März 2017 · Julia Manske und Dr. Tobias Knobloch

Toolbox

Ansätze und Instrumente für die verantwortungsvolle Öffnung von Verwaltungsdaten

Stiftung
Neue
Verantwortung

Toolbox: Ansätze und Instrumente zum Schutz der Privatsphäre bei der Öffnung von Verwaltungsdaten

1. Definition von open by default

Open by default ist die standardmäßige Bereitstellung derjenigen Daten, die die Verwaltung bei der Erfüllung ihrer öffentlich-rechtlichen Aufgaben erhebt, und zwar standardisiert, in maschinenlesbarer Form und zur kostenlosen, uneingeschränkten Weiterverwendung (siehe Open-Definition).

2. Ausnahmen von open by default

A. Ausnahmen, die sich aus der geltenden Rechtslage ergeben können

Folgende gesetzliche Einschränkungen geben den Rahmen für potenzielle Ausnahmen der standardmäßigen Bereitstellung vor:

- **Datenschutzgesetz(e)** und (demnächst) das **Open-Data-Gesetz** des Bundes („Erstes Gesetz zur Änderung des E-Government-Gesetzes“)
- Einschränkungen der Weitergabe und Weiterverwendung von Informationen, wie sie im **Informationsfreiheitsgesetz (IFG)** und **Informationsweiterverwendungsgesetz (IWG)** geregelt sind
- Einschränkungen der Veröffentlichung und Nutzung von Informationen in bereits bestehenden **Spezialgesetzen**: Umweltinformationsgesetz, Geodatenzugangsgesetz, Verbraucherinformationsgesetz
- **Geheimschutz**: Verschlusssachen im Sinne des Sicherheitsüberprüfungsgesetzes
- **Betriebs- und Geschäftsgeheimnisse**
- Sonstige Geheimnisschutz-relevante Bereiche: Statistikgeheimnis, Steuergeheimnis, Sozialgeheimnis, Berufsgeheimnisse (z.B. Ärzte)
- **Schutz geistigen Eigentums**: Urheberrechtsgesetz, Markengesetz, Patentrecht

Ferner muss das Veröffentlichungsinteresse grundsätzlich mit folgenden zu schützenden Rechtsgütern abgewogen werden:

- personenbezogene und personenbeziehbare Daten
- öffentliche Belange und Rechtsdurchsetzung
- Ablauf von Verwaltungsverfahren und bevorstehende behördliche Maßnahmen
- Angaben und Mitteilungen öffentlicher Stellen des Bundes oder anderer Länder
- Schutz des behördlichen Entscheidungsbildungsprozesses

B. Ausnahmen von den Ausnahmen

Die unter A genannten Ausschlussgründe geben den Rahmen vor. Sie sind kein abschließender Ausschlussgrund, sondern lassen erstens im Zuge der Abwägung mit dem Veröffentlichungsinteresse Spielräume für Ausnahmen. Dabei darf die Entscheidung über Veröffentlichung nicht der subjektiven Beurteilung von Behördenmitarbeitern überlassen werden, sondern muss auf ausformulierten Normen und Prinzipien beruhen. Zweitens können Daten generell so aufbereitet werden, dass sie dennoch öffentlich bereitgestellt werden können. So sind zwar personenbezogene Daten grundsätzlich von der Öffnung ausgeschlossen, anonymisierte Daten, also Daten, denen der ursprüngliche Personenbezug entzogen wurde, können nach juristischer Betrachtung aber bereitgestellt werden.

C. Der Kreis der zu veröffentlichenden Daten muss zum Schutz der Privatsphäre gegebenenfalls weiter eingeschränkt werden

Auch wenn dies rechtlich legitim wäre, können nach unserer Einschätzung im Sinne eines verantwortungsvollen Open-Data-Ansatzes nicht alle Daten, die beim Negativabgleich (vgl. A) übrig bleiben beziehungsweise im Zuge der Einzelfallentscheidung als gegebenenfalls veröffentlichungsfähig ermittelt werden (vgl. B), gefahrlos als Open Data bereitgestellt werden, auch dann nicht, wenn sie anonymisiert werden. Dabei können aus unserer Sicht auch nicht bedenkenlos die Kriterien des IFG & IWG auf Open Data übertragen werden, da die Gefahren für die Privatsphäre durch die Veröffentlichung von Informationen als Open Data in aller Regel sehr viel größer sind als durch eine gezielte Herausgabe an bestimmte Personen oder Institutionen und die entsprechende Weiterverwendung. Dies ist in der Hauptsache auf zwei Ursachen zurückzuführen (vgl. dazu auch unser Papier „Offene Daten und der Schutz der Privatsphäre“ vom Oktober 2016 sowie die diese Instrumentensammlung begleitenden Ausführungen):

- Technische Grenzen von Anonymisierungsverfahren
- Neue Risiken, die sich aus der maschinellen Verarbeitung und Kombinierbarkeit mit anderen Datensätzen ergeben

Aus diesem Grund bedarf es entsprechender Instrumente, mit denen mögliche Datenschutzrisiken durch offene Daten im Vorhinein abgeschätzt und im Zuge der Veröffentlichung minimiert werden können. Die Vorschläge und Ideen in dieser Übersicht dienen diesem Zweck.

