

# Digitalgespräch Folge 42

## Sensible Daten für die Wissenschaft: Weshalb ein Forschungsdatengesetz?

Mit Stefan Bender von der Deutschen Bundesbank, 10. Oktober 2023

<https://zevedi.de/digitalgespraech-042-stefan-bender/>

*[Vorspann mit Musik und Ausschnitten aus dem Gespräch beginnt.]*

**Marlene Görger [mg]:** Herr Bender, Sie sind Wirtschaftswissenschaftler, Sie leiten das Forschungsdaten- und Servicezentrum der Deutschen Bundesbank, kennen also die Aushandlungsprozesse zwischen Wissenschaft und Datengebern und deren jeweiligen Bedarfe sehr gut.

**Stefan Bender [Bender]:** Es gibt ja jetzt auch die Diskussion zu einem Forschungsdatengesetz, das aus der Taufe gehoben werden soll. Wir können uns natürlich jetzt über die Inhalte austauschen, die da drinstehen müssen.

**Petra Gehring [pgg]:** Kann man sagen, es gibt irgendwo eine Grenze? Oder ist das risikolos möglich, dass eine Gesellschaft alle ihre Behördendaten, Public-Sector-Daten der Forschung gibt?

**[Bender]:** Verkürzung des Abiturs um ein Jahr, hat das irgendwelche Auswirkungen auf den Erwerbsverlauf? Coronakrise, sagen wir: Schulabgänger, die jetzt in dieser Zeit in den Arbeitsmarkt gehen. Ich sehe ja, dass man zum Wohle der Gesellschaft bestimmte Sachen einfach auch erforschen sollte. Mir tut es weh, wenn ich sage: Och ich schmeiß irgendwas weg. Ich hätte das lieber, dass diese Daten dann irgendwo sicher lagern und der Personenbezug nicht mehr hergestellt werden kann.

*[Der Vorspann endet, das Gespräch beginnt.]*

**[mg]:** Mit der Digitalisierung hat sich das Verhältnis der allermeisten Menschen zu ihren persönlichen Daten verändert. Wir generieren und teilen sie heute täglich, oft unbewusst, vor allem mit privaten Unternehmen, manchmal aber auch mit Behörden. Vom Staat erwarten wir allerdings, dass er Bürgerdaten maximal schützt. Das Bundesverfassungsgericht etablierte in den 1980er Jahren das Grundrecht auf informationelle Selbstbestimmung und schuf damit die Grundlage für den Datenschutz, den wir heute viel diskutieren. Wobei es scheint, als würden die Deutschen ihre Daten dennoch lieber internationalen IT-Konzernen anvertrauen als den Meldeämtern. Gleichwohl entstehen natürlich auch bei Behörden zunehmend große Mengen an Daten. Und wer Zugang zu diesen Daten hat, kann vielfältigen Nutzen daraus ziehen. Insbesondere steckt darin potenzielles Wissen, das staatliche Leistungen verbessern kann. Naheliegendermaßen also, dass die Forschung, allen voran die Sozialforschung, einen geregelten Zugang zu Behördendaten fordert. Der ist aber auch für die Wissenschaft gar nicht immer einfach zu bekommen. Und nur den Datenschutz dafür verantwortlich zu machen, wäre zu einfach. Welche Daten sind das, an denen die

Forschung so interessiert ist? Und was macht ihren besonderen Status aus? Welche Fortschritte wären möglich, wenn Forscher:innen leichter Zugang hätten? Und welche Gründe gibt es trotzdem, genau zu überlegen, wie dieser Zugang gestaltet sein sollte? Darüber reden wir heute im Digitalgespräch. Mein Name ist Marlene Görger, ich bin Physikerin und Technikphilosophin und arbeite am Zentrum verantwortungsbewusste Digitalisierung.

**[pgg]:** Und ich bin Petra Gehring, Professorin für Philosophie an der Technischen Universität Darmstadt. Bei uns zu Gast im Digitalgespräch ist heute Professor Stefan Bender. Er ist zugeschaltet in unsere Videokonferenz aus Frankfurt am Main. Hallo und herzlich willkommen im Digitalgespräch, Herr Bender! Wir freuen uns sehr auf Ihre Expertise.

**[Bender]:** Hallo Frau Görger, Hallo Frau Gehring. Ich freue mich auch, dass ich dabei sein darf und freue mich auf den Austausch und das Gespräch.

**[mg]:** Herr Bender, Sie sind Wirtschaftswissenschaftler. Sie leiten das Forschungsdaten- und Servicezentrum der Deutschen Bundesbank. Zudem sind Sie Honorarprofessor an der Universität Mannheim. Ihre umfassende Expertise zu Datenzugang, Big Data, zu Verknüpfung von Daten und Datenqualität bringen Sie in verschiedene Beiräte ein. Zum Beispiel vertreten Sie die Deutsche Bundesbank als Datengeber im Rat für Sozial- und Wirtschaftsdaten, dem Rat SWD. Sie sind aber als Wissenschaftler auch aus der Forscherperspektive an praktikablen Regelungen für den Zugang zu Behördendaten interessiert. Sie kennen also die Aushandlungsprozesse zwischen Wissenschaft und Datengebern und deren jeweiligen Bedarfe sehr gut. Und Sie haben Erfahrung darin, hier Lösungswege zu finden. Bevor wir aber über Lösungen sprechen, lassen Sie uns erst mal die Diskussion verstehen, denn das ist ja ein totales Expertenthema. Was genau sind denn Behördendaten und was macht sie so besonders? Und warum interessiert die Forschung sich dafür?

