Oktober 2016 · Julia Manske

Offene Daten und der Schutz der Privatsphäre

Anregungen zur systematischen Integration von Datenschutzprinzipien in Open Data

Stiftung
Neue
Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Executive Summary

Die Autorin bedankt sich bei Dr. Alexander Dix, Walter Palmetshofer und Dr. Carlo Piltz sowie beim Team der Stiftung Neue Verantwortung, insbesondere Dr. Tobias Knobloch, Dr. Stefan Heumann und Jan-Peter Kleinhans, für konstruktive Kritik und Unterstützung.

Wovon viele Open-Data-Verfechter in Deutschland lange geträumt haben, scheint nun Wirklichkeit zu werden: Offene Daten werden langsam aber sicher politische Chefsache. Durch die Öffnung von Verwaltungsdaten schafft die Politik eine wichtige Grundlage für eine Dateninfrastruktur, von der Verwaltung, Unternehmen und Bürger in hohem Maße profitieren werden. Dies konnte man bereits in den Vorreiterländern beobachten. Bei der Umsetzung sollte Deutschland von den Erfahrungen dieser Länder lernen. Deutschland bietet sich die Chance, auch die Fehlentwicklungen im Bereich offene Daten zu vermeiden und von vornherein die Weichen bestmöglich zu stellen.

Dies betrifft insbesondere den Schutz der Privatsphäre bei der Öffnung von Verwaltungsdaten. Inzwischen häufen sich international Fälle, in denen Daten geöffnet wurden, die eine Re-Identifizierung von Personen in den Datensätzen zuließen. Die verantwortungsvolle Öffnung ist aber eine wesentliche Voraussetzung für den langfristigen Erfolg des Open-Data-Vorhabens. Gerade das datenschutzsensible Deutschland muss bei der Öffnung seiner Datenbestände den Datenschutz als integralen Bestandteil konsequent mitdenken, um das Vertrauen der Bürger nicht aufs Spiel zu setzen. Gleichzeitig würden aber auch andere Länder von einem deutschen Datenschutzrahmen für Open Data profitieren.

Für die Implementierung eines solchen Ansatzes sind die Vorzeichen hierzulande sehr günstig: In kaum einem Land ist der Datenschutz stärker verankert als in Deutschland. Diese Stärke gilt es für den Umgang mit offenen Daten zu nutzen und mit diesen Erfahrungen auch die internationale Debatte zur systematischen Öffnung von Regierungsdaten anzureichern.

Dabei zeigt sich, dass fünf Punkte für eine verantwortungsvolle Datenöffnung wichtig sind. Diese sollen als I**mpulse für zukünftige Arbeitspakete im Bereich Open Data & Privacy** dienen, die es von der Bundesregierung in den nächsten Monaten in Zusammenarbeit mit Datenschützern, Technikern und der Zivilgesellschaft auszugestalten gilt.

1. Das **Prinzip des Abwägens** zwischen dem öffentlichen Informationsinteresse und dem Schutz der Privatsphäre, das bereits aus dem Informationsfreiheitsdiskurs bekannt ist, muss systematisiert, harmonisiert und auf offene Daten übertragen werden. Ergebnisse dieser Abwägungsentscheidung sollten bindend festgelegt werden können. Dadurch würden Verwaltungsmitarbeiter bei der Entscheidung, welche Daten bedenkenlos veröffentlicht werden können, entlastet.

2. Die **Prinzipien der Europäischen Datenschutzgrundverordnung (EU-DSGVO)** sowie des dementsprechend zu novellierenden deutschen Datenschutz-

rechts sollten in der Open-Data-Strategie der Bundesregierung sowie im Entwurf eines Open-Data-Gesetzes verankert werden. Daraus resultiert ein Öffnungsansatz, bei dem zu veröffentlichende Datensätze zunächst einer groben Prüfung des Datenschutzrisikos unterzogen werden. Dieser gezielt beschränkte Öffnungsansatz ist sachgemäß als open by design zu bezeichnen. Darüber hinaus sollten die Übernahme der „Informierten Einwilligung" bei der Veröffentlichung bestimmter Regierungsdaten und die Integration eines technischen Datenschutzes (privacy by design) in Open-Data-Plattformen zur Regel gemacht werden.

3. **Starke Anonymisierungsverfahren** sollten fester Bestandteil von Datenveröffentlichungen werden. Dafür braucht es leicht verständliche Handreichungen und einschlägige Fortbildungen für Verwaltungsmitarbeiter, die im Austausch mit den Datenschutzbehörden erarbeitet werden sollten. Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit sollte entsprechende Anleitungen, die auch den Landesbehörden als Orientierung dienen können, anfertigen lassen.

4. Es gilt, einen verstärkten **Fokus auf die Limitierung von Anonymisierungsverfahren** zu legen. Einerseits umfasst das die **umfangreiche Aufklärung der Datennutzer** über die Grenzen der Anonymisierung auf den Datenbereitstellungsportalen selbst. Andererseits setzt dies Investitionen in die Erforschung und Entwicklung rechtlicher und technischer Lösungen einer wirksamen Anonymisierung voraus.

5. Die Diskussion muss, über den klassischen Datenschutz hinaus, die **potenziellen Nutzungsszenarien von Verwaltungsdaten und ihren möglichen Missbrauch** thematisieren. Zusammen mit sowohl der Datenschutz- als auch der Open-Data-Community muss die Regierung Regularien dazu erarbeiten, wie der Missbrauch dieser Daten eingeschränkt und eventuell auch sanktioniert werden kann, ohne dabei generell das Vorhaben, Offenheit zu fördern, zu gefährden.

Inhaltsverzeichnis

Vom Nachzügler zum Vorreiter

In den letzten Monaten ging ein deutlicher Ruck durch Deutschlands Open-Data-Bestrebungen. Aus einer Nische ist es nach ganz oben auf die digitalpolitische Agenda gerückt. Das ist sehr erfreulich, denn offene Daten[1] bergen ein enormes Potenzial. Sie können helfen, Verwaltungshandeln effizienter zu gestalten, bessere politische Entscheidungen zu treffen, Start-ups und andere Innovationstreiber aus einer reichen Datenquelle schöpfen zu lassen. Außerdem fördern sie die Zusammenarbeit zwischen Politik, Verwaltung und externen Akteuren, wodurch schneller Lösungen für die gesellschaftlichen Herausforderungen unserer Zeit entwickelt werden können.

Während Länder wie England, die USA oder Frankreich dies schon seit längerem erkannt haben, will Deutschland nun aufholen: Das Bundesministerium für Verkehr und digitale Infrastruktur hat erst kürzlich ein eigenes Metadatenportal für Mobilitäts-, Geo- und Wetterdaten aufgebaut[2] und stellt nun einen Fördergelder[3] für die Entwicklung von Open-Data-Anwendungen auf dieser Grundlage zur Verfügung. Das Bundesministerium für Wirtschaft und Energie veranstaltet Workshops und Konferenzen mit der Start-up-Szene[4], und das Innenministerium erarbeitet ein Open-Data-Gesetz[5].

Die Zeichen stehen gut, dass deutsche Verwaltungsdaten zukünftig auch in der Fläche geöffnet werden. Dabei kann Deutschland nun davon profitieren, sich dem Thema zunächst nur zögerlich gewidmet zu haben. Denn es muss das Rad nicht neu erfinden, sondern kann von den vielen erfolgreichen Ansätzen und Beispielen aus dem Ausland lernen.[6] Genauso wertvoll ist jedoch, dass die deutsche Bundesregierung auch aus den Fehlern anderer Länder lernt und diese vermeidet. **Deutschland bietet sich insofern nun die Chance, ganze Entwicklungsetappen im Bereich offene Daten zu überspringen und von vornherein die Weichen bestmöglich zu stellen.**

Einer dieser Bereiche, in denen Deutschland nicht nur aufholen, sondern sich nach wie vor als Vorreiter positionieren kann, ist die Integration angemessener Standards zum Schutz der Privatsphäre. Zwar findet der Datenschutz

1 Offene Daten zeichnen sich dadurch aus, dass sie für jedermann und für jegliche Zwecke genutzt, (maschinell) weiterverarbeitet und weiterverbreitet werden können. In diesem Papier beziehen wir uns auf offene Verwaltungs- und Regierungsdaten, wie etwa Umwelt- und Wetterdaten, Geodaten, Verkehrsdaten, Haushaltsdaten, Statistiken, Publikationen, Protokolle, Gesetze, Urteile und Verordnungen.

2 http://mcloud.de/

3 http://www.bmvi.de/DE/DigitalesUndRaumentwicklung/DigitaleAgenda/Modernitaetsfonds/modernitaetsfonds_node.html

4 Zum Beispiel den Kongress „Open Data – Potenziale für die Wirtschaft“ zusammen mit der Geoinformationswirtschaft.http://www.bmwi.de/DE/Service/Veranstaltungen/dokumentationen,did=751072.html

5http://www.sueddeutsche.de/wirtschaft/digitale-verwaltung-auf-schatzsuche-im-amt-1.3065789

6 Vgl. zum Beispiel unser Policy Brief Knobloch, T.; Manske, J. (2016). Das Datenzeitalter gestalten. Offene Daten als Schlüssel. http://www.stiftung-nv.de/publikation/das-datenzeitalter-gestalten; Einzelne Erfolgsbeispiele finden sich außerdem auf www.datenwirken.de, auf der Plattform http://odimpact.org oder in diversen Publikationen zu Best-Practice-Ansätzen auf http://theodi.org/publications.

meist eine Erwähnung, wenn über offene Daten gesprochen wird, doch de facto wurde dieses Feld von Regierungen und Open-Data-Communitys eher stiefmütterlich behandelt. Das Ergebnis: Nicht zuletzt aufgrund steigender Risiken für den Datenschutz durch die maschinellen Verarbeitungsmöglichkeiten herrscht vielerorts Unklarheit, welche Daten geöffnet werden dürfen und welche auf keinen Fall veröffentlicht werden sollten. Teilweise wurden Daten veröffentlicht, die nicht intendierte Rückschlüsse auf Personen zuließen. Derweil wächst der Graben zwischen Datenschutz-Verteidigern auf der einen und Open-Data-Befürwortern auf der anderen Seite.[7] Obgleich zumindest in Forschungskreisen die Aufmerksamkeit für die Problematik steigt[8], findet sie bislang kaum Eingang in einschlägige Open-Data-Foren und Konferenzen.

