Deutscher Bundestag

Ausschuss Digitale Agenda

Ausschussdrucksache
19(23)107

Stiftung Neue Verantwortung

Stellungnahme von Aline Blankertz, Stiftung Neue Verantwortung für die öffentliche Anhörung des Ausschusses Digitale Agenda am 24. Februar 2021 zum Thema „Datenstrategie der Bundesregierung“ (BT-Drs. 19/26450) verbunden mit „Eckpunkte einer Datenstrategie der Bundesregierung“ (BT-Drs. 19/16075) und dem Antrag der Fraktion der FDP „Datenpolitik für Selbstbestimmung, Wettbewerb und Innovation“ (BT-Drs. 19/26538)

In meiner Stellungnahme beziehe ich mich indirekt auf den interfraktionellen Fragenkatalog (ohne Nummerierung). Mein Schwerpunkt liegt auf Abschnitt 2 der Datenstrategie, der Datennutzung.

Die fünf wichtigsten Punkte zusammengefasst:

a. Für die Zielsetzung einer Dateninfrastruktur bedarf es messbarer Ziele, die abstrakte Ziele wie digitale bzw. Datensouveränität klar operationalisieren oder ersetzen.

b. Datentreuhandmodelle haben großes Potenzial, das insbesondere im Zusammenhang mit Verbraucher:innendaten noch stärker gefördert werden sollte, um über Personal Information Management Systems (PIMS) hinausgehende Modelle zu testen (z.B. in Reallaboren) und rechtlich zu ermöglichen (z.B. durch Delegierung von Datenrechten).

c. Das Nutzbarmachen von Daten, ohne vertrauliche Elemente (des Personenbezugs oder Geschäftsgeheimnisse) preiszugeben, ist ein wichtiger Forschungsbereich, der praxisnah erforscht werden sollte.

d. Neue und bestehende Formen des Datenteilens sollten immer von einer Risikoanalyse begleitet werden. Mit entsprechenden Maßnahmen zur Risikoverringerung können und sollten neben Forschung und Zivilgesellschaft auch die Wirtschaft als Adressat zusätzlicher Daten stehen.

e. Die Datenstrategie selbst zeigt deutlich die Herausforderungen der Politik und Verwaltung für einen offenen Umgang mit Daten, wie z.B. im intransparenten Prozess der Konsultation deutlich wurde.

1. Allgemein

Ich begrüße die Datenstrategie der deutschen Bundesregierung prinzipiell. Sie enthält viele teils lange überfällige Themen und Maßnahmen. Leider hat sie wenig strategischen, zukunftsweisenden und fokussierenden Charakter und setzt damit nicht die nötigen Impulse, die von einer Strategie zu erhoffen sind.

Konkret: Von den 234 Maßnahmen im Anhang laufen 62% bereits, sodass nur gut eine von drei Maßnahmen überhaupt einen explizit planenden Charakter haben. Hinzu kommt, dass viele der Maßnahmen darin bestehen, dass „geprüft“, „gefördert“, „unterstützt“, „sich […] eingesetzt“ oder „weiter“ „fortgesetzt“ wird, oder Projekte ohne klare Handlungsanweisung oder Zielsetzung aufgeführt werden. Auf nur 37% der Maßnahmen in den Themenbereichen 1-3 trifft dies nicht zu, d.h. sie beschreiben eine klar definierte Aktivität. Im Themenbereich 4,

Staat als Vorreiter, sind immerhin 73% der Maßnahmen klar definierte Aktivitäten, was angesichts der handelnden Rolle des Staates in diesem Bereich einleuchtet.

Ebenfalls fällt stark auf, dass die Datenstrategie weder messbare Ziele noch messbaren Aufwand beinhaltet. Wie es besser geht, zeigt z.B. die britische Datenstrategie:[1] Sie ist deutlich evidenzbasierter und liefert damit auch eine Grundlage, an der der Erfolg der Strategie gemessen werden kann. Es ist fast in sich paradox, eine Datenstrategie zu formulieren, ohne zeitgleich eine entsprechende Datenerhebung zur Evaluierung umzusetzen.

2. **Dateninfrastruktur**

Es ist begrüßenswert, dass die Bundesregierung sich der Dateninfrastruktur widmet als eine Voraussetzung für eine funktionierende Datenökonomie. Dabei kommt **dem Staat auch eine Rolle zu als Förderer einer Infrastruktur**, die eine gesellschaftlich wünschenswerte Datennutzung ermöglicht.