**[Bender]:** Behördendaten sind eigentlich Daten, die im Rahmen von Verwaltungsakten von eben öffentlichen Verwaltungen anfallen. Bei der Bundesagentur für Arbeit sind dieses zum Beispiel Daten zur sozialversicherungspflichtigen Beschäftigung, zur Arbeitslosigkeit oder auch zum SGB-II-Bezug. Bei der Bundesbank, wo ich arbeite, sind es eben Daten zu Banken oder zu Wertpapieren oder auch zu Unternehmen. Die Daten sind besonders, weil sie sehr präzise sind. Sie sind auch meistens die Grundgesamtheit. Also wir haben da für alle Personen, Unternehmen oder Banken diese Daten. Mein Lieblingsbezug zurzeit ist die Vergabe des Nobelpreises 2021, also 2021 wurde der Nobelpreis in Ökonomie an David Card, Joshua Enquist und Guido Imbens vergeben, die die Wissenschaft in diesem Segment vorangebracht haben, weil sie über kausale Zusammenhänge geforscht haben. Wir können uns vorstellen, wie wir Kausalität messen. In den Naturwissenschaften sind das Experimente. Ich habe also eine kontrollierte Umgebung. Wenn ich jetzt Zusammenhänge wie in Wirtschaft oder Verhalten von Personen mir anschau, kann ich auf der einen Seite natürlich auch wieder versuchen,

das in Experimente zu gießen. Auf der anderen Seite kann ich natürlich sagen, kann ich durch bestimmte Gegebenheiten experimentelle Designs in der Wirklichkeit finden, und bestimmte experimentelle Designs in der Wirklichkeit sind Gesetzesveränderungen, Naturereignisse, Kriege, bestimmte Punkte, wo sie dann an einem bestimmten Punkt etwas festmachen können. Ich kann jetzt mal ein altes Beispiel machen, also wirklich altes Beispiel: Die Wehrpflicht wurde in Deutschland eingeführt und das wurde bei einem bestimmten Geburtsdatum begonnen in Deutschland. Ab diesem Tag, wenn man älter war als dieses eine Datum, musste man dann zur Armee gehen, also Bundeswehr, und davor musste man nicht. Und jetzt kann man natürlich anschauen, hat das Auswirkungen auf den Erwerbsverlauf gehabt, dass diese Personen eben wehrpflichtig wurden im Vergleich zu denen. Und da würde man zwei Wochen vor diesem Geburtsdatum und zwei Wochen nach diesem Geburtsdatum einfach die Personen vergleichen, um zu sagen: Och, die sind ja nahezu gleich alt, schauen wir doch mal an, wie diese Wehrpflicht dann sich auf die Erwerbstätigkeit ausgewirkt hat. Ein aktuelleres Beispiel ist Verkürzung des Abiturs um ein Jahr. Hat das irgendwelche Auswirkungen auf den Erwerbsverlauf? Coronakrise: sagen wir Schulabgänger, die jetzt in dieser Zeit in den Arbeitsmarkt gehen, haben die andere Voraussetzungen als davor? Und da können Sie auch sagen: Auch der Jahrgang, der davor reinging, ist ja vergleichbar wie der Jahrgang, der danach. Und dann haben Sie quasi ein experimentelles Setting.

**[pgg]:** Als Philosophin muss ich natürlich dann sofort sagen, im Grunde sind es wahrscheinlich immer nur Indizienbeweise, keine echte Kausalität. Und ich vermute jetzt mal, dass dieser Nobelpreis seinen Grund darin hat, dass die Beweislage so genau war und so kleinteilig und so mehrfach belegt, dass es dann sich einer Naturkausalität annähert - die Schärfe, mit der man belegen kann.

**[mg]:** Dass da natürlich ein großes Interesse an möglichst vielen und dichten Datensätzen entsteht ist klar.

**[Bender]:** Ich habe mit David Card, hinterher auch gesprochen, ob ich das sagen darf, dass ohne die Art von Daten das nicht möglich war. Da hat er gesagt: ja. Also dieser Ansatz wäre nicht möglich gewesen, wenn sie nicht diese Art von Daten gehabt hätten oder wir nicht in den Sozial- und Wirtschaftswissenschaften diese Art von Daten hätten, weil für diese Art von Datenanalysen und Versuche von Zusammenhängen herauszufinden, brauchen wir zum Beispiel präzise Datumsangaben. Das heißt, wenn wir da zum Beispiel Personen befragen würden, aber ich bin jetzt schon ein bisschen älter, Erinnerungsprobleme zum Beispiel, wann das genau passiert ist. Und wir haben auch natürlich eine Herausforderung direkt an diesem Datum, oder sehr nah an dem Datum, brauchen wir dann auch sehr viele Messstellen, also sehr viele Personen oder Unternehmen. Das bekommen wir dann bei zum Beispiel Befragungen oder so nicht hin, deswegen brauchen wir hier Grundgesamtheiten, also möglichst viele Personen oder Unternehmen, damit wir darüber Aussagen treffen können.

**[pgg]:** Also ich habe jetzt verstanden: Die Behörden haben unheimlich vollständige Datensätze. Im Grunde sind es ja eigentlich alle, für die Sie zuständig sind, zu denen Sie dann auch Daten haben. Die Daten sind ziemlich genau und granular, und sie sind auch zeitlich engmaschig unter Umständen verfügbar, sodass man so Vorher-Nachher-Effekte hat. Leuchtet total ein, dass das aus wissenschaftlicher Sicht quasi die perfekte Datenerhebung ist. Warum teilt der Staat denn Daten nicht einfach mit der Wissenschaft? Öffentliche Forschung ist ja was im weiteren Sinne Staatliches und auch vertrauenswürdig, und da kann man ja auch sich verpflichten, Daten nicht weiterzugeben. Warum gibt es da so Barrieren?

