um diese nutzen zu können. Im generellen Fall sind klinische Studien in sich relativ abgeschlossen, also wir haben eigentlich noch nicht die Position, dass die Daten nachträglich mit anderen Akteuren geteilt werden. Ein weiterer noch nicht erschlossener Bereich ist der Bereich unstrukturierter Daten der medizinischen Dokumentation. Aber auch die wissenschaftlichen Erkenntnisse, die in Literatur publiziert werden und auch unstrukturiert sind. In der Zeit der Pandemie haben die Preprints und auch die sozialen Medien eine hohe Rolle gespielt, um Informationen relativ schnell zu extrahieren. Diese wäre wahrscheinlich bei Auswirkungen von Long-COVID noch eine Quelle, die man berücksichtigen müsste. Zu Roadblocker ist schon vieles erwähnt worden. Nochmal die Auffindbarkeit, Datenschutz, Recordlinkage, Datenharmonisierung und allgemein die Möglichkeiten des Daten Sharing.

#### **Universitätsklinikum Köln / AGENS**

Wir beschäftigen uns seit 30 Jahren mit Routinedaten der Krankenkassen und auch anderer Sozialversicherungsvertreter. Letztendlich ist das Problem der Daten dieser Sozialversicherungsträger, unter anderem auch der Krankenkassen, dass die Kooperationsbereitschaft auf freiwilliger Basis passiert. Wir haben eine Kohorte gehabt, wir haben eine Einwilligungserklärungen gehabt, in der auch die Nutzung der Daten vom Patienten erlaubt worden ist. Trotzdem haben wir bei den Krankenkassen keine Kooperationsbereitschaft dafür finden können, obwohl diese Daten natürlich einen wichtigen Beitrag zum gesamten Verlauf retrospektiv beitragen. Leider sind wir vor verschlossenen Türen stehen geblieben. Wir versuchen es weiterhin, aber das ist gelebter Forschungsalltag, dass vorhandene Daten nicht verlinkbar sind, weil es an der Bereitschaft



der Träger letztendlich scheitert. Zum Unique Identifier wurde schon viel gesagt. Als Ergänzung, es muss dringend ein zentrales Register existieren, wo man von bestimmten Personen den Unique Identifier nachträglich erfragen und erfassen kann. Das gilt insbesondere für Datenquellen, die den Unique Identifier bisher noch nicht implementiert haben. Es gibt Datensammlungen von den Testungen, Simba n<sup>3</sup> hat vorhin davon gesprochen. Auch beim RKI gibt es eine riesengroße Sammlung von Daten über das Impfgeschehen, wir sind da in einem Projekt beteiligt, da ist auch die Bundesdruckerei mit im Spiel. Leider ist das Pseudonymisierungsverfahren so gewählt worden, dass eine Verlinkbarkeit nur unter größter Problematik stattfinden kann. Wir sind dabei das aufzulösen, sodass diese Datenbasis für Forschungsfragen genutzt werden kann.

### **Eileen Fuchs**

Damit sind Sie mit der ersten Fragerunde durch. Ich danke Ihnen vielmals für Ihre Beiträge. Das war sehr dicht, sehr viel Info und für uns sehr aufschlussreich. Es wird alles aufgezeichnet und wir werden ein Transkript erstellen und ein untertiteltes Video online stellen, insofern geht nichts von Ihren Beiträgen verloren. Es kam die Frage auf, wie eigentlich die Rednerliste gestaltet wurde. Das war eine zufällige Auslosung, nicht dass Sie denken, wir hätten irgendjemand bevorzugt. Das ist im Kontext der Wettbewerbsgleichheit natürlich wichtig, genauso wie die limitierte Zeit für Ihre Beiträge. Gut, wir haben jetzt 15 Minuten Pause. Wir freuen uns, Sie dann wiederzusehen. Dann würden wir in einen zweiten Komplex einsteigen. Ich kommuniziere Ihnen schon mal, dann können Sie sich Gedanken machen. Wir würden wissen wollen, was für Use Cases Sie vielleicht schon im Kopf haben, die für ein Dateninstitut interessant sein könnten und welche Voraussetzungen an einen Use Case gekoppelt wäre. Falls Sie Ideen haben, nach der Pause gerne, dann entsprechend der Rednerliste in der gleichen Reihenfolge, wie wir sie gerade durchgegangen sind

