

#OpenCoronaData - Appell an das Robert Koch-Institut von 45 Datenjournalistinnen und Datenjournalisten

An den
Präsidenten des Robert Koch-Institut
Herrn Professor Dr. Lothar H. Wieler
Nordufer 20
13353 Berlin

Stuttgart / Hamburg / Baden-Baden, den 18. Mai 2020

Sehr geehrter Herr Prof. Wieler,

die Bekämpfung der Pandemie stellt Gesellschaft, Wissenschaft, Politik, Wirtschaft und Medien vor großen Herausforderungen. Sie haben stets betont, dass wir sehr lange mit diesem Virus werden leben müssen. Das bedeutet für uns als Journalistinnen und Journalisten, dass wir mitten in einer sehr lange andauernden Situation der Berichterstattung stehen.

Unsere Aufgabe ist es ausführlich darzustellen und kritisch zu hinterfragen, wie Entscheidungen fallen, ob Prozesse funktionieren und wie Akteurinnen und Akteure in der Bekämpfung der Epidemie handeln - auch Gesundheitsbehörden. Wir benötigen daher umfassende Informationen zu dieser Epidemie, um den gesellschaftlichen und politischen Diskurs über die richtigen Entscheidungen in dieser Krise evidenz- und faktenbasiert zu begleiten und zu bereichern. Das gelingt in besonderer Weise, wenn wir dazu verlässliche Daten darstellen und analysieren können. Leider reichen die Daten, die das Robert Koch-Institut bisher veröffentlicht, dafür nicht aus. Auch die Art der Veröffentlichung (z.B. in text- und bildbasierten Berichten, statt in maschinenlesbaren Datensätzen) hindert uns daran, die Bevölkerung ausreichend über die Epidemie zu informieren.

Deshalb wenden wir, die unterzeichnenden 45 Datenjournalistinnen und Datenjournalisten, uns mit einem gemeinsamen Anliegen an Sie und fordern Sie auf, offene Corona-Daten schnellstmöglich bereitzustellen und diesen Ansatz mit Nachdruck voranzutreiben und zu unterstützen.

Stellen Sie die Corona-Daten des Robert Koch-Instituts möglichst umfassend, vollständig, regelmäßig und zeitnah für die Öffentlichkeit zur Verfügung.

Bitte tragen Sie auch Sorge dafür, dass Ihre Behörde (und insbesondere Ihre Pressestelle) personell, technisch und inhaltlich in die Lage versetzt wird, diesem datenbezogenen Informationsinteresse der Medien Rechnung tragen zu können.

In den vergangenen Tagen und Wochen haben Sie aus vielen unserer Redaktionen zahlreiche Datenanfragen erhalten, die leider zu oft nur teilweise oder gar nicht beantwortet worden sind. Wir benötigen aber freien Zugang zu Informationen. Das schließt insbesondere auch die Daten ein.

Der diesem Brief anhängende Katalog beschreibt, welche Daten Medien brauchen, um die dringend nötige Aufklärung in der Coronakrise zu leisten. Natürlich ist diese Aufstellung nicht erschöpfend. Darum möchten wir dringend anregen, mit uns in einen kontinuierlichen Prozess einzutreten. Der Presseworkshop zur Reproduktionszahl war ein gutes Beispiel für einen vertrauensvollen, offenen und tiefgreifenden Austausch. Wir vertrauen darauf, dass Sie diese Art der Zusammenarbeit mit den Medien, insbesondere mit uns Datenjournalistinnen und Datenjournalisten, intensivieren.

Mit offenen Daten tragen Sie wesentlich dazu bei, dass das Vertrauen der Bevölkerung in unsere Berichterstattung und letztlich in die staatlichen Institutionen in dieser Krise nicht verloren geht. Wir beobachten mit größter Sorge, dass immer mehr Fake News auftauchen und sich viele Menschen rationalen Überlegungen verschließen und Verschwörungstheorien zuwenden. Wir hoffen inständig, das RKI dafür gewinnen zu können, dieser Entwicklung das Beste entgegenzusetzen, was einer offenen Gesellschaft passieren kann: Transparenz und Öffentlichkeit - allen voran auch bei den Corona-Daten.

Mit freundlichen Grüßen

Ihre
Datenjournalistinnen und Datenjournalisten*

**Die Initiative ist ein informeller Zusammenschluss von Datenjournalistinnen und Datenjournalisten deutschsprachiger Medien. Ansprechpartner für Rückfragen sind: Johannes Schmid-Johannsen, Björn Schwentker und Jan Georg Plavec.*