**[Bender]:** Ich muss jetzt erst mal natürlich Ihren Punkt aufgreifen: Wir haben alles. Das stimmt so nicht. Jede Organisation hat nur die Daten, die für ihre Zweckbindung gilt. Jetzt wieder bei der Bundesagentur für Arbeit zum Beispiel. Hier gibt es sehr präzise Informationen zu jeder einzelnen Person. Aber zum Beispiel bei Beschäftigung ist es die sozialversicherungspflichtige Beschäftigung. Das heißt, da gibt es keine Informationen über Beamte oder über Selbstständige. Das heißt, da fehlen schon bestimmte Teile. Sie haben auch die Daten dort in Personeninformation und wenn wir selber uns mal überlegen, wie passieren bei uns zum Beispiel Entscheidungen. Wir machen die zum Beispiel mit der Ehepartnerin oder dem Ehepartner oder mit der Familie. Und das geht da zum Beispiel auch nicht, weil man dort nämlich nur Verläufe von einzelnen Personen hat, aber nicht den Haushaltskontext. Bestimmte Informationen fehlen auch, die zum Beispiel wichtig sind, um zu sehen, wie ein Berufsverlauf zum Beispiel funktioniert. Bildungsinformationen sind nur rudimentär zur Verfügung. Oder auch Gesundheit fehlt vollständig. Das sind wesentliche Aspekte, zum Beispiel, die unser Berufsleben auch determinieren. Nun zurück zu Ihrer anderen Frage: Warum teilt der Staat das nicht? Wir haben hier einen Abwägungsprozess zwischen auf der einen Seite Forschungsfreiheit, oder ich nenne es auch gerne, zum Wohle der Gesellschaft bestimmte Ergebnisse zu ermöglichen, und auf der anderen Seite gibt es auch noch zwei andere Komponenten. Das eine ist natürlich der Datenschutz, dass Personendaten oder auch Informationen, die die Bundesbank jetzt hat zum Beispiel, sehr detaillierte Daten zu Banken haben wir ja hier, dass diese also besonders schützenswert sind, und was jetzt immer mehr von Bedeutung ist, ist natürlich auch, dass wir als Behörden Vertrauen generieren müssen. Das heißt also Personen, aber auch die Einheiten, die wir untersuchen, müssen ja auch Vertrauen zu uns haben, dass wir diese Daten auch in einer sinnvollen Art und Weise verwenden und nicht missbrauchen.

**[pgg]:** Das heißt, die Behörden vertrauen der Wissenschaft nicht?

**[Bender]:** Also ich gehe davon aus, die Behörden vertrauen der Wissenschaft. Und wir können ja jetzt auch den Spieß umdrehen und sagen: Wenn wir uns jetzt mal einige der öffentlichen Organisationen anschauen – also Bundesbank nenne ich jetzt mich als erster – statistisches Bundesamt, Deutsche Rentenversicherung, Bundesagentur für Arbeit haben ja Zugänge zu ihren Daten für die Forschung geschaffen. Diese Zugänge unter den bedingten, bestimmten Regelungen werden die durchgeführt. Wir können

dann darüber diskutieren, wie effizient das dann für die Forschung ist. Aber das heißt, es gibt natürlich ein Vertrauen. Das ist auch gesetzlich natürlich vorgegeben, deswegen würde ich das gar nicht so überspitzt hier sehen.

**[mg]:** Wie ist denn der Zugang bisher geregelt? Können Sie mal beschreiben, wie das im Einzelfall abläuft?

**[Bender]:** Jetzt wird's technisch, jetzt bin ich mal gespannt. Es gibt verschiedene Arten von Zugängen zu Daten und auch verschiedene Möglichkeiten, wie man Daten schützt. Letztes Beispiel, wo man dazu was findet, ist im Data Governance Act. Sie können auf der einen Seite Datensätze anonymisieren. Das heißt also: Sie verändern die Originaldaten in einer Art und Weise, dass sie geschützt sind. Das führt aber oftmals dazu, dass die Daten dann für die Forschung nicht mehr verwendbar sind. Weil ich hab vorhin gesagt, der Nobelpreis wurde vergeben, damit die Daten sehr präzise sind, wenn sie dann das Datum zerstören, also in dieser Anonymisierung, geht der Ansatz, den ich vorher genannt, nicht mehr. Sie können, wenn Sie dann sagen: Och die Daten sollten halbwegs noch genügend Informationsgehalt haben, auf der anderen Seite aber schützenswert sein, gibt es jetzt die Konstruktion sogenannter Forschungsdatenzentren. Das heißt also, das sind Räumlichkeiten, also die Datenproduzenten stellen Räumlichkeiten zur Verfügung, in denen die Forscherinnen und Forscher dann arbeiten können. Das sind sogenannte sichere Umgebungen. Forscherinnen, Forscher kommen hierher, arbeiten an einem Arbeitsplatz, der keine Verbindung zum Internet hat, das Handy ist verboten und andere Punkte, und arbeiten dann in dieser Umgebung. Das ist zurzeit die Praxis. Davor wird aber festgestellt, wird ein Vertrag dann geschlossen, für welchen Zweck Sie, also welchen Zweck Ihre Forschung hat. Und Sie müssen dann auch noch bestimmten Rahmenbedingungen zustimmen. Also es ist ein sehr geregelter Zugang für Forschende, manchmal als sehr bürokratisch wahrgenommen, aber es gibt ja eine spezifische Möglichkeit, an die Daten zu kommen, die Forschungsdatenzentren wie Bundesbank und andere haben ja auch einen bestimmten Aufwand zu betreiben, weil wir stellen natürlich Personal, um die Forschungen zu betreuen, um auch die Daten zu erstellen. Wir müssen auch Dokumentationen zu den Daten durchführen, weil einfach nur Daten ohne Beschreibung sind ja so gut wie wertlos. Und wir müssen auch hinterher sicherstellen, dass alles, was veröffentlicht wird, wiederum dem Datenschutz genügt. Also, das sind so die Aufgaben, die wir haben.

**[mg]:** Sie haben jetzt erwähnt, dass der Zweck der Forschung im Einzelfall dann angegeben werden muss. Hat das den Hintergrund, dass man dann überprüfen kann, ob auch nur dieser Zweck verfolgt wurde? Oder könnte es auch sein, dass Vorhaben abgelehnt werden, weil der Zweck nicht unterstützt wird?

**[Bender]:** Das ist ein guter Punkt. Also die unabhängige Forschung hat Datenzugang zurzeit in Deutschland zu den genannten Institutionen, zum Beispiel kommerzielle Forschung nicht. Das heißt, zurzeit dürfen zum Beispiel Universitäten oder Institute der Leibniz-Gemeinschaft zu diesen Daten, allerdings nicht zum Beispiel die Forscher von

kommerziellen Banken bei uns. Also das heißt, die haben diesen Zugang nicht. Nicht, weil sie nicht auch vergleichbare Forschung machen, sondern wir geben auch deswegen den Zugang nicht, weil die Daten so detailliert sind und wir nicht ausschließen können, dass dann doch eventuell geschaut wird, was der Mitbewerber angibt. Also hier kommt wiederum Datenschutz und auch Vertrauen, was wir vorhin genannt haben, wieder zum Tragen. Was zurzeit diskutiert wird, muss ich jetzt noch zusetzen, sind Datenzugänge, die Remote Access heißen, also die Fernverarbeitung. Ich habe gesagt, zurzeit müssen die Forschenden physisch bei uns arbeiten. Es ist technisch möglich und sollte auch durchgesetzt werden, dass man vom eigenen Arbeitsplatz über eine sichere Verbindung eben Auswertungen machen kann. Dazu gibt es Bestrebungen zurzeit.