Ein kritischer Aspekt ist allerdings, welche Ziele mit dieser Infrastruktur verfolgt werden. Souveränität ist in diesem Kontext ein häufiger verwandter, jedoch ambivalenter Begriff: Er steht gerade für territoriale Herrschaftslogiken und nicht für einen wertebasierten, offenen Datenmultilateralismus.[2] Abstrakte Ziele bedürfen einer sauberen Operationalisierung, damit klar definiert ist, **welche Ziele mit der Infrastrukturförderung verfolgt werden und welche Indikatoren ermöglichen, den Erfolg zu messen**. Das gilt insbesondere für das Großprojekt GAIA-X, das so stark politisch aufgeladen wurde, dass kaum ein Unternehmen nicht dabei sein möchte. Die konkreten Ziele von GAIA-X sind nicht klar definiert und umfassen gewisse Datenstandards und eine Cloud-Interoperabilität. Allerdings ist unklar, inwiefern diese Themen Debatten z.B. über die Herkunft der mitwirkenden Unternehmen rechtfertigen, was oft mit dem Ziel der Souveränität assoziiert wird.

Ein **offensichtlicher Anwendungsfall für GAIA-X wäre die Nutzbarmachung von Daten des Staates**, der sich in Bezug auf seine Daten modernisieren und öffnen will (was dem Abschnitt 4 der Datenstrategie vorgreift). Es ist kurios, dass die Bundesregierung einerseits ein abstraktes Cloud-Projekt fördert, während sie andererseits unsicher ist, ob dieses einen Mehrwert für bestehende oder geplante Cloud-Unterfangen wie die Bundescloud oder die föderale Verwaltungscloud bietet. Diese scheinbare Unkoordiniertheit bleibt durch die Datenstrategie ungelöst.

3. **Datennutzung**

Interessen von Verbraucher:innen, Delegation von Datenrechten und PIMS

Der Datenschutz reicht nicht aus, um eine Art der Nutzung von Daten zu ermöglichen, die Verbraucher:inneninteressen berücksichtigt. Dieser Aspekt wird von der Datenstrategie in Abschnitt 2.4 zwar treffend analysiert, jedoch sind die **Maßnahmen bei Weitem unzureichend**, um spürbare Veränderungen herbeizuführen.

Aktuell haben Verbraucher:innen **keine Werkzeuge, um ihre Datenrechte und Interessen effektiv durchzusetzen**, denn allein der Aufwand, um Datenschutzbestimmungen zu

[1] UK National Data Strategy, 9. Dezember 2020.
[2] Thiel, Thorsten (2021): "Das Problem mit der digitalen Souveränität", Frankfurter Allgemeine Zeitung, 26. Januar.

verstehen, stellt eine Überforderung von Individuen dar.[3] Von den Personal Information Management Systems (PIMS) wird gesprochen, als seien sie bereits eine praktikable Form, Datenrechte bzw. Verbraucher:inneninteressen durchzusetzen. Jedoch sind PIMS bisher allenfalls eine Randerscheinung mit wenig messbarem Erfolg bzw. Durchschlagskraft für die weitere Datenökonomie. Die einzelnen Anbieter sind weit davon entfernt, einen den großen Plattformen vergleichbaren Wachstumspfad einzuschlagen.[4] Dies liegt vermutlich u.a. daran, dass PIMS nicht in der Lage sind, für eine Masse an Nutzer:innen attraktive Dienste anzubieten, weswegen sie auch keine Abkehr von etablierten Datensammelpraktiken ermöglichen.

Da PIMS aktuell nur sehr begrenzt in der Lage sind, die schwache Position der Verbraucher:innen zu verbessern, ist es mindestens **fraglich, ob sie durch zusätzliche Regulierung weiter in ihrem Handlungsspielraum eingeschränkt werden sollten**. Der Vorschlag für das Daten-Governance-Gesetz der Europäischen Kommission sieht eine Anforderung der expliziten Nicht-Nutzung der Daten (im Rahmen der Auflagen an Dienste für die gemeinsame Datennutzung) bzw. der Nicht-Kommerzialität (im Rahmen des Datenaltruismus) vor.[5] Dabei ist nicht klar, ob sich aktuell problematische Tendenzen durch eine mögliche Zweitverwertung oder Kommerzialität der Angebote ergeben.