Doch wenn dem Schutz der Privatsphäre nicht hinreichend Aufmerksamkeit geschenkt wird, riskiert die Bewegung, das Vertrauen der Bürger zu verlieren oder möglicherweise sogar gesetzliche Grenzen zu verletzen. Insbesondere für die weitere Ausarbeitung der Open-Data-Agenda auf internationaler Ebene, etwa im Rahmen der International Open Data Charta[9] oder der Open Government Partnership[10] müssen insofern dringend neue Standards erarbeitet werden. Deutschland ist für seinen ausgereiften Datenschutz bekannt und kann insofern genau für diese Entwicklungen einschlägige Impulse liefern. Insbesondere kann die Bundesrepublik mit gutem Beispiel vorangehen, indem sie den Schutz der Privatsphäre bei der Ausgestaltung ihrer Open-Data-Strategie systematisch berücksichtigt. **Dieses Papier möchte einen kurzen Überblick darüber geben, welche Herausforderungen aktuell im Spannungsfeld von offenen Daten und dem Schutz der Privatsphäre bestehen. Zudem möchte es Impulse dafür geben, welche Aktivitäten in Zukunft angestoßen werden sollten, um adäquate Lösungen für diese Herausforderungen zu entwickeln.**

---

7 Vgl. zum Beispiel die Debatte zwischen Kate Craword und Open-Data-Verfechtern https://twitter.com/katecrawford/status/742471535064154116 oder einen Blogpost als Reaktion auf die Position des Open-Data-Verfechters Tim O'Reilley http://whimsley.typepad.com/whimsley/2011/09/data-anonymization-and-re-identification-some-basics-of-data-privacy.html

8 Tran, E.; Scholtes, G. (2015). Open Data Literature Review. https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview_2015-04-14_1.1.pdf; Zuiderveen Borgesius, F.; van Eechoud, M; Gray, J. (2015). Open Data, Privacy and Fair Information Principles: Towards a Balancing Framework. In: Berkeley Technology Law Journal, Forthcoming; Institute for Information Law Research Paper No. 2015-04; Amsterdam Law School Research Paper No. 2015-46. http://ssrn.com/abstract=2695005. Altman, M; Wood A., O'Brien, D., Vadhan S. , and Gasser, U. (2016). Towards a Modern Approach to Privacy-Aware Government Data Releases. In: Berkeley Journal of Technology Law. http://openscholar.mit.edu/sites/default/files/dept/files/modernopendataprivacy.pdf; Whittington J; Calo, R.; Simon, Mike; Woo, J.; Young, M.; Schmiedeskamp, P. (2015). Push, Pull and Spill: A Transdisciplinary Case Study in Municipal Open Government Seattle. In: Berkeley Technology Law Journal, Forthcoming University of Washington School of Law Research Paper No. 2015-23. http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf; Personal Data and Privacy Working Group der Open Knowledge Foundation http://personal-data.okfn.org/; Kapitel zu Privacy and Data Protection im Open Government Guide http://www.opengovguide.com/topics/privacy-and-data-protection/

9 http://opendatacharter.net/

10 http://www.opengovpartnership.org/

Zu betonen ist, dass der Datenschutz keinesfalls als Argument gegen die Öffnung instrumentalisiert werden darf.[11] Im Gegenteil sollte es vielmehr darum gehen, den Schutz der Privatsphäre als Rahmen zu nutzen, um eine verantwortungsvolle Öffnung von Verwaltungsdaten beharrlich voranzutreiben.

Datenschutz als Vertrauensbasis

Als die britische Regierung in 2013 die Plattform care.data eröffnete, verfolgte sie das Ziel, Gesundheits- und Sozialdaten auf einer zentralen Plattform zusammenzuführen, um so Forschung zu fördern und das angeschlagene britische Gesundheitssystem effizienter zu machen.[12] Doch das Vorhaben krankte seit Beginn an der Ablehnung der britischen Bevölkerung. Denn die Bürger, deren sensible Gesundheitsdaten in pseudonymisierter Form bereitgestellt wurden, waren weder über die Prozesse informiert, noch wurden sie zu Beginn um Einwilligung gebeten. Über Nacht erhielten somit nicht nur Forscher und Wissenschaftler, sondern auch kommerzielle Anbieter Zugriff auf ihre Daten.[13] Datenschützer, aber auch Ärzte kritisierten das Projekt. Bürger, die über die Medien von immer neuen Skandalen erfuhren, fühlten sich unzureichend informiert und ihrer aktiven Zustimmung beraubt. Ihr Vertrauen in die sinnvolle Nutzung von Gesundheitsdaten hatte die Regierung auf diese Weise verloren, noch bevor das Projekt richtig starten konnte. Vor wenigen Wochen wurde care.data offiziell eingestellt.[14]

In Bhutan stellte die Regierung im Rahmen ihrer Open-Data-Aktivitäten Datensätze von Bewerberinnen und Bewerbern der öffentlichen Verwaltung bereit, um Transparenz zu fördern. In den Datensätzen enthalten waren aber auch Telefonnummern und Adressen – mit der Folge, dass die weiblichen Bewerber von interessierten Männern kontaktiert und belästigt wurden.

Diese Fälle verdeutlichen, dass der Schutz der Privatsphäre wesentlich ist, um das Projekt „Offenheit“ zu einem langfristigen Erfolg zu führen.

Über die Informationsfreiheit hinaus: Offene Daten werfen neue Fragen zum Datenschutz auf

Die Diskussion zwischen Offenlegung, Transparenz und Datenschutz ist

11 Vgl. dazu etwa eine IFG-Anfrage zu Kulturdenkmalen in Freiburg, die mit Verweis auf den Datenschutz abgelehnt wurde https://fragdenstaat.de/anfrage/liste-der-kulturdenkmale-im-regierungsbezirk-freiburg/

12 Das Portal cara.data ist der Definition nach keine reine Open-Data-Plattform, da es zunächst einer Registrierung bedarf, um Zugang zu den Datensätzen zu erhalten. Allerdings ist das Registrierungsverfahren relativ niedrigschwellig. So war von Beginn an undurchsichtig, welche Akteure (Forschung, Privatwirtschaft) welche Nutzungsrechte haben würden.

13 Ramesh, R. (2014). NHS patient data to be made available for sale to drug and insurance firms. In: The Guardian, 19. Januar 2014. https://www.theguardian.com/society/2014/jan/19/nhs-patient-data-available-companies-buy

14 Boseley, S. (2016). NHS to scrap single database of patients‘ medical details. In: The Guardian, 6. Juli 2016.https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details

keinesfalls neu. Im Zuge der Informationsfreiheitsbewegung war der Datenschutz bei der Weitergabe personenbezogener Daten immer wieder Bestandteil von Debatten bis hin zu Anlass für richterliche Entscheide.[15] **Auch wenn das Recht auf Privatsphäre historisch zumindest im europäischen Rechtsrahmen etablierter ist, werden das Recht auf Informationszugang und das Recht auf Privatsphäre als gleichermaßen relevant und nicht hierarchisch gesehen.**[16] Auch in Deutschland enthalten die meisten Informationsfreiheitsgesetze (IFG) Regelungen, „die die Herausgabe personenbezogener Daten aufgrund einer Einwilligung oder aufgrund des Überwiegens eines allgemeinen Informationsinteresses vorsehen."[17] Im Spannungsfeld von Informationsfreiheit und Schutz der Privatsphäre hat sich das Prinzip der Abwägung zwischen dem öffentlichen Interesse an Information und dem Schutz der Persönlichkeitsrechte des einzelnen bewährt.[18]

Zwischen den Prinzipien, die hinter der Idee der Informationsfreiheit und jenen von offenen Daten stehen, gibt es zahlreiche Überschneidungen. Und doch ist es mit Blick auf den Schutz der Privatsphäre wichtig, ihre Unterschiede zu verdeutlichen: Hinter dem Prinzip der Informationsfreiheit steht ein Rechtsanspruch. Dieser soll dabei helfen, die Transparenz über Prozesse zu erhöhen und es der Zivilgesellschaft oder Journalisten erlauben, Regierungen oder Privatwirtschaft zur Rechenschaft ziehen zu können. Viele Anfragen, die im Rahmen von Informationsfreiheitsgesetzen gestellt werden, gelten dem Abrufen von Dokumenten (zum Beispiel von Protokollen, Verträgen). Aber auch Datensätze können Gegenstand von IFG-Anfragen sein oder dem Zweck der Transparenz und Rechenschaft dienen. Etwa wenn sie Haushaltsdaten, Daten über Beschaffung, Gehälter und Zuwendungen oder Lobbyregister enthalten. IFG-Anfragen vorausgeht aber in der Regel eine klare Fragestellung, die die Regierung oder Verwaltung in der Rechen-

15 Vgl. hierzu auch die Stellungnahme der Artikel-29-Datenschutzgruppe zur Veröffentlichung personenbezogener Daten für Transparenz im öffentlichen Sektor. Article 29 Data Protection Working Party (2016). Opinion 02/2016 on the publication of Personal Data for Transparency purposes in the Public Sector.http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2016/wp239_en.pdf

16 Janssen, K.; Hugelier, S. (2013). Open Data: A New Battle in an Old War? In: Hildebrand et al.: Digital Enlightenment Yearbook 2013. The Value of Personal Data, S. 192. Bedauerlicherweise zeichnet sich in der Praxis allerdings nach wie vor ein anderes Bild ab: So wurde im Fall des Google-Urteils durch den Europäischen Gerichtshof dem Schutz der Privatsphäre grundsätzlich Vorrang einräumt. Obwohl auf das Prinzip der Abwägung verwiesen wird, deutet das Urteil darauf hin, dass im Zweifel der Datenschutz vorgeht.

17 Schnabel, C. (2012). Der Schutz personenbezogener Daten bei informationsfreiheitsrechtlichen Ansprüchen nach § 11 HmbIFG. In: DuD – Datenschutz und Datensicherheit, 7/2012, S. 520-525. https://www.datenschutz-hamburg.de/uploads/media/Datenschutz_bei_IFG-Anspruechen_-_Aufsatz_in_DuD-Heft-7-2012.pdf

18 Auch wenn der Eindruck entstehen mag: Der Schutz der Privatsphäre ist keineswegs ausschließlich ein individuelles Recht. Er ist ein öffentliches Gut, das essentiell für funktionierende Demokratien ist. Insofern geht es nicht um den Ausgleich zwischen öffentlichem und individuellem Interesse, sondern zwischen zwei Rechten, die beide der Allgemeinheit dienen. Vgl. O'Hara, K. (2011). Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office. London, GB, Cabinet Office, S. 13f. http://eprints.soton.ac.uk/272769/3/272769OHARA11.pdf; Banisar, D. (2011). The Right to Information and Privacy: Balancing Rights and Managing Conflicts. World Bank Institute Governance Working Paper. http://ssrn.com/abstract=1786473

schaftspflicht sieht. Die Antworten werden zudem lediglich dem Anfragenden zur Verfügung gestellt und können so gegebenenfalls nur begrenzt veröffentlicht oder weiterverarbeitet werden[19]

Offene Daten umfassen ausschließlich Datensätze, die in standardisierter, maschinenlesbarer und automatisch weiterverarbeitbarer Form veröffentlicht werden[20]. Sie ermöglichen nicht nur einen punktuellen Zugang zu Informationen, sondern sie erbringen theoretisch einen vorab noch nicht absehbaren Mehrwert für eine Vielzahl an Akteuren. Etwa ermöglichen sie es Bürgern, ein umfangreiches Verständnis für das Regierungshandeln und von gesellschaftlichen Strukturen zu entwickeln.[21] Deshalb eignen sie sich auch so gut als Quelle für Innovationen und können ebenso Ausgangspunkt für wirtschaftlichen Mehrwert sein. Das Wesensmerkmal offener Daten – und das ist ein entscheidender Unterschied zur Informationsbereitstellung nach dem Grundsatz der Informationsfreiheit – ist also erstens, dass sie in der Regel **proaktiv von der Regierung bereitgestellt werden** (darauf aber kein Rechtsanspruch besteht). Zweitens können (und sollen) sie von vielen Akteuren in vielfacher Weise weitergenutzt werden. Und drittens kann das theoretisch denkbare Nutzungsspektrum dieser Daten bei Veröffentlichung noch unklar sein und sich erst in der Nutzungspraxis vollumfänglich offenbaren.