[Pause]

### **Christina Schmidt-Holtmann**

Meine sehr verehrten Damen und Herren, wir begrüßen Sie zu Teil 2 des Marktdialogs zum Aufbau des Dateninstitutes Modul 1. Und wie vor der Pause schon anmoderiert, würden wir jetzt gerne mit ihnen in die nächsten Fragestellungen gehen, die ihnen unsere Kollegin Lena Müller jetzt nochmal detaillierter vorstellen wird.

### **Lena Müller**

Zunächst würden wir gerne von ihnen hören, ob Sie schon konkretere Ideen für Use Cases haben. Diese muss auch nicht konkret sein, aber ob ihnen etwas in den Sinn kommt, was als Pilot Use Cases für das Dateninstitut machbar wäre. Ob das eine Datenerhebung, eine Anwendung mit Plattform oder Formate sind. Welche Themen dabei bearbeitet werden könnten. Zum Kontext nochmals die Erinnerung, diese Use Cases sollen in etwa einem Jahr umgesetzt und möglichst gemeinwohlorientiert entwickelt werden und Nachnutzung soll möglich sein. Zusätzlich wäre für uns interessant zu wissen, ob es aus ihrer Sicht Voraussetzungen gibt beispielsweise für die Teilnahme an der Challenge oder für die Umsetzung eines Use Cases. Die Redezeit ist auf anderthalb Minuten begrenzt

### **Christina Schmidt-Holtmann**

Bekanntes und bewährtes Verfahren aus der ersten Runde, Sie kennen das. Da sich einige schon verabschiedet haben, liegen wir gut in der Zeit und starten dann.

### **Charité - Koordinierungsstelle Netzwerk Universitätsmedizin**

Aufbauend auf das, was ich vorher schon sagte, schlage ich vor, dass man versucht, diese bereits vorhandenen, sehr detaillierten Patientenverläufe in Kliniken, also die tiefe klinische Perspektive, mit erfassten Kohortendaten und dem außerklinischen Bereich, wie Daten der Renten- oder Krankenversicherung, zu verzahnen. Wenn man diese drei Datenarten zusammenbrächte, wäre das ein großer Fortschritt für die Forschung. Das halte ich in einem Jahr nicht für völlig unrealistisch, zumal es bereits bezogen auf die Anbindung von Krankenkassendaten Vorarbeiten und Aktivitäten gibt, auf die man aufsetzen kann. Das kann man mit zwei Zielen machen. Das erste ist eben, dass man aus der Klinik heraus einen Eindruck bekommt: Was passiert eigentlich mit unseren Patientinnen, nachdem Sie das Krankenhaus verlassen haben. Und zweitens, umgekehrt kann man in der Gesamtpopulation Gruppen identifizieren, bei denen sich eine vertiefte klinische Betrachtung, eine klinische Studie oder ähnliches, empfehlen würde, die man sonst für die Forschung nie erreichen würde. Also beide Zielrichtungen sollte man im Kopf haben, wenn man so einen Datenschatz kreiert.

Welche Voraussetzungen müssten erfüllt sein? Aus meiner Sicht wäre wichtig, dass man nicht nur einen Use Case baut, der mit viel Aufwand und Engagement einzelner einmal funktioniert, sondern von vornherein das Thema Skalierung, Institutionalisierung, Plattformfähigkeiten mitdenkt. Damit nicht nur eine Forschungsfrage individuell beantwortet werden kann, sondern diese Daten zukünftig systematisch für viele Forschungsfragen, die wir heute noch gar nicht kennen, möglichst niedrigschwellig zugänglich gemacht werden. Diesen Institutionalisierungsaspekt halte ich für extrem wichtig. Es sollten Akteure in den Lead gebracht werden, die eine solche Institutionalisierung verantworten und organisieren können.