Wenn PIMS breit angeboten und angenommen werden sollen, ist mindestens in gleichem Maße zu zusätzlichen Auflagen eine **ermöglichende Regulierung** zu erwägen. Ein wichtiger Aspekt hierfür ist, dass Verbraucher:innen die Ausübung von (mindestens manchen) **Datenrechten an PIMS delegieren** können.[6] Ein häufiger Einwand ist, dass dies mit der Datenschutzgrundverordnung nicht vereinbar sei.[7] Die Grenzen der Delegierbarkeit im Rahmen der DSGVO abzustecken liegt außerhalb meiner Expertise; jedoch ist es wichtig, neben der rechtlichen Prüfung zu untersuchen, welche Formen der Delegierung von Datenrechten für Verbraucher:innen nützlich sein können. Hierbei können **Reallabore** eine wichtige Rolle spielen, um in einem abgegrenzten und überwachten Bereich mögliche Regulierungsausnahmen zu testen. Bestätigt sich die Nützlichkeit einer Delegierung, sollte das ein Grund sein, den bestehenden Rechtsrahmen einschließlich der DSGVO kritisch zu überdenken.

Im Zusammenhang mit **Verbraucher:innendaten besteht ein großes Potenzial von Datentreuhandmodellen**, zu denen oft auch PIMS gezählt werden. Allerdings sind auch weitere Modelle denkbar, die **erweiterte Funktionalitäten und Befugnisse** mit sich bringen könnten. So könnte eine Treuhand für Verbraucher:innendaten perspektivisch eine kollektive Verhandlungsfunktion übernehmen und damit die Verhandlungsposition von Verbraucher:innen stärken.[8] Solche Modelle können Teil einer strategischen Untersuchung

---

[3] Zum Beispiel dauert das Lesen einer deutschen Datenschutzerklärung für Sprachassistenten im Durchschnitt 15 Minuten, siehe Kettner, Sara Elisa, Christian Thorun und Jan-Peter Kleinhans (2018): „Big Data im Bereich Heim und Freizeit Smart Living“.

[4] So sind viele Beispiele einer umfassenden Marktstudie aus dem Jahr 2017 (Stiftung Datenschutz: „Neue Wege bei der Einwilligung im Datenschutz – technische, rechtliche und ökonomische Herausforderungen“) inzwischen nicht mehr am Markt, während bestehende Lösungen zwar ihre Nutzer:innenzahlen nicht veröffentlichen, doch bisher wenig Aufmerksamkeit auf sich ziehen.

[5] Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über europäische Daten-Governance (Daten-Governance-Gesetz), 25. November 2020, insbesondere Artikel 11 und 16.

[6] Blankertz, Aline (2020): „Designing Data Trusts. Why We Need to Test Consumer Data Trusts Now“.

[7] Funke, Michael (2020): „Die Vereinbarkeit von Data Trusts mit der Datenschutzgrundverordnung (DSGVO)“.

[8] Siehe Blankertz, Aline (2020), op. cit..

sein, über welche Wege Verbraucher:innen am besten effektivere Werkzeuge erhalten können, um ihre Interessen in der Datenökonomie durchzusetzen.

Der Entwurf des **Data Governance Acts** beschäftigt sich mit einer Reihe von Modellen von Datenintermediären und Datentreuhändern, darunter auch die Idee des Datenaltruismus/der Datenspende. Diese kann als die simpelste Version einer Datentreuhand angesehen werden, da sie aufgrund der Abwesenheit einer monetären Dimension keine Bedenken über den „Verkauf" von Privatsphäre weckt. Jedoch stellt sich auch bei diesem Modell das Problem, dass es nur mit weiteren Auflagen versehen wird, während eine entsprechende Erleichterung anderer Anforderungen nicht erfolgt.[9] Dabei ist zu erwarten, dass v.a. solche Modelle weitere Verbreitung finden werden, zu deren Adaption mehr Anreize geschaffen werden als Hürden.