Offene Daten als Teil eines größeren Datenspektrums

Darüber hinaus zeichnet sich international ein Trend ab, nach dem offene Daten gewissermaßen in eine zweite Phase eingetreten sind.[22] Blickt man etwa in die Open-Data-Vorreiterländer USA oder Großbritannien, so sollen Regierungsdaten, in hoher Frequenz und Umfang bereitgestellt, als Quelle für umfangreiche Datenanalyse dienen. Einerseits dienen sie der Regierung selber, indem sie als **Grundlage für politische Entscheidungen** dienen (evidence-based decision making). Wie etwa in Los Angeles, um potenzielle Kriminalitätshochburgen zu identifizieren und den Einsatz von Polizisten entsprechend zu steuern.[23] Oder in Neuseeland, wo Gesundheits-, Zensus- und Sozialabgabedaten kombiniert und ausgewertet werden, um den Bedarf an

19 Anders nur in den Ländern, deren Informationsfreiheitsgesetze dem Prinzip „Acces for one is access for all" folgen, zum Beispiel das Schweizer Öffentlichkeitsgesetz. Diese „Erga-omnes-Wirkung" gilt auch für die EU-Transparenzverordnung 1049/2001. Vgl. hierzu für Deutschland auch das Gesetz über die Weiterverwendung von Informationen öffentlicher Stellen. http://www.gesetze-im-internet.de/iwg/index.html

20 Vgl. zum Beispiel die Definition der Open Knowledge Foundation http://opendatahandbook.org/guide/en/what-is-open-data/

21 Vgl. O'Hara, K. (2011). Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office. London, GB, Cabinet Office, S. 20. http://eprints.soton.ac.uk/272769/3/272769OHARA11.pdf

22 Hier beschrieben als „Second Wave"; Sollazzo, G. (2015). Open data: where the movement started and where it's headed.,Computer World UK http://www.computerworlduk.com/data/open-data-where-it-started-where-its-headed-3626537/; aber auch Scassa, T. (2014). Privacy and Open Government. In: Future Internet. 6/2014, S. 397-413. www.mdpi.com/1999-5903/6/2/397/pdf

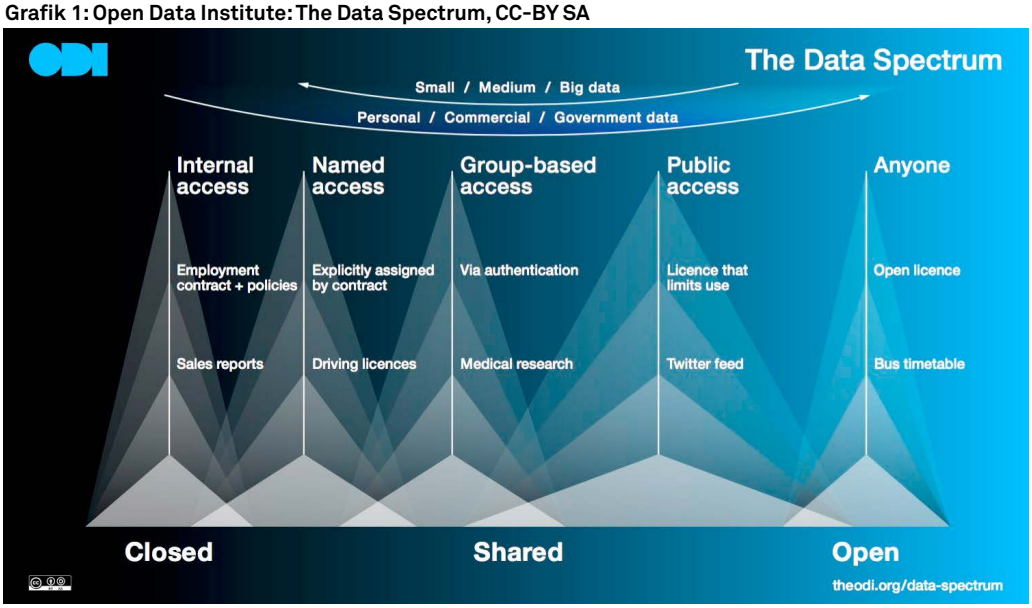
23 van Rijmenam, M. (2016). The Los Angeles Police Department Is Predicting and Fighting Crime With Big Data. https://datafloq.com/read/los-angeles-police-department-predicts-fights-crim/279

präventiven Maßnahmen gegen Kindesmissbrauch von Neugeborenen vorherzusagen.[24] Oder in San Francisco, wo Unfall-, Straßen-, und Verkehrsdaten für ein besseres Verkehrsmanagement genutzt werden.[25] Diese Beispiele verdeutlichen, dass der Mehrwert für die interne Nutzung besonders dann steigt, wenn Daten unterschiedlicher Quellen in einer zentralen Datenbank zusammengeführt werden können.

Andererseits werden Regierungsdaten aber auch verstärkt zu einer **Quelle externer Innovationen, insbesondere wenn diese mit unterschiedlichen Datenquellen** kombiniert werden. Das britische Start-up Geolytix beispielweise verknüpft offene Regierungsdaten mit Kundendaten und ermittelt so ideale Standorte für den Einzelhandel. Von dieser Warte aus betrachtetet, überrascht es also nicht, dass vermehrt Anreize für die privatwirtschaftliche Nutzung und Weiterverarbeitung von offenen Regierungsdaten geschaffen werden.

Diese Entwicklungen führen aber auch dazu, dass Verwaltungsdaten **Teil eines größeren Datenökosystems (oder Datenspektrums, siehe Grafik 1), bestehend aus Daten unterschiedlicher Quellen und Arten, werden.**[26] Und dies wiederum hat Implikationen für den Schutz der Privatsphäre.

Grafik 1: Open Data Institute: The Data Spectrum, CC-BY SA



24 http://www.msd.govt.nz/about-msd-and-our-work/publications-resources/research/predicitve-modelling

25 „Data-Driven Policy“: San Francisco just showed us how it should work. https://medium.com/@abhinemani/data-driven-policy-san-francisco-just-showed-us-how-it-should-work-c7725e0e2b40

26 Datenspektrum des Open Data Institute: The Data Spectrum, CC-BY SA, http://theodi.org/data-spectrum

Abwägen zwischen öffentlichem Interesse und Datenschutz

Generell lässt sich aus den etablierten Verfahren der Informationsfreiheitsbewegung, in Deutschland zuletzt auch im Zuge der Umsetzung des Hamburger Transparenzgesetzes[27], viel für eine datenschutzsensible Implementierung von Open-Data-Prinzipien lernen. Dies bietet sich nicht zuletzt deswegen an, weil Datenschutz und Informationsfreiheit hierzulande traditionell jeweils in ein und derselben Behörde angesiedelt sind.[28] Zum Beispiel lassen sich die Prinzipien der Abwägung direkt auf die Veröffentlichung als offene Daten übertragen. Dies zeigte sich etwa an der Entscheidung der Europäischen Kommission, die Datensätze zur Agrarsubventionen mit Personenbezug zu veröffentlichen.[29] Hier wurde zunächst entschieden, dass das öffentliche Interesse an diesen Datensätzen höher ist als der Schutz der Namen der Beteiligten, obgleich diese befürchteten, dass die Daten dann auch von kommerziellen Anbietern oder Konkurrenten genutzt werden können. Allerdings zeigt dieses Beispiel auch die Schwierigkeiten: Nach wie vor herrschen etwa im internationalen Vergleich große Inkonsistenzen in der Entscheidungsfindung vor.[30] Der Richtlinie der Europäischen Kommission zum Trotz nahm etwa die Bundesregierung jüngst ein Anpassung vor, nach dem die Daten deutscher Agrarsubventionsempfänger, die in IFG-Verfahren regelmäßig als Rohdaten angefordert wurden, lediglich für zwei Jahre zur Verfügung gestellt werden dürfen, während jede Nachnutzung nach diesem Zeitpunkt illegal ist.[31] Und selbst innerhalb nationaler Grenzen gelten unterschiedliche Bewertungskriterien: Auf eine Anfrage zur Veröffentlichung der geschäftlichen Telefonnummern von Jobcenter-Mitarbeitern, um Prozesse für Hartz-IV-Empfänger zu erleichtern, wurden diese nach richterlichem

27 Vgl. dazu Schnabel, C. (2012), Der Schutz personenbezogener Daten bei informationsfreiheitsrechtlichen Ansprüchen nach § 11 HmbIFG. In: DuD – Datenschutz und Datensicherheit, 7/2012, S. 520-525, hier S. 525: Die befürchteten Konflikte „zwischen Datenschutz und Informationsfreiheit sind nicht eingetreten. Beide Rechtsgüter sind keineswegs unversöhnliche Gegensätze, sondern zwei grundlegende Prinzipien der Informationsgesellschaft, die nur in einigen Fällen in Widerstreit geraten. Für diese Fälle hat der Gesetzgeber grundsätzlich gut handhabbare Regelungen geschaffen, die von der Rechtsprechung in sinnvoller Weise weiter ausgestaltet wurden."

28 Dass dies auch kritisch zu bewerten ist, zeigt die Ausstattung der Bundesdatenschutzbehörde: Auf 94 Planstellen für den Datenschutz, kommen gerade einmal vier Stellen, die mit der Informationsfreiheit betraut sind. Vgl. Kommentar von Semsrott zum aktuellen Tätigkeitsbericht zur Informationsfreiheit. https://irights.info/artikel/vosshoff-datenschutz-informationsfreiheit/27601 Zudem besteht in Deutschland weiterhin die Tendenz, den Datenschutz zu priorisieren. Entsprechend § 5 des Informationsfreiheitsgesetzes darf der „Zugang zu personenbezogenen Daten (...) nur gewährt werden, soweit das Informationsinteresse des Antragstellers das schutzwürdige Interesse des Dritten am Ausschluss des Informationszugangs überwiegt oder der Dritte eingewilligt hat. Besondere Arten personenbezogener Daten im Sinne des § 3 Abs. 9 des Bundesdatenschutzgesetzes dürfen nur übermittelt werden, wenn der Dritte ausdrücklich eingewilligt hat."

29 Vgl. hierzu das Urteil des Europäischen Gerichtshofs. http://www.lto.de/recht/hintergruende/h/eu-agrarsubventionen-zwischen-transparenz-und-datenschutz/

30 Pagallo, U.; Bassi, E. (2013). Open Data Protection: Challenges, Perspectives, and Tools. In: Hildebrand et al.: Digital Enlightenment Yearbook 2013. The Value of Personal Data; S.182.