### **Medizininformatik Initiative**

Ich komme nicht mit einer konkreten Fragestellung. Ich finde das Themenfeld Long- und Post-COVID sehr geeignet, weil es bereits wichtige etablierte Strukturen gibt. Es wurden in den ersten Redebeiträgen länglich Hürden und Lücken in der Erschließung und Verknüpfung von Daten und Datenquellen genannt, die fokussiert werden könnten. Ich fände es wichtig, dass dieser Use Case genutzt wird, um zu demonstrieren, wie man diese bestehenden, bereits aufgebauten Strukturen nutzen kann und gemeinsam mit den hier wichtigen und relevanten Stakeholdern Prozesse für eine übergreifende Datenverwertung entwickelt und abgestimmt die existierenden Hürden adressiert. Das sollte an diesem Use Case exemplarisch gezeigt werden, auf dem Weg hin zu einem späteren gesamten Dateninstitut.

### **Universität Leipzig – Medical Data Science**

Ich würde mich da dem Netzwerk Universitätsmedizin und der Medizininformatik Initiative anschließen. Wir sollten aufgrund der Kürze der Zeit Datenquellen nutzen, die besonders gut erschlossen sind. Das sind die Routinedaten über die Medizininformatik Initiative beziehungsweise NUM, die Krankenkassendaten und die Kohortendaten und da lassen sich verschiedene Themenstellungen finden. Wir würden uns für Therapieoptionen für bestimmte Konzeptionen der Patienten interessieren, für so einen Use Case lassen sich verschiedene Teilnehmer finden. Und wichtig: Bitte nicht einmalig, sondern es muss eine Plattform sein, so dass das routinemäßig durchgeführt werden kann.

### **DESAM ForNet**

Ich würde das gleich aufgreifen, was Universität Leipzig und Netzwerk Universitätsmedizin gesagt haben. Wir haben in der DESAM ForNet Initiative im Bereich der niedergelassenen Mediziner genau das realisiert. Wir haben eine Plattform, mit der wir Daten aus der Routine mit medizinischen Fragestellungen verknüpfen können. Und was wir festgestellt haben, ist, dass die Routinedaten oft nicht ausreichen, um eine Fragestellung wie den COVID Use Case ausführlich zu beantworten. Was wir über diese DESAM ForNet Plattform einbringen könnten wäre die Infrastruktur, mit der klinischen Studien, klinische Fragestellungen mit gezielten Studienfragestellungen in Arztpraxen bearbeitet werden und mit den Routinedaten die in der Arztpraxis zur Verfügung sind angereichert werden können. Wir mussten uns an der Realität in den Arztpraxen orientieren, was Vollständigkeit und Zuverlässigkeit der Dokumentation betrifft. Diese Infrastruktur auszuprobieren um zu verstehen, wie das Krankheitsgeschehen im Long-COVID Bereich läuft, was ja gerade in Hausarztpraxen aufschlägt. Da könnten wir einen entsprechenden Beitrag leisten, weil das eine bundesweite Initiative ist mit Netzen, die in mehreren Bundesländern etabliert werden und Daten, die zusammengeführt werden können. Die werden so aufbereitet, dass sie übergeordnet bereitgestellt werden können. Zum Beispiel im Rahmen von NUM oder der Medizininformatik Initiative.