**[pgg]:** Sind die meisten größeren deutschen Behörden so aufgestellt wie jetzt die Bundesbank oder die Arbeitsmarktforschung oder die Bundesagentur für Arbeit, dass tatsächlich so ein Forschungsdatenzentrum existiert und auch die Routinen da sind und Personen die Forschenden dann begleiten und diesen ganzen Controlling Workflow bearbeiten? Oder sind Sie sozusagen ein Premiumhaus?

**[Bender]:** Ich könnte jetzt einfach ja sagen. Nein, aber es gibt ja beim Rat für Sozial- und Wirtschaftsdaten, also ich rede jetzt mal für die Sozial- und Wirtschaftsforschung, 42 akkreditierte Forschungsdatenzentren, das heißt, da gibt es eine sehr große Kultur in Deutschland, würde ich sagen, oder sehr starke Kultur in Deutschland, diesen Zugang über Forschungsdatenzentren zu ermöglichen. Es gibt auch dann Forschungsdatenzentren der Wissenschaft, zum Beispiel die GESIS würde ich dann nennen wollen, die dann auch zu bestimmten Daten eben einen Zugang ermöglichen. Jetzt werde ich vielleicht übergriffig zu anderen Bereichen, ich denke, dass die Sozial- und Wirtschaftswissenschaft da ein, sagen wir, Leuchtturm ist. Ich würde mir wünschen, dass das bei anderen Fachdisziplinen ähnlich wäre. Also, bashe jetzt mal die Medizin oder andere. Also was ich aber betonen möchte, ist, dass jeder da seine eigenen Daten hat und es eigentlich dringend notwendig ist, dass diese Daten auch verknüpft werden, wo wir dann aber neue Schwierigkeiten haben.

**[mg]:** Ich würde gern, bevor wir darauf kommen, noch mal auf den Aufwand oder auf die Ressourcen, vielleicht auch Unterschiede eingehen. Sie sagten ja vorhin, dass Daten, die für die Aufgaben der Behörden notwendigerweise anfallen, wie auch immer, nicht unbedingt gleich auch die Daten sind, die Forscherinnen und Forscher gerne hätten. Kommt es denn vor, dass gezielt Daten für die Forschung erhoben werden, die in der Behörde für sich genommen jetzt gar nicht unbedingt erfassen würde?

**[Bender]:** Sind mir auch Fälle bekannt. Also muss aber allerdings auch die Behörde wieder Interesse daran haben. Ein schönes Beispiel wieder aus dem Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, wo das Forschungsdatencenter der Bundesagentur für Arbeit betrieben wird, ist, das IAB führt auch eigene Erhebungen durch und verknüpft die eigenen Erhebungen mit den

administrativen Daten, wenn eine informierte Einwilligung vorliegt. Also es wird dann dem Datenschutz Genüge getan und die Personen oder die Unternehmen werden befragt, und die, die zustimmen, wird verknüpft. Da gibt es zum Beispiel auch Beteiligungsmodelle, dass Forschende ihre Fragen einreichen können. Das ist ein Modell. Das andere sind auch Kooperationsprojekte. Also, die Deutsche Rentenversicherung hat mit dem sozioökonomischen Panel zusammengearbeitet. Das sozioökonomische Panel ist eine große Haushaltsbefragung, die von der Forschung organisiert wird. Und das IAB macht das Gleiche mit dem LfBi. Das ist das Bildungspanel, das große aus Bamberg. Und das ist ja ein extremer Vorteil, kann man jetzt sagen, Best of Both Worlds. Administrative Daten zeigen ja hauptsächlich Sachverhalte. Also jemand geht in die Arbeitslosigkeit oder geht in eine Beschäftigung. Was Behörden zum Beispiel nicht haben, sind Daten zu Erwägungen oder Erwartungen und andere Punkte, die man dann durch Befragungen bekommt. Oder, wie ich schon gesagt habe, Haushaltskontext fehlt zurzeit in den Daten der Bundesagentur wie auch bei der Deutschen Rentenversicherung und man kann dann über die Befragungen diese Kontexte abfragen. Schöner wäre es natürlich, man bekäme sie auch woanders noch her, aber das ist dann Best of Both Worlds, was man da hat.

**[pgg]:** Sind es eigentlich vor allem dann Bundesbehörden und Daten von Bundesbehörden, auf die die Forschung zugreifen kann? Oder gibt es auch die Möglichkeit, auf Landes- und Kommunalebene zu kooperieren mit Behörden?

**[Bender]:** Zurzeit sind überwiegend Bundesbehörden aktiv. Und das ist ja auch ein Manko, den zum Beispiel der Rat für Sozial- und Wirtschaftsdaten und ich glaube auch der Rat für Informationsinfrastrukturen angemahnt hat, nämlich dass wir bei bestimmten Datenbeständen, die jetzt in Länderhoheiten sind, wir sehr schlechte Zugänge haben. Also ein besonders prominentes Beispiel ist Bildung. Es wäre wünschenswert, wenn wir für Deutschland zentrale Informationen zur Bildung haben, also wissen, welche Schulabschlüsse haben die Personen, welche Art von Ausbildungsabschlüssen und sowas hat jemand. Wie ich vorhin schon gesagt habe, ist das ja dann wichtig, weil das ja den weiteren Erwerbsverlauf determiniert. Also dazu haben wir wenig Informationen. Wo wir sehr viel Information – wenn wir jetzt auf Bildung gehen – das ist eine der, nenne ich mal Sternstunden, wie alles begann, weil das damalige Arbeitsministerium gesagt hat: Wir wollen mehr über die Wirkung von arbeitspolitischen Maßnahmen haben, das heißt also Umschulungen und was es dort so alles gibt. Da gab es Daten, bei der Bundesagentur für Arbeit fielen die ohnehin an, aber dann gab es eine sehr umfassende Wirkungsforschung von arbeitspolitischen Maßnahmen. Also da ist Deutschland wirklich top, aber man sieht, dass man das zum Beispiel bei Bildung und anderen Dingen machen könnte. Aber da entstehen, persönliche Auffassung von mir, auch wiederum Ängste bei den Ländern, weil man dann plötzlich auch Ländervergleiche machen könnte, um dann zu sagen: Dieses Bundesland hat ein besseres Abitur als jenes oder ist erfolgreicher. Ich sehe aber, dass man zum Wohle der Gesellschaft auch bestimmte Sachen einfach auch erforschen sollte.

