

## Antwort

### der Bundesregierung

**auf die Kleine Anfrage der Abgeordneten Dr. Wieland Schinnenburg, Michael Theurer, Jens Beeck, weiterer Abgeordneter und der Fraktion der FDP – Drucksache 19/21959 –**

### Nutzen der Corona-Datenspende-App zur Bekämpfung von Corona

#### Vorbemerkung der Fragesteller

Seit Anfang April 2020 ist die Corona-Datenspende-App des Robert-Koch-Instituts (RKI) im App-Store von Apple verfügbar (<https://apps.apple.com/de/app/corona-datenspende/id1504705422>), ebenso existiert eine Version für Android-Geräte (<https://play.google.com/store/apps/details?id=de.rki.coronadatenspende&hl=de>).

Im Corona-Datenspende-Blog des RKI wird angegeben, dass mehr als 500 000 Nutzer die App heruntergeladen hätten und mehr als 70 Millionen Datensätze gesendet worden wären (<https://corona-datenspende.de/science/>). In dem Blog sind etwa Auswertungen über Pulsfrequenzen, Schrittzähler, Effekte des Wetters auf die Aktivität der Datenspender und der Körpertemperatur aufgeführt.

Bislang haben die Analysen der App-Daten allerdings für die Fragestellerinnen und Fragesteller keinen erkennbaren Beitrag zur Bekämpfung der Corona-Pandemie leisten können. So bereitet die Analyse der Fieberkurven Probleme (<https://corona-datenspende.de/science/reports/fevercurve/>), zudem ist nach Ansicht der Fragestellerinnen und Fragesteller unklar, wie bei der geplanten Fieberkarte Corona-Fälle von anderen Fällen unterschieden werden sollen.

#### Vorbemerkung der Bundesregierung

Nach Auffassung der Bundesregierung ist es mit der Corona-Datenspende-App gelungen, ein Instrument zu programmieren, das es erlaubt, bevölkerungsweit verschiedenste Gesundheitsdaten und Vitalwerte unter datenschutzrechtlich guten Bedingungen zusammenzutragen. Inzwischen spenden über 500.000 Personen täglich oder mit Unterbrechungen Daten, die von den unterschiedlichsten sogenannten Wearables gemessen werden. Durch die Auswertung der pseudonymisierten Daten – auch mit Hilfe Künstlicher Intelligenz – entsteht für die weit verbreiteten smarten Geräte ein überindividueller Nutzen, der im Idealfall Hinweise auf ein zunehmendes Ausbruchsgeschehen erlaubt.

Datenspenderinnen und Datenspender können zudem im Blog, das für die Corona-Datenspende-App angeboten wird, nachverfolgen, welche Aussagen sich aus den gesendeten Daten nach der Auswertung gewinnen lassen. Die hohe Zahl von Personen, die dauerhaft die Daten ihrer Wearables zur Verfügung stellen, ist ein klarer Indikator dafür, dass das Konzept der App nachvollzogen wird.

1. Welche Kosten sind für die Corona-Datenspende-App bisher angefallen?
  - a) Welche Kosten sind für die Programmierung der Apps für iOS und Android jeweils angefallen?
  - b) Welche Kosten sind für den Betrieb der Apps für iOS und Android jeweils angefallen, und mit welchen weiteren Kosten rechnet die Bundesregierung?

Initial sind für die Entwicklung der iOS-App für den Zustand „verschlüsselt und getestet“ Kosten in Höhe von 10.500 Euro (Netto) entstanden. Für die Entwicklung der Android-App in diesem Zustand wurden Kosten in Höhe von 15.750 Euro (Netto) veranschlagt. Hinzu kommen Kosten für die weitere technische Umsetzung, Datenschutz- und Datensicherheit sowie für die Aufbereitung von Informationsmaterialien. Insgesamt sind so Kosten für Einmalleistungen in Höhe von 95.200 Euro (Netto) entstanden. Hinzu kommen, wie vertraglich vorgesehen, Aufstockungen für ergänzende Leistungen (technische Umsetzung, Datenschutz- und Datensicherheit, Kommunikation) in Höhe von 38.200 Euro (Netto).

Für den Betrieb der App entstehen monatliche Kosten in Höhe von 81.000 Euro (Netto). Diese setzen sich aus Serverkosten (45.000 Euro Hardware, Housing, Monitoring, Lizenz) und Support und Maintenance zusammen (36.000 Euro).