#### **Ernst & Young GmbH**

Es wurden in der ersten Session bereits viele Themenfelder genannt, die alle sehr relevant sind. Ich denke, dass der Zeitrahmen den Rahmen für die Machbarkeit setzen wird und es ist wichtig, die Machbarkeit des Use Cases gut zu definieren und sicherzustellen, dass die Umsetzung innerhalb von einem Jahr möglich ist. Viele Anwendungsfelder im Data Science Bereich sind in der Umsetzung schwierig und daher müssen die Themen, die genannt wurden, Datenschutz, Ethik, frühzeitig geklärt werden und auch das Thema der Nutzung für die Endanwender von Datenprodukten. Es ist sehr wichtig, dass diese Gruppen von Anfang an eng mit einbezogen werden, sowohl in der Definition von dem Use Case als auch hinsichtlich der Möglichkeiten regelmäßig Feedback zu geben. Es sollten verschiedene Austauschformate vorgesehen werden. Wichtig für das Thema Machbarkeit und Umsetzung sind auch Besonderheiten in einer KI gestützten Umsetzung. Insbesondere das Thema Trainingsdaten, technische Infrastruktur und Schnelligkeit der Umsetzung. Hinsichtlich des Themas gab es heute Vormittag eine Reihe von sehr relevanten Möglichkeiten und Zukunftsthemen.

#### **NAKO Gesundheitsstudie**

Inhaltlich schließe ich mich meinen Vorrednerinnen vollkommen an. Wir sollten die Quellen, die wir haben, poolen und nutzen. Um die zweite Frage aufzugreifen: Um das Ganze in dem angegebenen Zeitrahmen umzusetzen, sollten bei verschiedenen Datenquellen ein gestreamlineter Use and Access Prozess frühzeitig etabliert werden, um den Zugang zu diesen Daten zeitnah zu ermöglichen und die Beantragung und Herausgabe nicht unnötig zu verzögern.

#### **Materna Information & Communications SE**

Mir also mir wäre wichtig zu betonen, dass die entstehende Methodik, mit diesen vielen wertvollen Akteuren, auch speziell ist. Man muss immer wieder drauf schauen, was man hier in dieser wichtigen Rolle als Think- & Do-Tank, als der man sich etablieren will, für andere Themen Gebiete erschließen wird. Das identifizieren der wichtigen Akteure, Sie haben hier eine ganze Reihe mit eingeladen, ist sehr wichtig. Die Methodik wie man vorgeht, damit man

das für andere Aufgabengebiete, beispielsweise Nachhaltigkeit, Ableitungen treffen kann. Dann die Identifikation der relevanten Formate, sowohl Austauschformate, die man schaffen möchte, als auch Plattformen und wie man mit vorhandenen Assets umgehen will. Was hier aus dem Vormittag herauskam ist, dass es schon sehr viele Plattformen gibt. Das Dateninstitut hat die Rolle, Dinge zu ermöglichen und Lösungen zu schaffen, wenn es sie noch nicht gibt. Rechtsrahmen, organisatorischen Rahmen aber auch technischen Rahmen und dass diese Aufgaben dann immer wieder neutralisiert werden müssen. Aus dem Gesundheitssystem heraus dann wieder in andere Aufgaben, Zyklen, Verkehr oder Umweltschutz, was auch immer für Themen uns umtreiben.