31 Vgl. hier Schreiben vom 30.12.2015 der Bundesanstalt für Landwirtschaft und Ernährung auf eine IFG-Anfrage durch die Plattform fragdenstaat.de. https://fragdenstaat.de/anfrage/empfanger-des-eu-agrarfonds-2014/#nachricht-26443

Beschluss in einigen Ländern freigegeben, in anderen nicht.[32] Hier steht das Grundsatzurteil des Bundesverwaltungsgerichts noch aus.[33]

Diese Inkonsistenzen zeigen, dass klarere Bewertungsmuster helfen könnten, um zwischen dem Schutz der Privatsphäre und der Transparenz abzuwägen.[34] **Im Bereich offene Daten fehlt es noch deutlicher an standardisierten Verfahren. Mitarbeiter des öffentlichen Dienstes werden mit der Entscheidung, wann Daten geöffnet werden können und sollten, alleingelassen. Diese Unsicherheit führt nicht zuletzt in Deutschland dazu, dass viele Datensätze gar nicht geöffnet werden.**[35] Erschwerend kommt hinzu, dass sie bei Fehlveröffentlichung dienstrechtlich belangt werden können.[36]

Die Kategorisierung von Datensätzen, also die Investition in gute Metadaten, wäre dafür zumindest eine hilfreiche Grundlage – auch um zu erläutern, warum bestimmte Daten nicht geöffnet werden.[37] Nun, da durch das Open-Data-Gesetz des Bundes bald eine Umstellung auf open by default zu erwarten ist, die eine standardisierte Öffnung von Verwaltungsdaten vorsieht, ist es umso wichtiger, klare Schemata darüber zu erarbeiten, welche Daten von der Öffnung ausgeschlossen werden müssen. **In diesem Sinne wäre der Begriff open by design, wie er jüngst von der Stadtverwaltung Seattles nach einer Revision der Open-Data-Aktivitäten verabschiedet wurde,[38] eigentlich sachangemessener. Danach sind nicht alle Daten grundsätzlich offen, sondern sie durchlaufen zunächst ein Prüfungsverfahren auf mögliche Datenschutzrisiken (Privacy Impact Assessment).** Klar ist, dass bei diesen Überlegungen die Datenschutzbehörden von Anfang an eingebunden werden müssen.

---

32 Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (2016). Tätigkeitsbericht zur Informationsfreiheit für die Jahre 2014 und 2015. 5. Tätigkeitsbericht, S.39ff. http://www.bfdi.bund.de/SharedDocs/Publikationen/Taetigkeitsberichte/TB_IFG/5TB06_16.pdf?__blob=publicationFile&v=2; dazu das entsprechende Urteil des Oberverwaltungsgerichts in NRW, vgl. Telefonliste des Jobcenters bleibt geheim. Plattform: Legal Tribune Online 17. Juni 2015. http://www.lto.de/recht/nachrichten/n/ovg-nrw-urteil-8-a-2429-14-telefonliste-jobcenter-kein-anspruch-informationsfreiheitsgesetz/

33 In einem ähnlichen Fall entschied das BVerwG allerdings gegen die Veröffentlichung der Telefonnummern von Verwaltungsmitarbeitern. http://www.bverwg.de/entscheidungen/entscheidung.php?ent=291015B1B32.15.0 Vgl. dazu auch Piltz, C. (2015). Gericht: Namen und Kontaktdaten von Amtsmitarbeitern dürfen nicht im Internet veröffentlicht werden. Siehe De Lege Data, 8. Mai 2015. https://www.delegedata.de/2015/05/gericht-namen-und-kontaktdaten-von-amtsmitarbeitern-duerfen-nicht-im-internet-veroeffentlicht-werden/

34 Vgl. hier etwa die Handreichung des Information Commissioner's Office in Großbritannien: How to disclose information safely. https://ico.org.uk/media/1432979/how-to-disclose-information-safely.pdf

35 Vgl. auch die Studie: Wirtz, B. W.; Piehler, R.; Thomas, M-J.; Daiser, P. (2016). Resistance of Public Personnel to Open Government: A cognitive theory view of implementation barriers towards open government data. In: Public Management Review. 18/2016, Heft 9, S. 1335-1364.

36 Etwa aufgrund von Vertraulichkeit nach § 30 VwVfG im Rahmen des Dienstrechts; bei personenbezogenen Daten auch nach § 5 BDSG; § 201 StGB, selten aber ggf. auch Normen des BDSG (43,44). Mit der EU-DSGVO wird der Strafrahmen für Mitarbeiter ggf. noch höher.

37 Vgl. hier auch als Positivbeispiel die OGD-Cockpit der Stadt Bonn: http://ogdcockpit.bonn.de/index.php/OGD_Cockpit_Bonn; leider kann an dieser Stelle nicht hinreichend darauf eingegangen werden, wie relevant generell die Datenqualität und die Einhaltung von Datenstandards ist, um eine Systematisierung vorzunehmen und so beispielsweise eine Bewertung möglicher Datenschutz-Risiken vorzunehmen.

38 http://www.routefifty.com/2016/02/seattle-open-data-policy/126268/

Adaption eines Open-by-Preferences-Ansatzes in Seattle

Die Stadt Seattle verabschiedete im Februar ihre neue Open-Data-Strategie. Diese hatte sie zusammen mit der University of Washington und der Sunlight Foundation erarbeitet. Im Fokus der Revision der vorangegangenen Arbeit und Öffnung von Daten stand die Frage: Wie können mehr und relevantere Datensätze veröffentlicht werden. Das Team nutzte die Gelegenheit aber auch, um die möglichen Datenschutzrisiken der bestehenden Datensätze zu prüfen. Dabei stellte sich heraus, dass viele der bereits veröffentlichten Daten in Kombination Rückschlüsse auf Individuen zuließen. Insbesondere offene Daten auf Stadtebene sind meist granular oder enthalten Geo-Referenzen, was die Risiken erhöht. Insofern empfahl das Forschungsteam, den Schutz der Privatsphäre als elementaren Bestandteil in die Strategie zu integrieren. Das Ergebnis ist die neue Open-by-Preference-Strategie. Im Gegensatz zum vorigen Open-by-default-Ansatz werden nun alle zu veröffentlichenden Datensätze zunächst einer Datenschutz-Prüfung unterzogen. Ziel bleibt nach wie vor die weitestmögliche Öffnung von Datensätzen, um Transparenz und Innovation zu fördern. Dies geschieht nun allerdings weitaus strategischer und kontrollierter.

http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf

Damit ist es aber nicht getan. Die beschriebene Automatisierung und Möglichkeiten der Weiterverarbeitung stellen im Kontext offener Daten weitreichende Herausforderungen an den Datenschutz dar und erfordern die Beachtung potenzieller Datenschutzrisiken, die über die Erfahrungen mit der Informationsfreiheit hinaus gehen.

Grenzen einer Kategorisierung von Daten

Natürlich wäre es bequem, wenn sich Handreichungen und Schemata sich ausschließlich an Datentypen orientieren könnten. Genau das macht die Rechtsprechung beispielsweise im Zuge der Europäischen Datenschutzgrundverordnung (EU-DSGVO), die Daten anhand unterschiedlicher Risikostufen klassifiziert. Warum dieser Ansatz problematisch ist, wird im Folgenden erläutert.

Zunächst ist festzuhalten, dass in der Regel zwischen personenbezogenen Rohdaten, pseudonymisierten, anonymisierten Daten und nicht personen-

bezogenen Daten[39] unterschieden wird. Generell gilt, dass in Deutschland im Kontext offener Daten ohnehin ausschließlich Daten ohne Personenbezug veröffentlicht werden.[40] Keine Daten also, die mit vertretbarem Aufwand einer natürlichen Person zugeordnet werden können.[41] Wobei hier festzuhalten ist, dass zumindest in Deutschland der Auslegungsbereich weit ist und etwa IP-Adressen auch als personenbezogene Daten gelten.[42]

Folglich finden sich aktuell auf deutschen Open-Data-Portalen vorrangig Daten, die keinerlei Personenbezug aufweisen. Etwa Wetterdaten, Infrastrukturdaten, Schuldaten, Standortdaten (beispielsweise von Toiletten, Defibrillatoren, Kitas, Denkmälern, Behindertenparkplätzen etc.). Auch pseudonymisierte Daten werden bislang nicht veröffentlicht. Dies ist insofern zu begrüßen, da sich diese durch Heranziehen zusätzlicher Informationen mit niedrigem Aufwand ein Personenbezug herstellen lässt (nach Erwägungsgrund 23 der EU-DSGVO, dann wenn die Bestimmbarkeit der Person durch den Einsatz zusätzlicher Informationen hergestellt werden kann[43]).

International zeigt sich jedoch bereits eine andere Realität. Obgleich der ursprüngliche Fokus der Open-Data-Bewegung auf Daten ohne Personenbezug lag[44], werden inzwischen vielfach Datensätze mit direktem Personenbezug veröffentlicht. In vielen Ländern werden beispielsweise Unternehmensregister als offene Daten zur Verfügung gestellt[45] oder Informationen über Landrechte[46]. Daten also, die ebenfalls einen wichtigen Beitrag zu Accountability oder Korruptionsbekämpfung beitragen können. In den USA werden in einigen Städten Wählerregistrierungsdaten inklusive der Adresse

39 Zuiderveen Borgesius, F.; van Eechoud, M; Gray, J. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. In: Berkeley Technology Law Journal, Forthcoming; Institute for Information Law Research Paper No. 2015-04; Amsterdam Law School Research Paper No. 2015-46, S. 34. http://ssrn.com/abstract=2695005

40 Vgl. zum Beispiel die Aussage von Thomas Jarzombek, MdB, zur Überlegung eines Open-Data-Gesetzes in der Süddeutschen Zeitung: Jannasch, S. (2016). Auf der Schatzsuche im Amt". Süddeutsche Zeitung. http://www.sueddeutsche.de/wirtschaft/digitale-verwaltung-auf-schatzsuche-im-amt-1.306578; oder von Jens Köppen, MdB, im Interview mit dem Handelsblatt: Delhaes, D. (2016). Union und SPD öffnen das Milliardengeschäft mit Daten. http://www.handelsblatt.com/politik/deutschland/open-data-gesetz-union-und-spd-oeffnen-das-milliardengeschaeft-mit-daten/13840960.html

41 Vgl. § 3 Abs. 1 des Bundesdatenschutzgesetzes http://www.gesetze-im-internet.de/bdsg_1990/__3.html

42 Hierzu steht eine Entscheidung des EuGH aufgrund einer Vorlage des BGH allerdings noch aus.

43 Piltz, C. (2016). Die Datenschutz-Grundverordnung und Open Data.https://www.delegedata.de/2016/01/die-datenschutz-grundverordnung-und-open-data/

44 Vgl. Open Data Handbook: The key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals. http://opendatahandbook.org/guide/en/what-is-open-data/

45 Vgl. Plattform Open Corporates, ein umfangreiches Portal aller verfügbarer Unternehmensregister https://opencorporates.com/

46 Beispielsweise in Großbritannien über das nationale Datenportal https://data.gov.uk/ und auf dieser Plattform (aktuell im Beta-Stadium). http://landregistry.data.gov.uk/

und der politischen Ausrichtung der Bürger online zur Verfügung gestellt.[47] Gleiches gilt in den USA für die hierzulande umstrittenen Register von Sexualstraftätern.[48] Oder man betrachte die jährliche Offenlegung von jährlichen Steuerrückzahlungen auf individueller Ebene in Norwegen, die klare Rückschlüsse auf das Einkommen zulässt.[49] Natürlich wird es kulturell bedingt immer Unterschiede im Verständnis davon geben, welche Informationen schützenswert sind und welche einen wesentlichen gesellschaftlichen Nutzen erbringen. Denkbar – und mitunter wünschenswert – ist aber, dass auch in Deutschland die Forderung nach der standardmäßigen Veröffentlichung von Rohdaten mit Personenbezug zunehmen, wenn diese von hohem öffentlichen Interesse sind (zum Beispiel Lobbyregister, Landregister, Nebeneinkünfte von Politikern oder Richtern etc.).[50] Eine Ausnahme bot, zumindest bis vor kurzem, der bereits erwähnte Datensatz zu den EU-Agrarsubventionen, der zwar nicht proaktiv von der Regierung bereitgestellt wurde, aber durch den IFG-Mechanismus von der Zivilgesellschaft regelmäßig als Open Data aufbereitet wurde.