2. Welchen Beitrag hat die Corona-Datenspende-App bisher zur Bekämpfung des Corona-Virus geleistet?

Die Corona-Datenspende-App und die damit zusammenhängende wissenschaftliche Auswertung der gesendeten Massendaten ermöglicht es mittlerweile tagesaktuell in einer Stichprobe von mehr als 500.000 Datenspenderinnen und Datenspendern mit hoher Genauigkeit Fiebersymptomatik zu erkennen. Die Aggregation der Daten ermöglicht somit den zeitlichen Verlauf der Fieberinzidenz in der Gesamtpopulation oder z. B. auf Landkreisebene zu schätzen. Da Fieber eine Kernsymptomatik der Covid-19 Erkrankung sein kann, werden überproportionale Zuwächse in den Covid-19 Infektionen über die Corona-Datenspende früher als über das Meldesystem erkannt. Durch die vollständige Automatisierung der Prozesskette Datenaggregation, -bereinigung, -voranalyse, -analyse, und -visualisierung ist die Corona-Datenspende das weltweit erste vollautomatisierte Gesundheitsmonitoringsystem auf Basis des Citizen-Science-Prinzips und in dieser Größenordnung, das in Realzeit für die Covid-19-Pandemie ein wichtiges, quantitatives prognostisches Frühwarnsystem liefert, insbesondere in Kombination mit anderen Surveillancesystemen, die am Robert Koch-Institut (RKI) betrieben werden. Die bisher geleisteten Fortschritte sind detailliert im Blog des Projekts beschrieben (<https://www.corona-datenspende.de/science>), für die o. g. Frage sind die Informationen im Bericht „COVID-19 & Fiebertmessungen“ (<https://corona-datenspende.de/science/reports/c19detect/>) relevant.

3. Welchen Beitrag soll die Corona-Datenspende-App in Zukunft zur Bekämpfung des Coronavirus leisten?
  - a) Welche Daten sollen besonders zur Corona-Bekämpfung genutzt werden?
  - b) Wie sollen Fieberdaten genutzt werden, und wie kann sichergestellt werden, dass erfasste Fieberdaten in einem Zusammenhang mit Corona stehen?

Die Corona-Datenspende-App hat das Potential, im weiteren Verlauf der COVID-19-Pandemie in verschiedenen Bereichen Informationen zu liefern, die zu einer verbesserten Lagebeurteilung beitragen können. So „läuft“ seit einigen Wochen die automatisierte Fieberdetektion bundesweit und auf Landesebene stabil; es wurde hierfür der täglich aktualisierte Fiebermonitor entwickelt (<https://corona-datenspende.de/science/monitor/>). Das Monitor-System wird weiterentwickelt, um noch präziser und hochaufgelöster Informationen zu liefern. Außerdem werden momentan neben den statistischen Verfahren, auf denen der Detektionsprozess beruht, Methoden aus dem Bereich der Mustererkennung und des maschinellen Lernens implementiert, die basierend auf publizierten Ergebnissen die Detektionsqualität noch erhöhen kann. Zweitens wird gerade ein Programm-Code implementiert, der als positiv detektierte Fieberfälle mit noch höherer zeitlicher Auflösung (15 Minuten statt 1 Tag) die Zeitreihen klassifiziert, um ggf. COVID-19-spezifische Anomalien in den Kurven zu identifizieren.

4. Wie werden die Messtoleranzen der unterschiedlichen Datenspenden durch unterschiedliche Messgeräte in der Datenanalyse berücksichtigt?

Bei der Corona-Datenspende-App kommen sehr verschiedene Wearables, Fitnessstracker und Smartwatches zum Einsatz. Obwohl diese Geräte „im Prinzip“ ähnliche Daten zum Ruhepuls und zur täglichen Schrittzahl oder zum Schlafrhythmus erfassen, gibt es deutliche Unterschiede im Nutzungsrhythmus, der Fragmentierung der Daten und statistischen Anomalien, die die verschiedenen Geräte aufweisen. Z. B. werden Applewatches nahezu nie nachts getragen. Ein ganz wichtiger und sehr zeitaufwendiger Bestandteil der wissenschaftlichen Auswertung ist die Berechnung individueller Baselines, also der Normalwerte individueller Spenderinnen und Spender und die Berücksichtigung statistischer Verschiebungen und Verzerrungen, die nicht nur durch die unterschiedlichen Geräte, sondern auch durch sozioökonomische, geographische und klimabedingte Einflüsse auftreten. Diese Einflüsse werden in den Analysen berücksichtigt. Hierzu wurden einige erklärende Blogbeiträge publiziert. Der wichtigste Punkt zum Thema „Messung“ ist die Fragmentierung individueller Datenströme. Da nur die allerwenigsten Spenderinnen und Spender jeden Tag und ohne Unterbrechung Daten spenden, sind die meisten Zeitreihen lückenhaft. Daher wurde ein Algorithmus entwickelt, der quantitativ bestimmt, bei welcher Fragmentierung noch valide Detektionen berechnet werden können. Die Ergebnisse zu dieser Studie werden demnächst in einer renommierten wissenschaftlichen Zeitschrift publiziert.

5. Welchen Beitrag leisten Daten von Schrittzählern oder Pulsmessern zur Corona-Bekämpfung insbesondere vor dem Hintergrund, dass viele Corona-Infektionen ohne oder nur mit geringen Symptomen verlaufen?