### **McKinsey & Company, Inc.**

Ich möchte mich auch meinen Vorrednern anschließen, insbesondere was Datenquellen und Elemente für die Nachnutzbarkeit, Standardisierung, Skalierung des Use Cases selbst angeht. Das kann man nicht überbetonen, der Kollege der Universität Leipzig hatte auch schon einige valide Datenquellen genannt. Was hier darüber hinaus noch sehr wichtig ist, ist, dass es dabei nicht nur um die Umsetzung eines Anwendungsfalls geht, der nachnutzbar sein soll, sondern auch darum, dass das Dateninstitut davon möglichst viel lernen soll und diese Learnings dann auf andere Bereiche übertragen werden können. Es ist wichtig einen Anwendungsfall zu nehmen, der zwar im Laufe der Zeit umgesetzt werden kann und zu wertvollen Ergebnissen in sich selbst führt, aber nicht unterkomplex ist, sondern eine gewisse Komplexität mit sich bringt, wo man in gewisse Hürden reinläuft, die dann wiederum dazu führen, dass man Elemente identifiziert, wo das Dateninstitut dann im Endeffekt beitragen könnte, diese für zukünftige Fälle zu lösen. Deswegen würde ich dafür plädieren, dass man keinen Anwendungsfall der zu Straight Forward ist auswählt, sondern etwas, das durchaus gewisse Komplexitäten, Schnittstellen und so weiter mit sich bringt. Um nicht nur den Anwendungsfall selbst nachnutzbar zu gestalten, sondern auch das, was man daraus lernt nachnutzbar für andere Fälle und als übertragbar dahin betrachtet.

### **Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin**

Ich schließe mich meinen Vorrednern bezüglich der Datenquellen vollständig an. Aber ich bin mal ein bisschen provokativ und würde vorschlagen, über den Scope und über den eigentlichen Use Case nachzudenken. Wie schon eingangs gesagt, wir können Post-COVID als Erkrankung schwer beschreiben und haben sehr viele Störfaktoren. Dazu kommt, dass es natürlich gut ist die Datenquellen in einem Dateninstitut zu verknüpfen, aber ganz ehrlich, COVID und Post-COVID-Daten liegen quasi überall auf der Straße, viel mehr als man auswerten kann. Wir haben europäische und internationale Datenbanken, wir haben in Deutschland Datenbanken, die frei zur Verfügung stehen, heute schon beforscht werden können. Da ist eher die Auswertung als die Datenverfügbarkeit momentan der Bottleneck, zumindest in einigen der Forschungsbereiche. Wir könnten den Mehrwert eines Dateninstitutes besser zeigen, wenn wir das thematisch mehr ausweiten. Wenn wir ein bisschen zurück in die Vergangenheit gehen und versuchen eine Datensammlung zu initiieren, die vielleicht 2017 ihren Anfang nimmt. Nehmen dann die NAKO Daten rein, die Versichertendaten, die Versorgungsdaten aus den Datenintegrationszentren und sagen, wir haben hier eine Datensammlung zur nationalen Gesundheit in Pandemiezeiten. Dann können wir uns dort Post-COVID anschauen, wir können uns aber auch Post-Vac anschauen, wir können uns anschauen wie sich Homeoffice auf unsere Gesundheit auswirkt und viele andere Sachen, die sich in den letzten drei Jahren gesellschaftlich massiv verändert haben und komplexe Auswirkungen auf die Gesundheit haben. Wir haben dann nicht mehr dieses

vielleicht politisch attraktive Post-COVID-Label. Aber wir könnten wirklich Forschung ermöglichen, die ohne ein Dateninstitut nicht möglich wäre und vielleicht Antworten finden, die uns momentan dringend fehlen.

#### **Charité – Institut für Medizinische Soziologie und Rehabilitationswissenschaften**

Ich möchte mich im Prinzip gleich der Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin anschließen, ich finde das ist sehr guter Input gewesen und eine ganz spannende Sache, um den Mehrwert eines größer gedachten Datenzentrums zu zeigen. Bestärken möchte ich die Rentenversicherungsdaten, wo es um Arbeitsunfähigkeit geht, Frühverrentung. Das ist besonders bei Post-COVID relevant in Ergänzung zu den Krankenkassendaten und den Versorgungsverläufen. Das besonders Spannende ist, dass man die Versorgungsverläufe hat. Von wem wird welche Diagnose zu welchem Zeitpunkt gegeben, wann tritt sowas wie eine Post-COVID Diagnose auf, was gibt es noch als alternative Diagnosen. Da hat ja auch der Kollege der Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin schon ein bisschen exploriert, dass dieses Krankheitsbild gar nicht so klar umrissen ist. Fatigue oder andere Sachen sind dort mitzudenken. Es wäre spannend eine breitere Basis zu haben, mit der man mit Post-COVID zusammenhängende Fragestellungen bearbeiten könnte. Das wäre ein wirklich großer Mehrwert.