**[pgg]:** Es leuchtet ein, wenn man ein bisschen drüber nachdenkt, dass gute Forschung, die wirklich auch Transparenz herstellt, natürlich für die öffentlichen Akteure auch zumindest einen sanften Kontroll- und Qualitätskontrollaspekt hat. Und letztlich zumindest die Forschungsergebnisse werden ja dann öffentlich. Natürlich nicht die Daten, aber die Ergebnisse und dann kann schon diskutiert werden: Wie wirksam war eigentlich welche Art von Politik oder was ist die Folge von Unterlassungen?

**[Bender]:** Es stellt sich natürlich jetzt die Frage, diejenigen, die vielleicht ähnlich wie ich bei einer Behörde arbeiten, was jetzt eigentlich der Vorteil ist, zuweilen für eine Bundesbank, die Daten zur Verfügung zu stellen. Die sind vielfältig. Wir bekommen natürlich Forschungsergebnisse, die wir wiederum selber verwenden können. Also als einen Punkt. Aber wir bekommen auch zum Beispiel eine höhere Datenqualität, weil ja mehr Personen diese Daten nutzen. Dann fallen auch bestimmte, sagen wir, Unzulänglichkeiten der Daten besser auf. Dadurch, dass wir die Daten auch dokumentieren, erhöhen wir auch indirekt die Datenqualität, weil man dann auch wiederum durch das Beschreiben der Daten die eigenen Grenzen oder sowas sieht. Ein Punkt, den ich auch immer gerne mache, ist: Mir ist es lieber, die Forscherinnen und Forschern arbeiten dann mit sehr guten Daten, die ich auch kenne - jetzt ich als Bundesbank - als wenn sie mit Daten von irgendwelchen Dritten forschen und wir dann in der öffentlichen Diskussion eine Diskussion zum Beispiel auf unzureichenden oder schlechterer Datenqualität haben. Also dann hätte ich als Bundesbank nur ein zusätzliches Problem, das ich ja dann erst auch noch mal eine Diskussion über die Qualität der Grundlagen, in dem Fall der Daten, führen muss. Wenn wir da auf einer gemeinsamen Grundlage diskutieren, die wir akzeptieren, dass ist dann zum Beispiel eben die Daten der Bundesbank. Dann ist der Diskurs hinterher einfacher über die Inhalte. Und auch zielführender. Ist ein Punkt, den wir machen. Wir, heißt in dem Fall, noch zwei Kollegen kommen von mir: Jannik Blaschke und Christian Hirsch. Wir vertreten die Auffassung, dass Data Sharing für alle Beteiligten einen Nutzen haben muss. Das Positive sollte nicht nur bei der Forschung, den Nachfragenden nach den Daten, passieren, sondern auch bei den Datengebenden. Also in dem Fall die Institutionen, sollte da was Positives dabei rauskommen. Und dann natürlich für die Gesellschaft. Das heißt also, in dem Fall sollten die Ergebnisse, die die Forschung produzieren, auch wenigstens zu Teilen, wo sinnvoll und notwendig, in den gesellschaftlichen Diskurs überführt werden.

**[pgg]:** Das ist doch mal eine wichtige Botschaft an die Länder und Kommunalbehörden. Sie können Ihre Datenprobleme, ihr Datenhandling, die Datenqualität vielleicht gemeinsam mit der Wissenschaft besser in den Griff kriegen, als wenn Sie alleine so ein bisschen verschämt auf Ihren Daten sitzen.

**[mg]:** Sie hatten ja vorhin schon angekündigt, dass wir vielleicht auch über die Verknüpfung von Daten sprechen sollten. Das ist jetzt vielleicht ein guter Punkt, wenn wir beim gesellschaftlichen Nutzen sind, der ist wahrscheinlich am schwierigsten



vorherzusagen und auch am schwierigsten zu messen. Vielleicht können wir da am besten mit einem Beispiel mal arbeiten, das Sie im Kopf haben, wenn Sie an diese Verknüpfung von Daten denken, die im Augenblick noch isoliert sind.

**[Bender]:** Ich kann Ihnen zwei, drei Beispiele liefern, um zu sagen, wo Verknüpfungen gut wären. Wir hatten ja Corona, und da gab es eine Menge von Hilfsprogrammen für Unternehmen. Also die entweder Kurzarbeitsgeld bei der Bundesagentur für Arbeit beantragen konnten oder finanzielle Unterstützung vom Bund, aber auch von den Ländern. Schauen Sie mal nach, ob Sie irgendwo eine Info, eine Quelle finden, um zu sagen: Welches Unternehmen hat denn in Summe welche Maßnahmen bekommen? Also, wir wissen bei der Bundesagentur für Arbeit, welche Unternehmen Kurzarbeit bekommen haben. Wir wissen dann vielleicht irgendwo, welche Unternehmen haben von den Ländern Geld bekommen. Aber wir haben nirgendwo eine Gesamtschau, zu sagen: Dieses Unternehmen hat Maßnahme eins, zwei und drei bekommen. Frankreich hat das durchgeführt, die haben das hinbekommen, dass die zeigen konnten, welche Unternehmen haben welche Hilfsmaßnahme. Es kann ja sein, dass ein Unternehmen eine Maßnahme bekommen hat, aber die zweite gebraucht hat, um zu überleben. Also jetzt sagt man: Och das Programm war effizient oder nicht effizient oder diese Maßnahme, aber man müsste eigentlich die anderen anschauen. Das andere Beispiel zum Linkage habe ich vorhin schon kurz angerissen. Das ist, wenn man sich Erwerbsverläufe anschaut, habe ich ja gesagt, es wäre wichtig, dort noch Informationen zu Haushalten zu haben, zu Bildung und auch zu Gesundheit. Anderer Punkt, obwohl da sehr gute Arbeit auch im IAB gemacht wird, ist Mindestlohn. Da war ich damals auch ein bisschen beteiligt, den Versuch: Mindestlohn sind ja Stundenlöhne. Die Bundesagentur für Arbeit hat keine Stundenlöhne, sie hat in einem Hilfskonstrukt Tageslöhne. Stundenlöhne hätte das Statistische Bundesamt. Damals wurde versucht, in einem Gesetz eine Zusammenführung dieser beiden Daten zu ermöglichen, und das war nicht von Erfolg gekrönt. Deswegen also, es gibt verschiedenste Inhalte, wo es danach schreit eigentlich, dass die Daten zusammengeführt gehören. Und das passiert zurzeit nicht, weil die Daten sehr enge Zweckbindungen haben und es auch an gesetzlichen Initiativen da ein wenig fehlen, um diese Verknüpfung zu ermöglichen. Neben natürlich noch den technischen Punkten, da sind wir jetzt dann wieder beim Datenschutz. Es gibt Länder, da gibt es einen Personenidentifikator, den man sein Leben mitträgt. Das haben wir in Deutschland nur bedingt. Also es gibt die Sozialversicherungsnummer und solche Punkte, die Steuernummer, die treten aber immer nur ab einem bestimmten Zeitpunkt ein. Ich glaube, die Sozialversicherungsnummer kriegt man jetzt sehr früh im Leben, aber ich habe sie erst bekommen, zum Beispiel mit meiner ersten Erwerbstätigkeit, deswegen fehlt dann die Vorgeschichte bis zu meiner ersten Erwerbstätigkeit. Und zum Zusammenspielen braucht man die Erlaubnis. Man braucht aber auch dann, wie gesagt, die Informationen, die es ermöglichen, dass die Daten zusammengespielt werden können.