Risiken der Datennutzung, Personenbezug und Verschlüsselung

Es gibt **Risiken, die von der Erhebung bzw. Nutzung von Daten ausgehen, die keinen oder nur indirekten Personenbezug** haben. Die Datenstrategie nimmt diese Risiken zwar in einer Maßnahme („Verhinderung von Diskriminierung von Menschen durch die Auswertung aggregierter/anonymer/synthetischer Daten") in den Blick, doch räumt der Thematik nicht den nötigen Raum ein. Die Beschränkung auf die Frage, ob ein Personenbezug vorliegt und somit ein Missbrauchsrisiko, verliert sich leicht in der Frage, wo genau der Personenbezug ende. Diese Frage abschließend zu klären, ist angesichts komplexer Mischdatensätze kompliziert. Zudem geht sie am Ziel vorbei, da auch ohne Personenbezug Missbrauchsrisiken vorliegen. Diese können entstehen, wenn aufgrund von Korrelation selbst anonymisierte Daten Aufschluss über sensible Informationen von Individuen geben,[10] oder wenn z.B. (größere) Unternehmen den Großteil des Werts von Daten von anderen (kleineren) Unternehmen für sich beanspruchen können und damit den datengebenden Unternehmen schaden können,[11]

Die Analyse der Risiken von Daten ohne Personenbezug ist wichtig, um **nötige Schutznahmen zu treffen, die über den Datenschutz hinausgehen**. In Indien arbeitet die Regierung gerade an der zweiten Iteration eines Gesetzesvorschlags zur Datengovernance von nicht-personenbezogenen Daten, die weltweit erstmals „Community-Daten" als rechtlich relevant anerkennt.[12] Es ist absehbar, dass solche Risiken präsenter werden: Vorstöße beim sogenannten Federated Learning sollen das Trainieren von Algorithmen ermöglichen, ohne dass Daten geteilt werden bzw. indem der Algorithmus zu lokal gespeicherten Daten geschickt wird. Das ermöglicht ein datenschutzkonformes Algorithmentraining, doch verhindert nicht, dass Algorithmen sensible Daten aufnehmen oder Diskriminierung lernen. Beispielsweise ist Federated Learning (of Cohorts)[13] ein Ansatz, über den Google personenbezogene Cookies ersetzen und dennoch gezielte Werbung ermöglichen will. Auch

---

[9] Veil, Winfried (2020): „Datenaltruismus: Wie die EU-Kommission eine gute Idee versemmelt", CR-online.de-Blog, 1. Dezember.

[10] Bspw. Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian und Asuman Ozdaglar (2019): „Too Much Data: Prices and Inefficiencies in Data Markets", NBER working paper.

[11] Wie es Gegenstand ist z.B. des aktuellen Wettbewerbsverfahrens der Europäischen Kommission gegen Amazon wegen Amazons Nutzung der Daten der Verkäufer:innen, siehe Pressemitteilung der Europäischen Kommission vom 10. November 2020.

[12] Indian Ministry of Electronics and Information Technology (2020): "Report by the Committee of Experts on Non-Personal Data Governance Framework", 16. Dezember.

[13] GitHub-Eintrag zu "Federated Learning of Cohorts (FLoC)".

wenn keine individuellen Daten geteilt werden, bestehen weiterhin erhebliche Risiken, dass Werbung gezielt manipulativ bzw. auf Mitglieder bestimmter Gruppen zugeschnitten wird.

Verfahren, um Datensets ohne Preisgabe vertraulicher Datenpunkte nutzbar zu machen, sind hilfreich, um eine risikoarme Datennutzung zu ermöglichen. Dies gilt für Vertraulichkeit in Bezug auf sowohl den Personenbezug als auch Geschäftsgeheimnisse.[14] Daher ist es sehr zu begrüßen, dass hierzu eine bessere Vernetzung stattfinden soll, die auch die internationale Forschung einschließen sollte. Es ist sinnvoll, auf eine breit gefasste Entwicklung von Verfahren zu setzen, da verschiedene Verfahren verschiedene Datenrisiken adressieren. Eine Förderung von Forschungsvorhaben sollte neben Forschungsinstituten auch praktische Anwendungen mit anderen kommerziellen oder nicht-kommerziellen Organisationen einbeziehen.[15]

Zielgruppen für breiteren Datenzugang

Die Bundesregierung legt ein besonderes Augenmerk auf eine gemeinwohlorientierte Nutzung von Daten. Dabei sollte die Definition des Begriffs Gemeinwohl nicht zirkulär verstanden werden als das, was durch Forschung und zivilgesellschaftliche Aktivitäten entsteht. Denn **auch wirtschaftliche Organisationen tragen zum Gemeinwohl bei,** z.B., wenn sie die Entwicklung von Covid-19-Impfstoffen vorantreiben, die einen hohen gesellschaftlichen Nutzen über den kommerziellen Profit des einzelnen Unternehmens hinaus bringen.[16]