#### **Philipps Universität Marburg**

Ich möchte nochmal den Phänotypen-Gedanken reinbringen und das würde sich auch mit dem decken, was der Kollege der Nationalen Pandemiekohorte im Netzwerk Universitätsmedizin schon gesagt hat. Dass man nach Biomarkern sucht und versucht spezifischer in die klinische Beschreibung der unterschiedlichen Long-COVID Typen zu gehen, dann kann man daraus die Therapien ableiten. Ich glaube, das ist eigentlich der einzig richtige Zugangsweg, um diesen Wust von Long-COVID Patienten eben auch besser zu diagnostizieren und zu behandeln. Wir müssen die Psyche und die Somatik dort zusammenbringen und diagnostischen und therapeutischen schauen, ob hier auch Hybridkonzepte eine Rolle spielen können.

#### **Deutsche Rentenversicherung Bund**

Ich möchte die Vorrednerinnen und Vorredner bezüglich des möglichen Zusammenspiels von Daten aus SGB V oder VI zu Forschungszwecken unterstützen und möchte auf den im Chat bereits genannten, bislang unveröffentlichten Bericht, verweisen. Das BMG hat dazu eine Art Machbarkeit Studie in Auftrag gegeben, da sind wir relativ weit. Man könnte das in einem Förderschwerpunkt oder Forschungsschwerpunkt versuchen zu kreieren, um es auch für die Zukunft in die Regel zu bringen. Und ich möchte eine Lanze brechen, das über Post-COVID hinaus zu tun. Insbesondere chronische Erkrankungen haben eine gesellschaftliche Relevanz. Ich denke da an psychische Erkrankungen, ich könnte mir das auch bei bestimmten Krebsbildern vorstellen. Langfristige Verläufe auf Basis dieser Daten zu analysieren und insbesondere Möglichkeiten in Studien zu prüfen, ob eine Verknüpfung besteht zu wiederkehrenden Befragungen, zu Kohortenstudien. Das könnte ein guter Gewinn für die Versorgungssituation oder die Analyse der Versorgungssituation in Deutschland sein und wir als Rentenversicherung stehen zur Verfügung dafür.

#### **TMF e.V.**

Die Grundsatzfrage ist nicht, wie kann man eine ideale Kohorte zur Post-COVID Fragestellung machen. Dafür braucht es kein Dateninstitut, da sind genug Experten auch hier in der Runde,

die das können. Die spannende Frage ist, und das hat der Kollege von McKinsey angesprochen, an was für einem Use Case kann das Dateninstitut lernen, welchen Beitrag das Dateninstitut dazu leisten kann, dass wir jenseits der Datenerhebung und -auswertung, die im Wesentlichen eine Ressourcenfrage ist, die Datenverfügbarkeit und -verknüpfung verbessern. Und da sind wir bei den Datenquellen. Also Rentenversicherungsdaten, Sozialversicherungsdaten, primäre Versorgungsdaten stationärer Sektor, primäre Versorgung ambulanter Sektor und der ist sehr vielfältig. Das sind die Punkte, die hier adressiert werden müssen. Und deshalb ist die Frage der Akteurslandschaft oder der Plattform ganz wichtig. Wir haben im Gesundheitswesen viel unterwegs und hier geht es für den Use Case darum, eine intelligente Verknüpfung auch den Akteuren und der Expertisen, die es gibt zu schaffen. Es sind im Chat mehrfach diese insgesamt vier Gutachten zu dem Thema genannt, die teils veröffentlicht, teils unveröffentlicht sind. Dass man die aufarbeitet. Und natürlich muss man sich am Ende an die regulative Frage ran trauen, denn es gibt gerade bezüglich Identifikation und Verknüpfung eine große regulatorische Hürde, nicht nur technisch. Das ist ein Punkt, der kann am Ende eines Use Cases als Lessons Learned oder Empfehlung an die Weiterentwicklung des Rahmens herauskommen. Dann kriegt man raus, an welcher Stelle das Dateninstitut eine geeignete Ergänzung, eine geeignete Hilfestellung für diesen Forschungssektor erbringen kann. Das ist das Wesentliche, was aus diesem Use Case herauskommen müsste.