**[pgg]:** Sie haben jetzt das Stichwort Gesetz schon genannt. Der Rat SWD und andere wissenschaftliche Akteure fordern ja schon lange, dass der Gesetzgeber was tun muss.

Also das Beispiel war jetzt die Zusammenführung von Daten. Auch dieser Remote Access könnte möglicherweise leichter ins Werk gesetzt werden, wenn ein Gesetz da wäre. Die jetzige Regierung hat ein Forschungsdatengesetz angekündigt. Wie sehen Sie das?

**[Bender]:** Jetzt gebe ich wieder meine persönliche Meinung, nicht der Bundesbank. Wie sehe ich das? Wir haben ziemlich viele zurzeit europäische Initiativen, Gesetzesinitiativen wie den Data Governance Act, aber auch die Datenschutzgrundverordnung und Data Act, die es jetzt ja ermöglichen, dass es, sagen wir, umfassenderen Zugang der Forschung zu Daten geben könnte oder können auch sagen, müsste. Und es gibt ja jetzt auch die Diskussion zu einem Forschungsdatengesetz, das im Koalitionsvertrag steht und wo es ja jetzt auch erste Bestrebungen gibt, dass ein solches Forschungsdatengesetz aus der Taufe gehoben werden soll. Ein notwendiger und sinnvoller Schritt, den da die Politik geht und eigentlich gehen muss, meiner Meinung nach. Wir können uns natürlich jetzt über die Inhalte austauschen, die da drinstehen müssen. Und da gibt es ja verschiedenste Vorschläge. Sie haben ja den Rat für Sozial- und Wirtschaftsdaten genannt, Rat für Informationsinfrastrukturen, die Deutsche Forschungsgemeinschaft und andere Forschungsinstitutionen haben Meinungen oder auch Expertise eingebracht. Es sind aber auch andere Interessenverbände, -träger wie die Privatwirtschaft und andere, die sich da an der Diskussion beteiligen. Wichtig ist, dass aus meiner Sicht, wir haben jetzt die ganze Zeit darüber geredet: Forschungsdaten. Wir müssen Forschungsdaten definieren, und das muss flexibel definiert werden. Ich habe ja vorhin ganz bewusst auch das Beispiel des Nobelpreises 2021 gesetzt, weil die haben mit Daten geforscht, die es vor 20, 30 Jahren gar nicht gab. Und das sind auch überwiegend Daten, die nicht aus der Forschung kommen. Also haben ja die Behördendaten genannt, und also administrative Daten. Die werden ja nicht für die Forschung produziert, sondern für andere Zwecke, und die Forschung ist Zweitverwender. Da sollte man offen sein, welche Daten die Forschung in Zukunft zweitverwendet. Wir können das Fass jetzt mit privat-erhobenen Daten aufmachen. Andere Punkte, Sie haben schon Remote Access und Zugangsmöglichkeiten genannt. Die sollten auch geregelt werden, wie natürlich aber auch das Löschen von Daten. Weil durch den Datenschutz werden auch Daten nach einem bestimmten Zeitpunkt gelöscht: Manche nach 10, 20 oder 30 Jahre, andere, wie Hilfsmerkmale, zum Beispiel zur Volkszählung, nach nur wenigen Jahren. Diese Daten sind allerdings wiederum für die Forschung von großer Relevanz, und wenn man diese als Forschungsdaten definieren würde, bestimmte Anonymisierungen durchführen, kann man ja, verschiedene Anonymisierungsverfahren, da sollten diese Daten dann noch länger gespeichert werden.

**[pogg]:** Das heißt, es wird auch zu viel weggeschmissen hier und da im behördlichen Alltag?

**[Bender]:** Da ist noch nicht so viel weggeschmissen worden oder sowas. Aber, sagen wir so, die Gefahr besteht - also konkret: Volkszählungen sind Hilfsmerkmale

weggeschmissen worden. Also, und da gab es Daten wie ein Adressregister, wurde mühsam aufgebaut, was dann jetzt gelöscht werden musste. Oder man hat dort auch Daten zusammengeführt und das wurde dann zum Beispiel beim Namen durchgeführt, und dann gab es Synonymtabellen, wo man dann sagt, wenn man einen Namen schreibt, kann man den auch bestimmt falsch schreiben oder man kann ihn dann in einer anderen Form schreiben. Stephanie zum Beispiel Steffi. Meines Wissens sind auch diese Informationen gelöscht worden. Ist auch wieder schlecht, weil man dann immer wieder bei Null anfangen muss. Das ist ja auch irgendwie Verschwendung von Geld und Arbeitszeit.