Kein Personenbezug, kein Datenschutzproblem?

Durch das Verfahren der Anonymisierung, also der Veränderung von Daten in einem Maße, dass diese nicht mehr einer Person zuzuordnen sind, kann Daten der Personenbezug entzogen und können diese dann veröffentlicht werden. Nach deutscher (und zukünftig europäischer) Rechtsprechung fallen anonymisierte Daten nicht mehr unter den datenschutzrechtlich abgesicherten Bereich. Eine Form der Anonymisierung ist etwa die Aggregierung personenbezogener Daten.[51] Bei dieser werden Personen nach einer Generalisierung von Merkmalswerten zusammengefasst (etwa durch Intervallwerte wie „20-25 Jahre“ bei der Angabe des Alters). Diese können darüber hinaus konsolidiert angegeben werden. Auf den deutschen Datenportalen finden sich viele Beispiele solch aggregierter und konsolidierter Datensätze, etwa der Bevölkerungsanteil mit Migrationshintergrund, die Aufteilung der Bevölkerung nach Altersgruppen, Wahlergebnisse nach Bezirk etc. Generell

47 Ethan, C. (2016). Why the D.C. government just publicly posted every D.C. voter's address online., Fusion, 14. Juni 2016. http://fusion.net/story/314062/washington-dc-board-of-elections-publishes-addresses/

48 https://www.nsopw.gov/?AspxAutoDetectCookieSupport=1

49 Collinson, P. (2016). Norway, the country where you can see everyone's tax returns. In: The Guardian, 11. April 2016. https://www.theguardian.com/money/blog/2016/apr/11/when-it-comes-to-tax-transparency-norway-leads-the-field

50 Hier ist allerdings zu erwägen, ob diese Daten tatsächlich als offene Daten (im Sinne der Definition) oder aber mit eingeschränkten Zugangs- und Verbreitungsrechten veröffentlicht werden sollten. Vgl. dazu Zuiderveen Borgesius, F.; van Eechoud, M; Gray, J. (2015). Open Data, Privacy and Fair Information Principles: Towards a Balancing Framework. Berkeley Technology Law Journal, Forthcoming; Institute for Information Law Research Paper No. 2015-04; Amsterdam Law School Research Paper No. 2015-46, S. 34. http://ssrn.com/abstract=2695005

51 Eine umfangreiche Übersicht und Bewertung gängiger Anonymisierungsverfahren wie etwa Randomisierung, k-anonymity, l-diversity, bietet die Article 29 Data Protection Working Party (2014). Opinion 05/2014 on Anonymisation Techniques. http://www.cnpd.public.lu/fr/publications/groupe-art29/wp216_en.pdf

ließen sich auf diese Weise auch andere ursprünglich granulare Datensätze, wie etwa Verbrechens- oder Verkehrsdaten, veröffentlichten. Allerdings verändert sich die Nutzbarkeit signifikant mit dem Aggregationslevel.[52] Denkt man etwa an Daten über Unfälle, so zeigt sich schnell, dass die aggregierten Daten der Unfallstatistik eines Jahres für eine Kommune zwar grundsätzlich von Interesse sein können. Der Erkenntnisgewinn und die Nachnutzbarkeit sind aber deutlich höher, wenn diese Daten beispielsweise auf Wochenbasis und für einen bestimmten Bezirk zur Verfügung stehen. **Insbesondere für Daten, die etwa im Kontext von Smart-City-Konzepten oder intelligenten Verkehrssystemen relevant werden, steigt der Nutzen mit der Granularität deutlich.**

Vielerorts werden Daten daher auf granularer Ebene veröffentlicht. Der Trend geht dahin, dies in immer kürzeren Abständen zu tun. Etwa die monatlich veröffentlichten Kriminaldaten[53], die durch die Polizei in Großbritannien veröffentlicht werden, oder Daten über Strafzettel54 und Taxifahrten in New York55. Bislang werden auf den deutschen Portalen kaum anonymisierte Daten veröffentlicht.[56]

Mit der Granularität der Daten steigen allerdings auch die Möglichkeiten, Rückschlüsse auf einzelne Personen ziehen zu können, was bis zur Re-Identifikation geht. Insbesondere dann, wenn diese Daten Georeferenzen, wie Postleitzahlen, GPS-Kennungen oder Kartenreferenzen, enthalten. Dies kann selbst dann der Fall sein, wenn ein Individuum gar nicht in den Daten auftaucht, oder eben wenn jemand zu einer Minderheit in den Datenpunkten gehört.[57] Hier denke man etwa an Geodaten: Selbst wenn diese aggregiert werden, können sie, beispielsweise in Regionen mit geringer Häuserdichte, recht genaue Rückschlüsse zulassen. Dies zeigt die Analyse von Daten aus 1,1 Milliarden anonymisierten Taxi- und Uber-Fahrten in New York: Aufgrund

52 Wie Zuiderveen Borgesius et. al. (2015) richtig anmerken, heißt dies zudem noch keineswegs, dass diese Daten, auch wenn sie keinen Personenbezug aufweisen, nicht dennoch negative Folgen für Einzelne haben können. Ein klassisches Beispiel ist die Auswertung von Verbrechensstatistiken, die dann dazu führt, dass in einer amerikanischen Stadt vermehrt Streifen in den Bezirk geschickt wurden, und die Bürger sich durch die erhöhte Polizeipräsenz beobachtet fühlten. Vgl. zum Beispiel Schmith, J. (2016). ‚Minority Report' Is Real – And It's Really Reporting Minorities. MicTech. https://mic.com/articles/127739/minority-reports-predictive-policing-technology-is-really-reporting-minorities#.ByMlHVk42

53 https://data.police.uk/

54 https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2016/kiv2-tbus

55 http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

56 Allerdings gibt es auch hierfür bereits Ausnahmen, etwa auf der Open-Data-Plattform der Stadt Moers, wo Datensätze über Bußgeldbescheide und fließenden Verkehr zugänglich sind, oder die Flugzeug-Unfalldaten auf mCloud des BMVI.

57 Hier sei erneut auf das Papier der Article 29 Data Protection Working Party (2014) verwiesen. Die Working Group unterteilt die Gefährdung nach dem Risiko des Singling Out, dem Herausgreifen oder Aussondern einer oder mehrerer Individuen in einem Datensatz, der Linkability, der Möglichkeit, Daten über ein Individuum aus verschiedenen Datensätzen zu verknüpfen, und der Inference, der Möglichkeit, aufgrund von Informationen andere relevante Informationen in einem Datensatz abzuleiten; S. 11f. http://www.cnpd.public.lu/fr/publications/groupe-art29/wp216_en.pdf

der Längen- und Breitenkoordinaten konnte ein Programmierer die exakten Adressen und das Taxi-Fahrverhalten einiger Kunden der reicheren, dünn besiedelten Vororte identifizieren, da ihre Häuser die einzigen im Umkreis waren.[58]

Es gilt insofern abzuwägen, wann welches Aggregationslevel sinnvoll ist. Dazu ein weiteres Beispiel: Daten, die über Kfz-Kennzeichenleser automatisiert gesammelt wurden, haben Behörden in Minneapolis zunächst für interne Zwecke genutzt. Die Polizei veröffentlichte nach Erlass einer neuen Open-Data-Gesetzgebung mindestens 2,1 Millionen Aufzeichnungen inklusive Datum, Ort und Standort der Wagen.[59] Diese Daten wurden jedoch auch von kommerziellen Anbietern, wie Fahrzeug-Zwischenhändlern aber auch Datenhändlern[60], genutzt. Diese Nutzung war weder von der Regierung noch von der Polizei intendiert. Vor allem aber deckte sie sich nicht mit der öffentlichen Vorstellung, warum Regierungsdaten veröffentlicht werden (hier stand klar das Argument Transparenz im Vordergrund). Mediale Beiträge führten dann zu Beschwerden der Bürger bezüglich potenzieller Datenschutzbedenken und schließlich zur Entfernung der Datensätze. Das öffentliche Interesse an den Informationen hätte in diesem Fall auch durch aggregierte Daten befriedigt werden können. So aber hat von der Datenveröffentlichung ausschließlich die Privatwirtschaft profitiert, nicht jedoch die Bevölkerung.[61]

Insbesondere mit Geo- und Standortdaten erhöht sich also das Risiko für den Datenschutz. Wenn die Daten unzulänglich anonymisiert sind, ist es beispielsweise ein Leichtes, Profile über einzelne Nutzer zu erstellen. Als in London das ÖPNV-Unternehmen Transport for London (TfL) die Daten über die Nutzung öffentlicher Fahrräder veröffentlichte, zeigte ein Hacker anhand dieser Daten, wie leicht die Route eines einzelnen Nutzers, und damit auch sein Wohnort, sein Arbeitsort bis hin zu Lebensgewohnheiten identifiziert

58 http://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance/#data-privacy-concerns

59 Farivar, C. (2012). Found: Secret location of Minneapolis police license plate readers, Ars Technica, 18. Dezember 2012. http://arstechnica.com/tech-policy/2012/12/found-secret-location-of-minneapolis-police-license-plate-readers/

60 Datenhändler (engl. data broker) sammeln Informationen aus diversen Quellen über Individuen und verkaufen diese in gebündelter Form an Unternehmen. Sie stellen diese Informationen Werbetreibenden für zielgerichtetes Marketing zur Verfügung, indem sie beispielsweise Profile über Nutzer erstellen. Datenhändler nutzen die Daten aber auch, um die Identität einer Person zu überprüfen oder das potenzielles Betrugsrisiko und die Kreditwürdigkeit eines Individuums zu bewerten. Kritiker, etwa Verbraucherschützer, werfen der Branche Intransparenz über Bewertungsverfahren und Geschäftspraktiken vor und fordern stärkere Regulierung. Einen umfangreichen Bericht legte dazu die FTC vor: Federal Trade Commission (2014). Data Broker. A Call for Transparency and Accountability. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

61 Altman, M; Wood A., O'Brien, D., Vadhan S. , and Gasser, U. (2016). Towards a Modern Approach to Privacy-Aware Government Data Releases. In: Berkeley Journal of Technology Law. 30/2016, S. 1967-2072, S. 2033f. http://openscholar.mit.edu/sites/default/files/dept/files/modernopendataprivacy.pdf

werden können.[62] Zwar nahm TfL daraufhin die Datensätze vom Portal und stellte sie erst wieder online, nachdem eine stärkere Anonymisierung vorgenommen wurde. Allerdings besteht die Gefahr weiterhin, denn während die unzureichend anonymisierten Daten online gestellt waren, konnten sie heruntergeladen und dann weiterhin zu solchen Zwecken genutzt werden.