Die Corona-Datenspende und die wissenschaftliche Analyse ist ein populationsbezogenes Werkzeug. Wenn in der Referenzpopulation die COVID-19-Erkrankungen zunehmen, dann wird, selbst wenn die Mehrzahl der Infektionen

fieberfrei verlaufen, es dennoch zu einem Anstieg der Fieberinzidenz kommen und damit ein Signal erfasst, dass die Fallzahlen ansteigen.

6. Welche Anzahl an Datenspendern hat in den einzelnen Monaten des Betriebs der App welche Menge an Datensätzen gespendet?

Momentan sind etwas mehr als 500.000 Datenspenderinnen und Datenspender aktiv. Aktiv heißt in diesem Zusammenhang für die wissenschaftliche Auswertung, dass diese Datensätze zur Auswertung beitragen.

Für die Fieberanalyse sind wenig fragmentierte Zeitreihen am wertvollsten und liefern die besten Ergebnisse. Ebenso gilt: je länger die Zeitreihen, desto besser können die Basiswerte einer Person berechnet, also die täglichen und wöchentlichen Schwankungen herausgemittelt werden, was die Detektion verbessert. Für die bisher genaueste Fieberdetektionsanalyse können etwas mehr als 120.000 Datensätze verwendet werden. Um die Einflüsse von klimatischen Schwankungen berechnen zu können, werden ca. 340.000 Datensätze verwendet. Um die landkreis aufgelösten Ruhepulsmittelwerte zu berechnen, können über 450.000 Datensätzen verwendet werden.

7. Welche Änderungen und neuen Funktionen sind in Zukunft für die Corona-Datenspende-App geplant?

Zum jetzigen Zeitpunkt sind die Corona-Datenspende-App und die wissenschaftliche Auswertung der Daten ein Projekt, das zum 31. Dezember 2020 endet. Es wurden und werden verschiedene Funktionen und Erweiterungen diskutiert, deren Umsetzung allerdings nur auf Basis einer nachhaltigen Finanzierung bis mindestens Ende des Jahres 2021 und darüber hinaus erfolgen kann. Eine wesentliche Erweiterung der Corona-Datenspende-App ist die Verbindung mit anderen am RKI verankerten Surveillance-Systemen, wie der GrippeWeb-Plattform. Ganz allgemein werden wissenschaftliche Pläne für Citizen-Science-Projekte in der Gesundheitsforschung entworfen, für die die Corona-Datenspende-App als Schablone dienen kann.

8. Wie wird der Datenschutz der gesammelten Daten sichergestellt und gewährleistet, dass diese Daten nicht anderweitig verwendet werden?

Vor Inbetriebnahme wurde ein Datenschutzkonzept erarbeitet und eine Datenschutz-Folgenabschätzung gem. Artikel 35 der Datenschutzgrundverordnung (DSGVO) durchgeführt. Darüber hinaus gab es im Vorfeld umfangreiche Konsultationen des Bundesbeauftragten für den Datenschutz und die Informationsfreiheit (BfDI). Die vom BfDI formulierten Anregungen wurden umgesetzt.

Die Daten werden sicher in einem (ISO 27001) zertifizierten Rechenzentrum in Deutschland gespeichert. Der Server wird in der Open Telecom Cloud („OTC“) der deutschen Telekom gehostet. Die Server der OTC befinden sich in Deutschland und sind hinsichtlich Qualitätsmanagement und Sicherheit mehrfach extern auditiert und zertifiziert. Die pseudonymisierten Daten werden logisch und physisch von Datensätzen anderer Mandanten des Dienstleisters getrennt.

Bestehende Zertifikate sind u. a.:

ISO 27017 (Prozesse für Informationssicherheit in der Cloud),

ISO 27018 (Prozesse für Schutz persönlicher Informationen in der Cloud),

TCDP 1.0 (Trusted Cloud Datenschutzprofil).

Auf dem Server gespeicherte Daten werden täglich automatisch über die Backup-Funktion der Telekom Cloud gesichert. Für Zugang und Betrieb gelten hohe Sicherheitsstandards.

Lediglich die Auswertungsergebnisse werden zu Forschungszwecken in aggregierter und anonymisierter Form, die keinen Rückschluss auf Einzelperson zulässt, dauerhaft in einer wissenschaftlichen Forschungsdatenbank des RKI gespeichert.

Mit dem Dienstleister wurde im gesetzlich erforderlichen Umfang eine Vereinbarung über die Datenverarbeitung im Auftrag abgeschlossen, die den Anforderungen des Artikel 28 DSGVO entspricht. Das RKI erteilt dem Dienstleister entsprechende Weisungen zum Umgang mit den Daten.

*Vorabfassung - wird durch die lektorierte Version ersetzt.*

*Vorabfassung - wird durch die lektorierte Version ersetzt.*

*Vorabfassung - wird durch die lektorierte Version ersetzt.*