#### **Labvantage - Biomax GmbH**

Ich würde gerne drei Use Cases vorschlagen, anknüpfend an die erste Runde. Erstens, eine Sache, die wir auch leisten können, einen zentralen Prozessierungsservice für neuroimaging Daten in diesem Bereich, um hier einen Blindspot zu adressieren und eine interpretierbare Datenbasis zu haben im Vergleich zu ca. 1000 Referenzgehirnen. Ein zweiter Use Case wäre ein Harmonisierungsservice durch eine Semantic Layer, der unterschiedliche Datenressourcen miteinander verknüpfbar und auswertbar macht und damit einen unmittelbaren Mehrwert hat. Und der so ausgestaltet sein sollte, dass man es auf andere Non Post-COVID Themen erweitern kann. Und drittens, um etwas für die Mediziner zu liefern, die ja wesentliche Datenbereitsteller sind, eine Art Cockpit, welches mir erlaubt die Daten des Patienten im Kontext der bereits erfassten und verfügbaren Daten auch longitudinal zugänglich zu machen. So dass ich sehen kann: Wie sieht mein Patient im Verhältnis zu den bereits gesehenen Patienten und den jeweiligen Therapieerfolgen aus?

#### **Open Government Netzwerk Deutschland**

Also naturgemäß interessiert uns besonders das Thema Datenaltruismus. Was sich daraus erwächst steht im Data Governance Act und hat Bezug zu Krisenresilienz, zur Selbstwirksamkeitserkenntnis. Auch was McKinsey ansprach, die Übertragbarkeit. Im besten Falle schaut man da, wo schon Altruismus existiert, das wäre bei der nationalen Kohorte. Ich will jetzt nicht vorgreifen, aber ich fände das einen Ansatzpunkt zu schauen, wie wir im digitalen Datenaltruismus weiterkommen und datenrechtliche Hemmnisse bearbeiten. So wie sie Winfried Veil gelayoutet oder aufgebracht hat. Inwiefern zum Beispiel aus den Sozialwissenschaften das Five Safes Prinzip anwendbar ist in dem Bereich und inwiefern sich die Anwendungen, die bereits von RKI und Paul Ehrlich existieren, sich nochmal heranziehen lassen um einen Demonstrator für Altruismus zu bauen. Das Thema objektorientierte Identifier ist wunderbar, ich weiß aber nicht, ob sowas in einem Jahr drin ist. Aber dieses FTO Forum wird wohl demnächst mit eigener Finanzierung ausgestattet sein. Insofern

könnte das ein interessanter Partner sein, um in diesem speziellen Thema die Objekte Indikatoren auszuarbeiten. Das wären so die Dimensionen.

### **Universitätsklinikum Köln / AGENS**

Als letzter in der Runde kann ich sagen, es ist fast alles gesagt worden. Breite Fragestellungen finde ich sehr gut, eine breite Datenbasis finde ich auch sehr gut. Das heißt also alle Datenbasen die bekannt sind erstmal einzubeziehen, unabhängig davon, ob wir davon überzeugt sind, dass sie etwas beitragen können oder nicht. Ich würde die Datenbank von Testungen, die Simba n<sup>3</sup> Datenlösungen vorhin vorgestellt hat, mit einbeziehen und darauf verweisen, dass sie nicht personenbezogen verlinkbar sind. Möglicherweise ergibt das beim nächsten Mal, wenn man sowas aufbaut, dann den Hinweis: „Bitte prüft auch die Verlinkbarkeit von solchen Datenbanken“. Und ich muss noch darauf hinweisen, die beim RKI befindlichen Impfdaten sind verlinkbar und sollten miteinbezogen werden und können dann entsprechend in der Fragestellung berücksichtigt werden.