**[mg]:** Ich denke gerade auch so ein bisschen an Geld und Arbeitszeit und andere Ressourcen. Sind da die Infrastrukturen da, um, ich sag jetzt mal ein bisschen provokativ, vielleicht auf Verdacht, mal alles aufzuheben? Ist das einfach so möglich? Muss man sich nur dafür entscheiden oder müssten da erstmal auch Archive aufgebaut werden?

**[Bender]:** Also es gibt schon gesetzliche Vorgaben, wann was zu löschen ist. Das hängt ja auch wiederum mit der Zweckbindung ab. Also den Datenschutz muss man da natürlich auch wieder im Blick haben. Eine Diskussion zum Beispiel: Wenn ich gestorben bin, wie lange sollten denn dann bestimmte Sachen von mir noch aufgehoben werden?

**[pgg]:** Ein Beispiel ist ja das Verkehrssünderregister. Das fällt mir sofort ein, dass nach einiger Zeit da die Verstöße oder die Punkte gelöscht werden. Ich weiß gar nicht, ob die wirklich gelöscht werden, aber bei solchen Dingen, so Negativeinträgen ist das mit der Löschfrist ja sehr plausibel.

**[Bender]:** Das stimmt. Auf der anderen Seite muss man natürlich sagen, wenn da eine gewisse Verhaltensdisposition man hinbekommt, um zu sagen, ich fahre ständig bei Rot über die Ampel, kann das natürlich auch bestimmte Indikatoren für Risikofreudigkeit oder bestimmte andere Punkte sein, die man vielleicht dann wiederum für bestimmte Analysen verwenden möchte. Mir tut es weh, wenn ich sage: Och ich schmeiß irgendwas weg. Ich hätte das lieber, dass diese Daten dann in irgendeinem Modell, sagen wir in einem Datentreuhändermodell, irgendwo sicher lagern und aber der Personenbezug nicht mehr hergestellt werden kann. Das heißt also, man könnte sich dann vorstellen, es gibt irgendwo eine zentrale Stelle, die alle Daten von uns sammelt, aber dann, die muss ja dann auch bestimmte Sicherheitsbestimmungen und bestimmte Grundlagen und so erfüllen. Aber das ist in gewisser Weise, wenn ich den Data Governance Act richtig verstanden habe, ja eine der Grundideen jetzt, dass es da auch solche zentralen Stellen geben soll.

**[pgg]:** Kann man sagen, es gibt irgendwo eine Grenze dessen, was die Forschung interessiert? Oder zumindest eine Grenze sollte es geben, was die Forschung bekommen kann an Daten? Oder ist das eigentlich aus Ihrer Sicht risikolos möglich,

dass eine Gesellschaft alle ihre öffentlichen Daten, also Behördendaten, Public-Sector-Daten der Forschung gibt?

**[Bender]:** Ich habe ja vorhin etwas zu Datenschutz und zu Vertrauen gesagt. Also mir ist es wichtig, dass diese zwei Punkte eingehalten werden. Das heißt, für Bundesbank kann ich sagen: Wir haben so detaillierte Daten zu Banken oder auch teilweise zu Unternehmen, dass wir sehr nahe an Geschäftsgeheimnisse kommen und man auch zum Beispiel Strategien von Unternehmen oder Banken ableiten könnte. Da würde ich eine Grenze setzen wollen. Deswegen ja auch die Zweckbindung dieser Daten, wobei man die Zweckbindung weiter fassen könnte, als sie jetzt zurzeit ist. Wenn man sagt, für nichtkommerzielle Forschung als eine Zweckbindung, als bei einigen Gesetzen eine neue Zweckbindung, würde das sehr hilfreich sein, auch zur Weitergabe. Ich denke auch, wir schauen hier zu arg nur auf Gesetze. Die Konstruktion von zum Beispiel Datentreuhänder oder Forschungsdatenzentren ermöglichen natürlich einen zusätzlichen Schutz über die Grenzen von Gesetzen hinaus. Also ich denke, die Konstruktionen, die wir hier jetzt haben, helfen, dass weiterhin Vertrauen in diese Institutionen bestehen, als auch, dass der Datenschutz eingehalten wird. Und ich habe ja auch gesagt, kommerzielle Forschung oder auch die Kommerzialisierung von diesen Daten, die jetzt öffentlich erhoben sind, sehe ich schon Grenzen darin.

**[pgg]:** Wie wäre das mit Datenjournalismus? Das ist ja auch ein Feld, das sich entwickelt im Moment. Ist es dann keine Forschung, keine Wissenschaft? Wenn man so will, sind Massenmedien ja auch ein Markt.

**[Bender]:** Ich kann die Bestrebungen der Datenjournalisten verstehen, es ist ein Für und Wider. Zum einen finde ich es wichtig, habe ich auch gesagt, dass nach wissenschaftlichen Kriterien geprüfte Informationen aus diesen Daten auch in die Öffentlichkeit kommen. Auf der anderen Seite habe ich allerdings auch einen gewissen Zugang zu regeln und einen wichtigen Punkt, als wir das mal untersucht hatten oder darüber diskutiert haben, ist: Was ist eigentlich ein Datenjournalist? Bei unabhängiger Forschung können wir das klar definieren, was da ein Forscher ist, weil sie oder er ist an einer Uni beschäftigt oder an einem, sagen wir, zertifizierten Forschungsinstitut. Ein Journalist – tue ich denen vielleicht Unrecht – kann man sehr leicht werden, wenn es da nicht eine Arbeitsteilung zwischen Datenjournalisten und Forschung gibt. Was man sich ja auch vorstellen kann: Warum muss ein Datenjournalist an die Mikrodaten der Bundesbank kommen? Ist die Frage: Kann er nicht an Zwischenergebnisse kommen, die die Forschung zur Verfügung stellt? Wenn er oder sie aber an die Daten kommen muss, muss geregelt werden, was einen Datenjournalisten ausmacht. Das hört sich jetzt sehr bürokratisch an, aber da sehe ich dann eine gewisse Grenze oder plädiere da eher auch für eine Art von Arbeitsteilung. Die Daten auszuwerten braucht man ja auch ein gewisses Grundwissen. Daten sind nicht unkompliziert in ihrer Art und Weise, so dass man da doch auch gewisse Zeiten rein verwenden muss, bis man sie versteht und bis man sie dann auch zu einer sinnhaften Auswertung führt. Da weiß ich nicht, ob sich das Datenjournalisten auch wirklich antun wollen.