Ein breiteres Verständnis von Gemeinwohl spricht dafür, **Daten eher breiter zugänglich zu machen, während mögliche Risiken der Datennutzung systematisch gering zu halten sind**. Dementsprechend sollten (öffentliche oder andere) Daten auch großen Digitalkonzernen zugänglich gemacht werden: Auch wenn diese möglicherweise bessere Möglichkeiten haben, Wert aus den Daten zu generieren, sollte dies – im Interesse ihrer Nutzer:innen – nicht bewusst unterbunden werden. Ein Beispiel: Google ist eher in der Lage, Mobilitätsdaten über verschiedene Quellen zu erheben und einzukaufen. Alternative Mobilitätsdienste profitieren stärker davon, dass Daten von öffentlichen Verkehrsdiensten leichter zugänglich gemacht werden. Wenn Daten für alle besser verfügbar sind, können Wettbewerber und auch z.B. zivilgesellschaftliche Akteure einfacher entsprechende Datenkompetenzen aufbauen.

4. **Datenkompetenz**

Es ist begrüßenswert, dass die Bundesregierung das Potenzial von Zivilgesellschaft in der Vermittlung von Datenkompetenzen erkennt. Allerdings halte ich es für wünschenswert zu hinterfragen, ob das die die einzige wesentliche Rolle für die Zivilgesellschaft sein soll. Außer in Abschnitt 3 wird die Zivilgesellschaft lediglich als eine der Adressat:innen der Datenöffnung (neben Wissenschaft und Wirtschaft) in Abschnitt 4 erwähnt. Es gibt bereits einige zivilgesellschaftliche **Organisationen, deren Aktivitäten eher unter (die Ermöglichung von) Datennutzung fallen als unter die Vermittlung von Datenkompetenz**, wie z.B. die Open Knowledge Foundation (insb. mit dem OK-Lab), und

[14] Leider sind die Begriffe Anonymisierung (da nicht vollständig möglich oder sinnvoll) und Depersonalisierung (da es auch um Geschäftsgeheimnisse gehen kann) nicht hilfreich, um zu erfassen, was mit den Daten geschieht. Deswegen sind neue Begriffe nötig, siehe Schäfer, Christin (2021): „Depersonalisierung – eine Sprachkritik“, Tagesspiegel Background Digitalisierung & KI, 29. Januar.

[15] Beispielsweise engagiere ich mich aktuell als Teil des Vorstands von SINE Foundation e.V., die an Open-Source-Verschlüsselungsmodulen arbeitet, um nachhaltiges Datenteilen zu ermöglichen.

[16] Blankertz, Aline (2020): „Die Datenökonomie aus gesellschaftlicher Perspektive denken“, Heinrich-Böll-Stiftung, 17. Dezember.

CorrelAid. Eine Förderung der Zivilgesellschaft in der Datennutzung ist ebenfalls wünschenswert.

5. **Staat als Vorreiter**

Open Data und Bereitstellung von Daten durch den Staat sind keine neuen Themen. Es ist unkontrovers, dass eine breitere Bereitstellung von öffentlichen Daten sinnvoll ist. Leider greift die Datenstrategie bei der Analyse der Probleme zu kurz. Ein wesentliches Problem bei der Bereitstellung von Daten ist nämlich, dass **Verwaltungsstrukturen so statisch und verschlossen sind**, dass sie schwer mit einer offenen und dynamischen Bereitstellung von Daten vereinbar sind. Allerdings sind die aus diesen Strukturen erwachsenden Herausforderungen deutlich größer als die unzureichende Datenoffenheit, weswegen sie auch nicht im Rahmen allein einer Datenstrategie zu lösen sind.

Es sei nur angemerkt, dass sich diese **Verschlossenheit auch bei denen zeigt, die die Datenstrategie zu verantworten haben**: Denn nicht nur wurden erst auf externes Drängen die Konsultationsfragen in einem Dokument zugänglich gemacht, auch die Antworten auf die Konsultation wurden in keiner Weise veröffentlicht, obwohl dies in anderen Ländern durchaus üblich ist. Die fehlende Transparenz verhindert demokratische Rechenschaft und untergräbt auch die Glaubwürdigkeit der eigenen Maßnahmen, um den Staat als Vorreiter für offeneres Datenteilen zu etablieren.