### **Eileen Fuchs**

Ja, vielen Lieben Dank für Ihren Input, sehr dicht und für uns sehr aufschlussreich. Wie machen wir jetzt weiter mit dem Modul 1? Wir werden auf der Basis Ihrer Eingaben und dem, was wir uns sonst an Wissen und Hintergrund aneignen, einen Use Case präzisieren und ausarbeiten, den wir dann mit den Ressorts diskutieren werden und abstimmen werden. Wir brauchen natürlich ein gemeinsames Verständnis, wie wir mit dem Dateninstitut im Modul 1 antreten innerhalb der Bundesregierung. Und wir hoffen, dass wir jetzt genug Material haben, um uns eine Meinung zu bilden, ob es im Bereich Post-COVID einen Use Case geben kann, der realisierbar ist entsprechend der in dem Konzept genannten Maßgaben. Wenn das so kommt, wird schnellstmöglich eine Ausschreibung in Form einer Challenge, wie Frau Schmidt-Holtmann am Anfang erläutert hat, stattfinden. Für diese Challenge können sich verschiedene Teams bewerben. Das weitere Prozedere wird noch im Einzelnen bekannt gegeben. Das ist erstmal das Modul 1 um das es heute ging. Vielleicht zu Ihrer Info, was die weiteren Module angeht: Modul 2 hatte Frau Schmidt-Holtmann anfangs auch beschrieben, da geht es um den Use Case aus dem Energiebereich. Der wird durch die Deutsche Energieagentur, die dena, als inhouse Vergabe seitens des BMWK umgesetzt. Das Modul 3, da gab es ja schon den Marktdialog, da geht es um die Frage, wie sieht das Dateninstitut aus und da ist natürlich ganz maßgeblich, was in Modulen 1 und 2 passiert. Welche Ableitungen können getroffen werden aus den Use Cases in Modulen 1 und 2 und wie beeinflusst das die spätere Tätigkeit des Dateninstituts? Denn in der Tat, es geht um einen Think & Do-Tank. Wir wollen eine möglichst agile Organisation aufbauen, die sich sukzessive immer unabhängiger entwickelt. Da sind die Ableitungen dessen, was in den Use Cases gemacht wird, essentiell. Insofern ist Ihre Zulieferung und das, was sie uns reingegeben haben, für uns sehr wichtig. Wir bedanken uns herzlich und würden dann so schnell wie möglich eine Veröffentlichung des Marktdialoges und der Materialien zur Verfügung stellen, sind da schon in der Vorbereitung und würden diese Ihnen nochmal zukommen lassen, wenn sie einverstanden sind. Aber ich habe jetzt noch kein gegenteiliges Votum gehört. Dann wäre es das von unserer Seite. Wir würden Sie in eine verdiente Mittagspause entlassen, es sei denn, Sie haben jetzt noch dringende Fragen, die Sie an uns loswerden wollen im Hinblick auf die weiteren nächsten Schritte oder das, was wir heute besprochen haben. Das scheint derzeit nicht der Fall zu sein. Dann bedanke ich mich für Ihre Teilnahme.

**Christina Schmidt-Holtmann**

Dem bleibt mir nichts hinzuzufügen. Vielen Dank für Ihre konstruktiven und sehr disziplinierten Beiträge im Zeitrahmen und wir freuen uns, den einen oder die andere im weiteren Verlauf wiedersehen zu dürfen, so will ich es mal formulieren, Dankeschön.