Dies verdeutlicht, dass der Kompetenzaufbau solider Anonymisierungsverfahren ein wesentliches Element bei der Veröffentlichung offener Daten sein muss. Eine entsprechende Handreichung, wie sie etwa vom Information Commissioner's Office (ICO) in England veröffentlicht wurde, bietet dafür eine gute Orientierung.[63] **Die Förderung und Nutzung entsprechender technischer Lösungen,** wie etwa der in München entwickelten Software Arx – Data Anonymization Tool (http://arx.deidentifier.org/), können dafür ebenfalls wertvoll sein.

Verknüpfung von Datensätzen unterschiedlicher Quellen

Das Beispiel der Fahrraddaten aus London verdeutlicht aber noch ein anderes Problem: Das von den dort veröffentlichten Daten ausgehende Risiko steigt exponentiell, wenn diese mit anderen (öffentlich verfügbaren) Datensätzen, etwa Standortdaten des Social-Media-Dienstes Instagram, verknüpft werden. Auf diese Weise werden Rückschlüsse bis hin zum Namen einzelner Personen möglich. Mit den New Yorker Taxidaten war genau dieses bei einer vorigen Veröffentlichung in 2013 passiert: Ein Forscher konnte anhand dieser Daten nicht nur die Route von Prominenten in der Kombination mit Paparazzi-Fotos rekonstruieren, sondern auch den Wohnsitz von Besuchern eines Strip-Clubs enthüllen.[64]

Das Risiko, das sich durch die Möglichkeiten der Verknüpfung mehrerer Datensätze ergibt, wird auch als Jigsaw-Effekt bezeichnet. Eines der bekanntesten Beispiele ist die Arbeit von Sweeney et al., die in Datensätzen mit Erbgut-Profilen, die auch Postleitzahlen, Geburtstag und Geschlecht enthielten, 84 bis 97 Prozent der Profile mit Namen identifizieren konnten, indem sie

62 Mirani, L. (2014). London's bike-share program unwittingly revealed its cyclists' movements for the world to see. Quartz, 16. April 2014. http://qz.com/199209/londons-bike-share-program-unwittingly-revealed-its-cyclists-movements-for-the-world-to-see/

63 https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

Vgl. auch die erwähnte Stellungnahme der Article 29 Data Protection Working Party (2014). Opinion 05/2014 on Anonymisation Techniques. Diese liefert zwar inhaltlich wichtige Impulse, ist allerdings nicht sehr praxisorientiert aufbereitet.

64 Tockar, A. (2014). Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset. Neustar Research, 15. September 2014. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/; die Veröffentlichung der Taxi-Datensätze stand ohnehin unter Kritik, da die Taxifahrer aufgrund unzureichender Anonymisierung identifiziert werden konnten und die Daten auch Rückschlüsse auf Wohnort und Einkünfte zuließen.

sie mit dem öffentlich zugänglichen Wählerregister verknüpften.[65] Es ist sehr wahrscheinlich, dass auch die oben erwähnten New Yorker Datensätze über Strafzettel in Verbindung mit anderen Datensätzen ähnliche Rückschlüsse zulassen.

Die Evaluierung des Open-Data-Portals der Stadt Seattle durch ein interdisziplinäres Wissenschaftler-Team ergab beispielsweise, dass schon eine relativ eingeschränkte Analyse genügt, um reichhaltige Informationen für eine Profilbildung von Individuen, etwa für Marketingzwecke, zu erhalten. Und lediglich vier Datensätzen reichten aus, um Individuen in den anonymisierten Datensätzen aufgrund von sich überschneidenden Merkmalen wiederzuerkennen.[66]

Wie die einschlägige Forschung zeigt, sind also, entgegen der gängigen Meinung und der juristischen Einschätzung, etablierten Verfahren der Anonymisierung in Zeiten stetiger Automatisierung und Verknüpfung von Datensätzen durchaus Grenzen gesetzt.[67] **Nur weil ein Datensatz zu einem bestimmten Zeitpunkt rechtlich nicht Gegenstand des Datenschutzes ist, heißt dies eben nicht, dass er auf Dauer kein Potenzial für Missbrauch enthält.** Und wie die Beispiele verdeutlichen, nimmt das Datenschutzrisiko mit steigender Nutzbarkeit von Daten, etwa durch eine höhere Granularität, zu.[68] Damit ist auch die Abgrenzung einzelner Kategorien (Rohdaten, Anonymisierte Daten etc.) in der Praxis unzureichend. Sie können höchstens unterschiedliche Abstufungen von Datentypen aufzeigen, bieten aber noch kein Raster für bestimmte Ansätze zum Schutz der Privatsphäre.[69]

Solche Risiken müssen bewertet und bei der Öffnung anonymisierter Daten aus dem Fundus der öffentlichen Hand unbedingt bedacht werden. Bereits in einer Open-Data-Machbarkeitsstudie im Auftrag des BMI aus dem Jahr 2012 empfehlen die Autoren aufgrund dieser Problematik, „organisatorische

65 Sweeney, L.; Abu, A.; Winn, J. (2013). Identifying Participants in the Personal Genome Project by Name. Harvard University. Data Privacy Lab. White Paper 1021-1. http://dataprivacylab.org/projects/pgp/1021-1.pdf

66 Whittington J; Calo, R.; Simon, Mike; Woo, J.; Young, M.; Schmiedeskamp, P. (2015). Push, Pull and Spill: A Transdisciplinary Case Study in Municipal Open Government Seattle. 30/2015, S. 44. http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf

67 Einige Experten fordern gar, den Begriff Anonymisierung nicht mehr zu nutzen; vgl. zum Beispiel O'Hara, K. (2011). Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office. London, GB, Cabinet Office, S. 13f. http://eprints.soton.ac.uk/272769/3/272769OHARA11.pdf; Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review, Vol. 57, S. 1701, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. http://ssrn.com/abstract=1450006

68 Tran, E.; Scholtes; G. (2015). Open Data Literature Review. https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview_2015-04-14_1.1.pdf

69 Zuiderveen Borgesius, F.; van Eechoud, M; Gray, J. (2015). Open Data, Privacy and Fair Information Principles: Towards a Balancing Framework. Berkeley Technology Law Journal, Forthcoming; Institute for Information Law Research Paper No. 2015-04; Amsterdam Law School Research Paper No. 2015-46, S. 38f. http://ssrn.com/abstract=2695005

Prozesse für die Reanonymisierung zu etablieren"[70]. Damit einhergeht insofern auch, dass mit Datenveröffentlichung der Prozess nicht beendet ist. **Es sollten Mechanismen eingeführt werden, die eine regelmäßige Prüfung der veröffentlichten Datensätze nach möglichen Datenschutzrisiken vorsehen. Über technische Ansätze hinaus sollten insofern auch entsprechende gesetzliche Regelungen für die Veröffentlichung und den Umgang mit den Daten zu schaffen.** Interessant ist hier etwa das Beispiel Australien, wo kürzlich der Versuch einer Re-Idenfitizierung aus anonymisierten Daten als Straftat erklärt wurde.[71]

Aktuelle Maßnahmen gegen De-Anonymisierung in Australien

Die australische Regierung entfernte kürzlich einen Gesundheitsdatensatz von ihrer Open-Data-Plattform. Ein Forschungsteam der Universität Melbourne hatte darauf hingewiesen, dass die Daten über pharmazeutische und medizinische Vorsorgeleistungen die Kennzeichennummer von Versicherern und Ärzten enthielten. Entschlüsselt könnten diese darüber Auskunft geben, was einzelne Ärzte verschrieben haben. Ärzte fürchten sogar, dass in dünnbesiedelten Gebieten Rückschlüsse über einzelne Patienten gezogen werden könnten. Inzwischen wurden die Daten von der Plattform entfernt. Die Datenschutzbehörde ermittelt nun, ob das zuständige Gesundheitsministerium die Daten fahrlässig veröffentlichte, ob die vorgegebenen Anonymisierungsschritte befolgt wurden, und welche Missbrauchspotenziale sich aus den Daten tatsächlich ergeben. Kurz nachdem das Forschungsteam auf die Lücke hinwies, gab die australische Bundesanwältin bekannt, Anpassungen am nationalen Datenschutzgesetz vornehmen zu wollen. Danach wird die De-Anonymisierung anonymisierter Datensätze fortan als Straftat eingestuft werden. Dies beinhaltet auch die Beratung, Förderung oder Ermöglichung von De-Anonymisierungsverfahren sowie die Veröffentlichung oder Verbreitung von de-anonymisierten Datensätzen.

http://www.smh.com.au/national/public-service/privacy-watchdog-called-after-health-department-data-breach-20160929-grr2m1.html

http://www.zdnet.com/article/brandis-to-criminalise-re-identifying-anonymous-data-under-privacy-act/

Hier wäre es außerdem wünschenswert, dass verstärkt in Forschung und Innovation im Bereich Tragfähiger Anonymisierungsverfahren investiert wird. Aktuell ist leider festzustellen, dass bahnbrechende Forschung, etwa

70 BMI (Hrsg.) (2012). Open Government Data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Inneren, S. 439. https://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile

71 Reichert, C. (2016). Brandis to criminalise re-identifying anonymous data under Privacy Act. ZDNet, 29. September 2016. http://www.zdnet.com/article/brandis-to-criminalise-re-identifying-anonymous-data-under-privacy-act/

im Feld der sogenannten Differential Privacy[72], vorrangig in den USA erfolgt. Auch um dem in der EU-DSGVO geforderten Prinzip des privacy by design gerecht zu werden, sollte unbedingt ausreichend in die Entwicklung technischer Datenschutzmaßnahmen investiert werden.

Aber auch Transparenz und proaktive Aufklärung über den Datenschutz, die angewendeten Anonymisierungsverfahren, aber auch über die Limitierung von Anonymisierung als Gewährleistung des Schutzes der Privatsphäre sind erforderlich. **Aktuell gibt keines der deutschen Open-Data- oder Metadaten-Portale Auskunft über offene Daten und Datenschutz im Allgemeinen oder darüber, wie die dort veröffentlichten Datensätze aufbereitet werden.**[73] Ein positives Beispiel ist hier die erwähnte Plattform der Kriminalitätsdaten aus Großbritannien. Statt den Datenschutz unter den Teppich zu kehren, gehen sie proaktiv auf der Plattform mit der Herausforderung um: „Trying to find a balance between providing granular crime data and protecting the privacy of victims has been one of the biggest challenges involved in releasing this data". Sie stellen dann detailliert dar, wie die Daten aufbereitet wurden.[74]

Über den Datenschutz hinaus

Zuletzt ist festzuhalten, dass sich vermehrt Fragen stellen, die weit über den klassischen Datenschutz hinausgehen. Beispielsweise wenn Daten als Grundlage für Versicherungs- oder Kreditdienstleistungen genutzt werden und zur Diskriminierung von (oftmals bereits benachteiligten) Bevölkerungsgruppen führen. Auch offene Regierungsdaten können grundsätzlich als Quelle für solche Verfahren herangezogen werden. Die erwähnte Studie aus Seattle, wo deutlich schwächere rechtliche Rahmenbedingungen als in Deutschland gelten, zeigte, dass offene Regierungsdaten von Datenhändlern genutzt wurden, um in Kombination mit anderen Daten Profile von Bürgern zu erstellen und diese beispielsweise an Werbetreibende weiterzuverkaufen.[75]

Solche Entwicklungen stehen im Zusammenhang mit grundsätzlichen Fragen einer verantwortungsvollen Nutzung von Daten. Ein entsprechender Dis-

72 Definition aus Wikipedia: "Differential Privacy (engl. für ‚differentielle Privatheit') ist ein Maß für das Risiko einer einzelnen Person, an einer statistischen Datenbank teilzunehmen. Der Begriff fällt in den Bereich des sicheren, Privatsphären erhaltenden Veröffentlichens von empfindlichen Informationen. Mechanismen, die Differential Privacy erfüllen, verhindern, dass Angreifer unterscheiden können, ob eine bestimmte Person in einer Datenbank enthalten ist oder nicht." https://de.wikipedia.org/wiki/Differential_Privacy

73 Auf govdata.de findet sich lediglich der Randverweis: „Datenbereitsteller sollten darauf achten, dass ihre Daten auch zum Beispiel durch eine Verknüpfung mit anderen Daten nicht de-anonymisiert werden können." https://www.govdata.de/faq

74 https://data.police.uk/about/#anonymisation

75 Whittington J; Calo, R.; Simon, Mike; Woo, J.; Young, M.; Schmiedeskamp, P. (2015). Push, Pull and Spill: A Transdisciplinary Case Study in Municipal Open Government Seattle. http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf

kurs findet inzwischen statt.[76] Er umfasst etwa eine Rechenschaftspflicht für Algorithmen, die Überprüfbarkeit datenbasierter Entscheidungen und des sogenannten Profilings[77], aber auch die Regulierung des Datenhandels. **Es ist nun, da in Deutschland an einer gesetzlichen Grundlage für offene Daten gearbeitet wird, äußerst wichtig anzuerkennen, dass auch offene Daten von solchen Fragen betroffen sind** und dass die diesbezügliche Verantwortung der Verwaltung mit der zunehmenden Öffnung von Regierungsdaten steigt.