Die Unterzeichnenden

1. Jan Georg Plavec (Stuttgarter Zeitung)
2. Johannes Schmid-Johannsen (SWR)
3. Björn Schwentker (NDR)
4. Ulrich Lang (SWR)
5. Claus Hesselning (NDR)
6. Frederik von Castell (Universität Mainz, hr)
7. Robert Schöffel (BR)
8. Simon Haas (ZDF)
9. Michael Kreil (BR)
10. Vanessa Wormer (Süddeutsche Zeitung)
11. David Hilzendege (Südkurier)
12. Uwe Springfeld (freier Wissenschaftsjournalist)
13. Hinnerk Feldwisch-Drentrup (Mitgründer von MedWatch und freier Journalist)
14. Christian Endt (Süddeutsche Zeitung)
15. Manuel Mohr (Mitteldeutscher Rundfunk)
16. Götz Gringmuth-Dallmer (Rundfunk Berlin-Brandenburg)
17. Lars Koppers (Karlsruher Institut für Technologie und Gastwissenschaftler am Science Media Center Germany)
18. Sophie Rotgeri (Science Media Center Germany)
19. Holger Dambeck (DER SPIEGEL)
20. Marcel Pauly (DER SPIEGEL)
21. Robert Meyer (ZDF, heute.de)
22. Jan Eggers (hr)
23. Anna Behrend (NOZ/mh:n Medien)
24. Haluka Maier-Borst(rbb)
25. Yannik Buhl (Stuttgarter Zeitung)
26. Marie-Louise Timcke (Funke Mediengruppe)
27. Marco Maas (Datenfreunde GmbH)
28. Daniel Kraft (Mannheimer Morgen)
29. Christopher Ophoven (WDR)
30. Hendrik Lehmann (Tagesspiegel)
31. Philipp David Pries (Münchner Merkur / Ippen-Digital-Zentralredaktion)
32. Jochen Schäfer (SWR)
33. James Daniell (Risklayer GmbH, CEDIM / Karlsruher Institut für Technologie)
34. Frieder Kümmerer (SWR)
35. Phil Ninh (Rheinische Post)
36. Laura Stresing (t-online.de)
37. Claudia Kohler (BR)
38. Nico Heiliger (SWR)
39. Christian Schellenberger (BNN)
40. Achim Tack (DER SPIEGEL)
41. Martina Schories (Süddeutsche Zeitung)
42. Elias Harlan (Bayerischer Rundfunk)
43. Jakob Vicari (tactile.news)
44. Colin Büchtmann (NDR)
45. Natalie Sablowski (WDR)

*sowie
ddjbw (Netzwerk für
Datenjournalismus und Open Data
in Baden-Württemberg)*

Datenkatalog

- A. Offenlegung der Datenspeicherstruktur der Corona-Daten des Robert Koch-Instituts bzw. (soweit in Abstimmung mit dem RKI erstellt) der Gesundheitsbehörden in Deutschland (Namen der Datenbanken, Angaben der Merkmale bzw. Spalten, Metadaten und Dokumentationen)

- B. Die folgenden Daten sollten generell tagesaktuell, als Zeitreihe, und regional so fein wie möglich aufgegliedert veröffentlicht werden. Das Datenformat sollte maschinenlesbar und immer gleich strukturiert sein. Als Vorbild kann der Daten-HUB "RKI COVID19" dienen (https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0/data). Wo immer möglich, wäre es wünschenswert, in diesen Datensatz die genannten neuen Merkmale einzupflegen. Datensätze sollten nicht überschrieben, alte Versionen erhalten bleiben.
 1. Bereitstellung aller anonymisierten Datensätze der Epidemie insbesondere des Infektionsgeschehens als vollständige (auch historische) Zeitreihen, insbesondere:
 - a. Fallzahlen mit imputiertem Erkrankungsbeginn (wie im Nowcast verwendet)
 - b. Sterbefälle mit Sterbedatum
 - c. Fallzahlen mit zusätzlichen Kategorien nach IfSG (z.B. §§ 23, 33, 36 oder 42, wie im Lagebericht dargestellt) z.B. nach Arbeitsstatus oder Wohnverhältnissen der positiv getesteten Personen
 - d. Genesene mit dem bereits vorliegenden echten oder geschätzten Genesungsdatum (ersetzt durch die Angaben aus den erweiterten Meldepflichten, sobald diese vorliegen)
 - e. Altersgruppen so genau wie möglich (idealerweise 5 Altersjahre)
 - f. Alle Zeitstempel im Ablauf des Meldeverfahrens (z.B. Eingang der Meldung beim Gesundheitsamt, Erkrankungszeitpunkt, Datenweitergabe an Land, Datenweitergabe an RKI, Veröffentlichung durch das RKI)
 2. Tägliche Angaben zu Testungen und deren Ort (Z.B. Arztpraxis oder Stationstyp im Krankenhaus), getesteten Personen (mit Geschlecht und Altersgruppe), positiven Testergebnissen, Anzahl übermittelnder Labore und Testkapazitäten, Anzahl von Laboren mit Rückstau und dessen Ausmaß in abzuarbeitenden Proben, Ausmaß von Lieferschwierigkeiten und der betroffenen Artikel.
 3. Die genaue Schätzung von R mit Konfidenzintervallen, die anscheinend auch auf Länderebene zur Verfügung gestellt wird wie z.B. für Berlin.
 4. Der Code hinter dem Nowcast aber auch hinter Veröffentlichungen wie den "Beispielszenarien der SARS-CoV-2- Epidemie 2020 in Deutschland"
 5. Aus dem „R-Workshop“ ist uns ferner bekannt, dass das RKI regelmäßig Gesundheitsämter abfragt, wie groß die Belastung ist. Die Ergebnisse dieser regelmäßigen Umfrage würden wir auch gerne veröffentlicht sehen.