**[pgg]:** Das mit der Arbeitsteilung hat sicherlich Charme für beide Seiten, denn Journalisten haben wenig Zeit und wären sicher dankbar, wenn bestimmte Kärnerarbeit jetzt nicht jedes Mal selbst gemacht werden müsste, sondern wenn man da auf was zugreifen kann.

**[Bender]:** Und bei den Forschenden ist klar, dass: Der Forschende ist nicht am Einzelfall interessiert. Die Schwierigkeit beim Datenjournalisten ist, dass er vielleicht, wenn er die Daten dann sieht, plötzlich dann doch ein Interesse an einem Einzelfall hat. Da kann auch sehr schnell das Interesse sich ändern. Also wenn ich dann glaube: Och, ich sehe in den Bundesbankdaten irgendeinen Verstoß von irgendeiner Bank oder Unternehmen, kann man mir nicht sagen, dass man das nicht dann rausträgt. Also das ist dann auch noch mal eine gewisse Schwierigkeit. Der Forschende ist am Einzelfall nicht interessiert, der ja dann viel mehr Schlagzeilen bringt als jetzt irgendein nebulöser Zusammenhang. Das ist dann der gleiche Punkt, den wir machen bei kommerzieller Forschung, wenn man dann sagt: Och, ich schau doch mal, was mein Mitbewerber eigentlich macht, weil er das sieht, also in bestimmten Daten, jedenfalls bei uns. Also deswegen ist das eine spannende Diskussion.

**[mg]:** Sehr spannendes Thema. Sollten wir vielleicht mal einen anderen Podcast zu machen.

**[pgg]:** Das wird sicher ein Zukunftsthema sein. Das glaube ich schon.

**[mg]:** Aber was die Gesetzgebungsvorhaben betrifft, Sie haben ja vorhin beschrieben, dass im Einzelfall schon Einigungen zustande kommen zwischen Datenproduzenten und auch Wissenschaftlern, wo dann auch beide Seiten den Nutzen sehen und einen Nutzen haben. Was würde ein Gesetz denn darüber hinaus für Vorteile bringen?

**[Bender]:** Wenn man sich die deutsche Gesetzeslandschaft anschaut, haben wir ja verschiedene Gesetze für verschiedene Sachverhalte. Also ein Beispiel: Das Bundesstatistikgesetz regelt, wie das Statistische Bundesamt mit Daten umzugehen hat. Das Sozialgesetzbuch regelt, wie die Deutsche Rentenversicherung, die Krankenversicherung und die Bundesagentur für Arbeit mit den Daten umzugehen hat. Und die Gesetze sind unterschiedlich, obwohl da teilweise gleiche Sachverhalte drin sind. Es werden Personendaten verwendet oder es werden Unternehmensdaten genommen. Aber dadurch, dass das unterschiedliche Rechtskreise sind, gibt es dafür unterschiedliche Regelungen. Also Beispiel: Im Sozialgesetzbuch ist die Verknüpfung von Personendaten möglich. Das wird auch dann durchgeführt. Ich habe ja gesagt, die sozialversicherungspflichtige Beschäftigung wird mit der Arbeitslosigkeit, mit aktiver Arbeitsmarktpolitik verknüpft. Im Bundesstatistikgesetz gibt es die Personenverknüpfung nicht. Das heißt, das Statistische Bundesamt zum Beispiel hat spezifische Daten zu Personen, bekommt auch von anderen spezifische Daten zu Personen, zum Beispiel die Steuerdaten. Aber kann dann bestimmte Daten von sich auf der Personenebene nicht miteinander verknüpfen. Noch schöner, was wiederum bei dem Sozialgesetzbuch möglich ist und das IAB hat das auch durchgeführt, ist die

Verknüpfung von Personendaten mit Unternehmensdaten. Das heißt dann sogenannte Linked-Employer-Employee-Datensätze. Waren mal sehr en vogue und wahrscheinlich immer noch in den Wirtschaftswissenschaften, weil sie dann den individuellen Erwerbsverlauf haben und dann auch sehen, wie das Unternehmen quasi agiert oder die Bedingungen des Unternehmens. Also wenn Sie zum Beispiel Glück hatten, in ein Unternehmen zu kommen, was gerade floriert, und Sie haben da eine bessere Einkommensentwicklung oder machen dort auch eine bessere Karriere. Und es ist ein Nachteil, wenn Sie in eine Firma kommen, die dann plötzlich unerwartet Pleite macht, was ja dann ein wirklicher Treffer in ihrer Erwerbsbiografie ist. Also das ist da möglich auf der einen Seite, auf der anderen Seite nicht. Also es wäre schön, wenn zumindestens für den Bereich Forschung am deutschen Forschungsdatengesetz bestimmte Regelungen vielleicht einheitlich und zusammengeführt werden. Was mir wichtig ist, und das hat der Rat für Informationsinfrastrukturen gesagt: Man sollte durch ein neues Gesetz nicht hinter die jetzigen Gesetzesregelungen fallen. Da gibt es so eine Tendenz in Deutschland, dass man dann sagt: Wir haben jetzt diese Gesetze, und jetzt machen wir noch ein Gesetz obendrauf, und dann wird es noch komplizierter. Also das heißt, man sollte aufpassen, dass, wenn man jetzt da neue Gesetze entwickelt, dass die Klarheit schaffen, nicht die Situation, obwohl man sie verbessern wollte, dann insgesamt verschlechtern.

*[Der Abspann mit Musik beginnt.]*

**[mg]:** Und damit sind wir am Ende dieses Digitalgesprächs angekommen und bedanken uns bei Stefan Bender von der Deutschen Bundesbank für das spannende Gespräch und die interessanten Eindrücke. Viele Grüße nach Frankfurt am Main! Und natürlich wie immer auch vielen Dank an Sie, liebe Zuhörerinnen und Zuhörer, für Ihr Interesse und die Aufmerksamkeit. Wenn Sie mögen, hören wir uns wieder in drei Wochen zur nächsten Folge des Digitalgesprächs, einem Podcast von ZEVEDI, dem Zentrum verantwortungsbewusste Digitalisierung.



This work is licensed under CC BY-NC-ND 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>