Ein möglicher Lösungsansatz, um den Missbrauch von Regierungsdaten einzugrenzen, wäre eine Restriktion der zugelassenen Anwendungen. Anwender, die solche Daten beispielsweise für sogenanntes Profiling nutzen, könnten dann entsprechend sanktioniert werden. Auch restriktive Zugangsmöglichkeiten im Sinne des Data Sharing (siehe hierzu Grafik 1 auf Seite 10) wären ein möglicher Ansatz, um die Missbrauchspotenziale bestimmter sensibler Datensätze in Zukunft zu verringern.[78] Da dieses Vorgehen gegen die allgemeingültige Open-Definition stünde, muss bei der Ausdifferenzierung dieser Nutzungseinschränkung die Zivilgesellschaft unbedingt eingebunden werden, damit nicht umgekehrt neue Hintertüren für Intransparenz geschaffen werden. Eine Herausforderung stellen solche „Post-Veröffentlichung"-Verfahren auch deswegen dar, weil sie zusätzliche Kapazitäten bei ohnehin dürftig ausgestatteten Datenschutzbehörden voraussetzen.[79]

Wer solche Aspekte von Anfang offen artikuliert und diskutiert, hat gute Chancen, einen nachhaltigen Open-Data-Ansatz zu etablieren. Staatliche Stellen, die Daten öffnen, profitieren in jeder Hinsicht von der regelmäßigen Einbindung von und dem Austausch mit externe Experten, bzw. der "Community". Nicht nur können sie sich über mögliche negative Nutzungsszenarien und deren Implikationen verständigen, sondern gleichzeitig auch ein besseres Verständnis für Datennutzer und ihre Bedarfe, Fragen und Bedenken entwickeln. Dies fördert die Entwicklung eines guten Open-Data-Ökosystems.

---

76 Vgl. etwa Executive Office of the President (2016). Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf; Bericht der Federal Trade Commission (2016). A Tool for Inclusion or Exclusion? Understanding the Issues. https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf oder in Deutschland der Ministerdialog des BMI am 13.9.2016 zu „Algorithmen und Werte" sowie die neugegründete Initiative Algorithm Watch http://algorithmwatch.org/

77 Als Profiling bezeichnet man das Erstellen von Profilen von Nutzern auf Basis von meist passiv generierten Datenpunkten. Im Rahmen der EU-DSGVO, Art. 22.1 gilt zukünftig „Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt." Ausnahmen können unter gelten, wenn der Betroffene eine ausdrückliche Einwilligung gegeben hat.

78 Vgl. in diesem Zusammenhang auch als Reaktion auf den Data.care-Misserfolg den Prozess, den die britische Regierung in Gang gebracht hat, um gemeinsam mit der Zivilgesellschaft zu erarbeiten, welche Daten mit externen Organisationen geteilt werden sollten. www.datasharing.org.uk

79 Entschließung der Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder vom 25. Mai 2016: „EU-Datenschutz-Grundverordnung erfordert zusätzliche Ressourcen für Datenschutzbehörden". http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DS-BundLaender/Entschliessung_Ressourcen.pdf?__blob=publicationFile&v=2

Davon abgesehen würde der Ansatz der Offenheit auf diese Weise auch auf der Prozessebene gelebt.

Ausblick und Impulse

Deutschland ist im Begriff, eine neue Dateninfrastruktur aufzubauen. Deren wesentliches Element werden offene Verwaltungsdaten sein. Zukünftig werden immer mehr und immer granularere Daten veröffentlicht werden. Damit dieses Projekt nachhaltig ist, darf das öffentliche Vertrauen nicht durch einen fehlenden Schutz der Privatsphäre verspielt werden. Der hohe Datenschutzstandard ist ein wichtiger deutscher Wettbewerbsvorteil, der schon deswegen ausreichend Berücksichtigung finden sollte. Wie dargelegt, häufen sich in der internationalen Praxis Beispiele, bei denen auch mittels offener Daten Rückschlüsse auf schützenswerte persönliche Informationen gezogen werden können.

Mit einem datenschutzkonformen Open-Data-Modell kann Deutschland international einen wichtigen Beitrag leisten und hier eine Lücke schließen. Viele der besonders aktiven Open-Data-Akteure, etwa im globalen Süden, haben nämlich bislang überhaupt keine rechtliche (geschweige denn technische) Verankerung des Datenschutzes und könnten von der deutschen Expertise profitieren. **Die geplante Open-Government-Partnership-Mitgliedschaft Deutschlands kann die Bundesregierung beispielsweise nutzen, um das Modell zu exportieren und im internationalen Dialog weiterzuentwickeln**.

Auf nationaler Ebene würden die Mitarbeiter des öffentlichen Dienstes davon profitieren, indem sie bei der Datenaufbereitung und -veröffentlichung entlastet und abgesichert würden. Doch auch der Länder- und Kommunalebene, die für eine breite Öffnung von Verwaltungsdaten unerlässlich sind, kann dieses Modell zur Orientierung dienen und wertvolle Unterstützung sein.

Auch aus pragmatischen Gründen sollte das Open-Data-Gesetz mit den Vorgaben der EU-DSGVO bereits konsistent sein und die Vorgaben der zu novellierenden deutschen Datenschutzgesetzgebung beachten und einhalten. Darin formulierte technische Prinzipien wie privacy by design sollten direkt auf die angestrebte Praxis der standardisierten Öffnung von Verwaltungsdaten übertragen werden. Dadurch würde ein sowohl nutzbringendes als auch datenschutzkonformes Open-by-design-Konzept erarbeitet. Ebenso wichtig ist aber, dass die verantwortungsvolle Öffnung von Regierungsdaten und die Entwicklung von entsprechenden Standards über den Rechtsrahmen hinaus diskutiert werden. Denn die im Rahmen der Rechtsprechung vorgenommene Kategorisierung von Daten bietet, wie oben aufgezeigt, nur eine sehr begrenzte Garantie dafür, dass die Privatsphäre der Bürgerinnen und Bürger gewährleistet bleibt.

Ein sinnvolles Modell profitiert insofern von **technischen Elementen (zum Beispiel Differential Privacy), einer guten Steuerung (restriktive Zugangs- oder Weiterverarbeitungsrechte, Beratungsgremien), einer proaktiven Kommunikation und dem Aufbau von Kapazitäten (Transparenz auf Open-Data-Plattformen, Weiterbildung) und der Adaption gesetzlicher Elemente (EU-DSGVO)**. Außerdem sollte das Modell die verschiedenen Phasen offener Daten (Sammlung, Verarbeitung, Veröffentlichung, Nutzung) adressieren.[80]

Dabei wird deutlich, dass die Regierung die Entscheidung, welche Daten unter Verschluss bleiben sollten, nicht im Alleingang treffen sollte. Entsprechende Steuerungsmechanismen, die wissenschaftliche und zivilgesellschaftliche Experten von außen in diese Entscheidungen einbeziehen, sind essentiell, um sowohl dem Recht auf Informationszugang und dem Recht auf Privatsphäre gerecht zu werden. Da der regelmäßige Austausch mit der „Community" für eine breite Wirkung offener Daten ohnehin unerlässlich ist, können diese Prozesse auch als Chance für den Aufbau einer nachhaltigen Open-Data-Infrastruktur verstanden werden.

Empfehlungen für nächste Schritte:

1. In Zusammenarbeit mit Datenschutzbehörden und der Open-Data- und der Datenschutz-Community **Erarbeitung klarer Standards, Richtlinien und Bewertungsmechanismen für Ministerien und Behörden entsprechend eines Open-by-design-Ansatzes** (in Anlehnung an existierende privacy impact assessment tools). In diesen sollten die Prinzipien der Europäischen Datenschutzgrundverordnung (EU-DSGVO) sowie des dementsprechend zu novellierenden deutschen Datenschutzrechts enthalten sein. An diesen sollten sich zukünftig auch Länder, Städte und Kommunen orientieren können.

2. **Capacity Building und Prozesse:** Weiterbildung und Handreichungen für Verwaltungsmitarbeiter, etwa für Anonymisierungsverfahren (ähnlich der Arbeit der ICO in Großbritannien).

3. **Investition in gute Metadatenstandards:** Kategorisierung von verfügbaren und veröffentlichten Daten (gute Datenstandards, Datenqualität und Datenpflege sind dafür wichtige Voraussetzungen).

4. **Organisatorische Prozesse zur Bearbeitung des Re-Anonymisierungsrisikos aufbauen** (Vgl. BMI 2012). Dies umfasst auch die Etablierung von Mechanismen zur kontinuierlichen Prüfung des Datenschutzrisikos bereits veröffentlichter Datensätze.

80 Vgl. hierzu als erste auszuarbeitende Ansätze: „Example categorization of privacy controls and interventions", In: Altman, M; Wood A., O'Brien, D., Vadhan S. , and Gasser, U. (2016). Towards a Modern Approach to Privacy-Aware Government Data Releases, Berkeley Journal of Technology Law. S. 2031. http://openscholar.mit.edu/sites/default/files/dept/files/modernopendataprivacy.pdf; und Wood, A.; O'Brien, D.R.; Gasser, U. (2016). Privacy and Open Data Research Briefing. Berkman Klein Publication Series, S. 8f. https://cyber.harvard.edu/publications/2016/OpenDataBriefing

5. Mehr Investitionen in **Forschung und Entwicklung technischer Ansätze für die verantwortungsvolle Veröffentlichung von Daten.** Entsprechend der EU-DSGVO also auch Privacy-by-design-Ansätze in die Entwicklung von Open-Data-Portalen integrieren und existierende Technologien prüfen.

6. **Kommunikation und Aufklärung** (zum Beispiel über Open-Data-Portale) über Datenschutz, angewandte Anonymisierungsverfahren und die Grenzen von Anonymisierungsverfahren.

7. Einbettung des Open-Data-Bereichs in den **allgemeinen Austausch über die Begrenzung oder Sanktionierung von Datennutzungsmodellen**, die über den klassischen Datenschutz hinausgehen, etwa durch Datenhändler. Hier auch proaktive Kommunikation über (nicht offenes) Data Sharing zwischen Regierung und Privatwirtschaft.

8. Austausch über und **Advocacy für Datenschutz-Prinzipien bei Open Data auf der internationalen Agenda**, wie etwa der Open Government Partnership, der International Open Data Charter, aber auch im Rahmen der internationalen Zusammenarbeit (unter Berücksichtigung der kulturellen Besonderheiten sowie politischen und rechtlichen Rahmenbedingungen).

Glossar

Open Data: Als offene Daten werden Datensätze bezeichnet, die maschinenlesbar, unter freien Nutzungslizenzen und in der Regel kostenlos über das Internet (entweder zum Download oder über Schnittstellen) verfügbar sind. In diesem Papier beziehen wir uns auf offene Verwaltungs- und Regierungsdaten wie etwa Umwelt- und Wetterdaten, Geodaten, Verkehrsdaten, Haushaltsdaten, Statistiken, Publikationen, Protokolle, Gesetze, Urteile.

Datenspektrum / Dateninfrastruktur: Unter dem Begriff Datenspektrum (siehe Grafik S. 7) fasst das britische Open Data Institute (ODI) alle Daten unterschiedlicher Art und Quelle, die grundsätzlich zur gesellschaftlich produktiven Verwendung dienen können, zusammen. Dies umfasst Daten, die – etwa im Netz – öffentlich zugänglich sind, die privat sind oder die bei der öffentlichen Hand liegen. Diese können strukturiert, semi-strukturiert oder unstrukturiert sein. Sie können geschlossen (closed), teilweise geöffnet, bzw. mit Zugang für einen beschränkten Nutzerkreis versehen (shared) oder offen (open) sein. Alternativ verwenden wir hier auch den Begriff Dateninfrastruktur und meinen damit eben nicht Systeme zur Datenspeicherung und -verarbeitung, sondern die Gesamtheit der zur Weiterverwendung verfügbaren Daten, unabhängig von ihrer Art und Herkunft.

Evidence-based decision making: Entscheidungen über Programme, Strategien oder praktische Interventionen werden nach den jeweils besten verfügbaren (empirischen) Fakten getroffen.

Informationsfreiheit: Informationsfreiheit hat zum Ziel, das Vertrauen zwischen Staat und Bürgerinnen und Bürgern zu stärken, indem öffentliches Verwaltungshandeln für Bürger transparenter und nachvollziehbar gemacht wird. Das 2006 verabschiedete Informationsfreiheitsgesetz (IFG) gewährt jedem das Recht auf freien Zugang zu amtlichen Informationen der öffentlichen Stellen des Bundes und die Einsicht in deren Verwaltungsvorgänge.

Open Government Partnership (OGP): Eine multilaterale Initiative, die sich zum Ziel gesetzt hat, in ihren 70 Mitgliedsländern die Prinzipien offenen Regierens systematisch auszubauen. Dies umfasst die Förderung von Transparenz, die Bekämpfung von Korruption, die Stärkung der Rolle der Bürger und die Nutzung der Möglichkeiten neuer Technologien, um Regierungs- und Verwaltungshandeln effektiver und nachvollziehbarer zu gestalten.

International Open Data Charter: Sie ist eine internationale Initiative, die, aufbauend auf der G8-Open-Data-Charter, sechs Prinzipien für die Veröffentlichung von Regierungsdaten formuliert. Diese sind: 1. Open by default (standardmäßig offen), 2. Rechtzeitige und umfassende Bereitstellung, 3. Zugänglichkeit und Nutzbarkeit, 4. Vergleichbarkeit und Interoperabilität, 5. Für ein besseres Verwaltungshandeln und bürgerschaftliches Engagement, 6. Für eine integrative Entwicklung und Innovation.

Abwägungsprinzip: Das Abwägungsprinzip ist ein juristischer Grundsatz, der im Idealfall nicht zwingend zu einer Entscheidung zwischen zwei Rechtsprinzipien führt (in diesem Fall zwischen dem Recht auf Privatsphäre oder dem Recht auf Information). Vielmehr wird danach gestrebt, beiden Rechten unter Berücksichtigung der Grundsätze Verhältnismäßigkeit, Notwendigkeit und Öffentliches Interesse bestmöglich gerecht zu werden. Bei der Öffnung von Daten, die ein mögliches Datenschutzrisiko enthalten, müsste also beispielsweise gefragt werden: Ist es möglich, das gleiche oder ein ähnliches Ergebnis zu erzielen, wenn ein bestimmter Datensatz nicht oder in anderer Form veröffentlicht wird? Das Ziel ist maximale Transparenz ohne Gefährdung der Privatsphäre.

Europäische Datenschutzgrundverordnung (EU-DSGVO): Diese Verordnung regelt zukünftig die Verarbeitung personenbezogener Daten in EU-Ländern. Sie wird im Mai 2018 in Kraft treten und damit die EU-Datenschutzrichtlinie von 1995 ersetzen. Öffnungsklauseln ermöglichen in einigen Punkten eine national spezifische Ausgestaltung durch die Mitgliedstaaten, ansonsten ist das Verordnungsziel gerade eine Vereinheitlichung der Datenschutzstandards in der Europäischen Union und gilt für Unternehmen und öffentliche Stellen.

Open by default: Nach diesem Ansatz sind alle Daten und Informationen einer Institution bzw. Verwaltungsebene grundsätzlich offen und frei zugänglich, die nicht dem Datenschutz unterliegen (zum Beispiel personenbezogene Daten) und bei denen keine sonstigen Hinderungsgründe vorliegen. Das bisher gebräuchliche Vorgehen wird also umgekehrt, indem Nicht-Veröffentlichungen begründet werden müssen.

Privacy by design: Dies bedeutet so viel wie Datenschutz durch Technik. Es soll sicherstellen, dass der Schutz der Privatsphäre schon bei der Entwicklung von Technik von vornherein mitgedacht und implementiert wird.

Open by design: Dieses Prinzip bezieht sich hier auf die Data-Governance-Ebene und wird komplementär zu open by default verwendet: Nicht alle Daten sind grundsätzlich offen, sondern nur diejenigen, die im Governance-Modell als offen gekennzeichnet sind. Daten durchlaufen zunächst ein Prüfungsverfahren auf mögliche Datenschutzrisiken (ähnlich einem Privacy Impact Assessment). Zusätzlich kann dieses Prinzip durch einen datenschutzsensiblen technischen Aufbau von Open-Data-Plattformen (im Sinne von privacy by design) ergänzt werden.

Personenbezogene Daten: Laut § 3 Abs. 1 BDSG handelt es sich bei personenbezogenen Daten um Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person.

Aggregierung: Sie ist ein Anonymisierungsverfahren, bei dem Personen nach

einer Generalisierung von Merkmalswerten zusammengefasst werden (etwa durch den Intervallwert „20-25 Jahre“ bei der Angabe des Alters). Die Daten können dann als individuelle Daten aufgelistet oder mit den Durchschnittswerten (alle Personen dieser Altersgruppe) in konsolidierter Form dargestellt werden.

Anonymisierung: Durch das Verfahren der Anonymisierung werden Daten in einem Maße verändert, dass diese nicht mehr einer Person zuzuordnen sind. Nach deutscher (und zukünftig europäischer) Rechtsprechung fallen anonymisierte Daten nicht unter den datenschutzrechtlich abgesicherten Bereich. Galt Anonymisierung lange als sichere Lösung, um Daten für Personen sicher zu veröffentlichen, häufen sich nun die Fälle, in denen Daten wieder de-anonymisiert, bzw. Personen in Datensätzen re-identifiziert werden konnten.

Pseudonymisierung: Bei diesem Prozess wird ein Merkmal in einem Datensatz durch ein anderes ersetzt. Die natürliche Person kann daher nach wie vor relativ leicht indirekt identifiziert werden. Deswegen fallen pseudonymisierte Daten weiterhin in den Anwendungsbereich der Rechtsvorschriften für den Datenschutz.

Profiling: Als Profiling bezeichnet man das Erstellen von Profilen von Nutzern auf Basis von meist passiv generierten Datenpunkten, die etwa über das Surfverhalten von Nutzern beim Besuch von Webseiten generiert werden. Solche Nutzerprofile werden beispielsweise von Werbetreibenden sehr häufig für die gezielte Ansprache potenzieller Kunden genutzt.

Datenhändler (engl. data broker) sammeln Informationen aus diversen Quellen über Individuen und verkaufen diese in gebündelter Form an andere Unternehmen. Beispielsweise stellen sie diese Informationen etwa Werbetreibenden für zielgerichtetes Marketing zur Verfügung, indem sie Nutzerprofile erstellen. Datenhändler nutzen solche Daten auch dazu, die Identität einer Person zu überprüfen, das potenzielle Betrugsrisiko oder die Kreditwürdigkeit einer Person zu bewerten. Verbraucherschützer und andere Kritiker werfen der Branche Intransparenz über Bewertungsverfahren und Geschäftspraktiken vor und fordern stärkere Regulierung.

Über die Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Neue Technologien verändern Gesellschaft. Dafür brauchen wir rechtzeitig politische Antworten. Die Stiftung Neue Verantwortung ist eine unabhängige Denkfabrik, in der konkrete Ideen für die aktuellen Herausforderungen des technologischen Wandels entstehen. Um Politik mit Vorschlägen zu unterstützen, führen unsere Expertinnen und Experten Wissen aus Wirtschaft, Wissenschaft, Verwaltung und Zivilgesellschaft zusammen und prüfen Ideen radikal.

Über das Projekt

Offene Verwaltungsdaten fördern die Entstehung neuer Geschäftsideen sowie neue Formen des zivilgesellschaftlichen und bürgerlichen Engagements. Die In-Wert-Setzung von Verwaltungsdaten stattet Behörden mit mehr Wissen und Kompetenzen aus und macht sie für's Digitalzeitalter fit. Das Projekt “Open Data & Privacy” treibt das Thema auf der politischen Agenda Deutschlands voran und bezieht den Datenschutz von Anfang an als Kernbestandteil mit ein. Im Austausch mit Stakeholdern aus Politik, Verwaltung, Zivilgesellschaft, Wirtschaft und Forschung entwickelt das Projekt konkrete Empfehlungen für einen Open-Data-Ansatz, der national tragfähig ist und die internationale Entwicklung auf diesem Feld voran bringt.

So erreichen Sie die Autorin:

Julia Manske
jmanske@stiftung-nv.de,
+49 30 8145 0378 92
Twitter: @juka_ma

Impressum

Stiftung Neue Verantwortung e. V.
Beisheim Center
Berliner Freiheit 2
10785 Berlin

T: +49 (0) 30 81 45 03 78 80
F: +49 (0) 30 81 45 03 78 97

www.stiftung-nv.de
info@stiftung-nv.de

Lektorat:
Antje Ziemer

Design:
Make Studio
www.make-studio.net

Layout:
Franziska Wiese

Kostenloser Download:
www.stiftung-nv.de