3. Vorschlag eines Drei-Phasen-Modells zur Risikoabschätzung für Open Data

Nachfolgend unterscheiden wir diese drei Prozessphasen der Datenöffnung und schlagen für die jeweilige Phase Datenschutzmaßnahmen vor:

1. **Vor der Veröffentlichung:** Entscheidung, ob Daten überhaupt geöffnet werden dürfen beziehungsweise sollten
2. **Bei der Veröffentlichung:** Datenschutzmaßnahmen im Zuge der Veröffentlichung von Daten
3. **Nach der Veröffentlichung:** Steuerung der Nutzung bereits geöffneter Daten

Sämtliche Maßnahmen bauen natürlich auf laufenden Geschäftsprozessen auf, in denen Datenschutzprinzipien – wie Datenvermeidung und Datensparsamkeit, etwa bei der Beschaffung von IT-System – ohnehin verankert sind.

Die unterschiedlichen Maßnahmearten werden nachfolgend wie folgt unterschieden und benannt:

- **Bewertungshilfe**: Hilfen zur Bewertung des Datenschutzrisikos eines oder mehrerer Datensätze
- **Institutionalisierung**: Schaffung oder Nutzung von Institutionen oder institutionalisierten Prozessen
- **Kapazitätsaufbau**: Maßnahmen zur Sensibilisierung und Fortbildung der beteiligten Mitarbeiter und zur Entwicklung von Verfahren / Methoden
- **Technischer Datenschutz**: Bereitstellung technischer Infrastruktur für erhöhten Datenschutz
- **Regulierungsansatz**: stark reglementierende Maßnahmen der Verwaltung / des Gesetzgebers

Zudem sind diese in **kurzfristig (KF), mittelfristig (MF)** und **langfristig (LF)** umsetzbare Maßnahmen eingeteilt (siehe Abkürzung in der linken Spalte).

Wir teilen außerdem Daten in Anlehnung an ein **Ampelsystem** wie folgt ein:

Grün = Datensatz kann bedenkenlos als Rohdaten (nicht-anonymisiert) geöffnet werden
Orange = Datensätze müssen zunächst eine Prüfung durchlaufen und können ggf. unter Berücksichtigung bestimmter Schutzmaßnahmen geöffnet werden
Rot = Datensatz darf so auf keinen Fall geöffnet werden

Anmerkungen

Die im Folgenden vorgestellten Ansätze setzen auf sehr unterschiedlichen Ebenen an. Einige lassen sich sehr kurzfristig und niedrigschwellig umsetzen, während andere eher als Anregung für die Weiterentwicklung des Themas dienen. Die Vorschläge sind nicht isoliert, sondern als Elemente eines Baukastens zu betrachten. Grundsätzlich entfalten sie erst im Zusammenspiel ihre volle Wirkung.
Generell gilt außerdem: Je intensiver externe Experten (sowohl aus der Open-Data- als auch der Datenschutz-Community) eingebunden werden und Transparenz über Prozesse und Vorgänge geschaffen wird, desto größer wird auf allen Seiten die Akzeptanz für eine breitflächige Veröffentlichung von Daten der öffentlichen Hand sein – auch dann, wenn einmal etwas schief gehen sollte.

1. Vor der Veröffentlichung: Entscheidung, ob Daten überhaupt geöffnet werden dürfen bzw. sollen

Mögliche Risiken, die bei dieser Entscheidung auftreten können:

- Daten, die nicht veröffentlicht werden dürfen oder sollten, werden irrtümlich bzw. infolge unzureichender Prüfung als Open Data veröffentlicht.
- Das Abwägungsprinzip wird unzureichend angewandt und das Recht auf Privatsphäre nicht hinreichend berücksichtigt.
- Das Datenschutz-Argument wird missbraucht, um relevante Datensätze mit hohem gesellschaftlichen Mehrwert nicht zu öffnen.

Beispiele aus dem Ausland, die diese Risiken verdeutlichen:

- In Bhutan stellt man die Daten von Bewerbern auf öffentliche Stellen inkl. Kontaktdaten als offene Daten auf der Regierungsplattform bereit (Hintergrundgespräch mit NGO-Vertreter aus Bhutan).
- Die Stadt Washington, D.C. veröffentlichte Wählerdaten inkl. Name, Adresse und politische Präferenzen. http://fusion.net/story/314062/washington-dc-board-of-elections-publishes-addresses
- Für einen Hackathon stellten die Ausrichter Mobilfunkdaten (Call Detail Records) auf einer Online-Plattform zum Download zur Verfügung. https://responsibledata.io/reflection-stories/open-data-hackathon
- Die polnische Regierung veröffentlichte Daten über die Empfänger bestimmter Sozialleistungen mit Namen und Adressen (Hintergrundgespräch mit NGO-Vertreter aus Polen).

Prozessvorschlag

Datenbereitsteller werden über Materialien aus einer Art Werkzeugkasten über Datenschutzrisiken und Maßnahmen informiert. Für die Bewertung der Daten steht dem Mitarbeiter eine Checkliste als Entscheidungshilfe zur Verfügung. Diese bildet ein Vorgehensmodell für die Datenöffnung ab und liefert die Bewertungsgrundlage, ob Daten geöffnet werden sollen oder nicht. Die Daten werden so vom Datenbereitsteller (Sachbearbeiter) anhand eines Ampelsystems im Sinne einer Risikoprüfung bewertet. Das Bewertungsschema muss zwingend in regelmäßigen Abständen auf Aktualität und Richtigkeit geprüft werden. **Rote Datensätze** werden nicht veröffentlicht, **grüne** werden im Sinne der Open-Definition veröffentlicht. **Orange** Datensätze durchlaufen eine Art abgeschwächtes Privacy Impact Assessment. Bei der Klassifizierung in unbedenklich (grün), nicht zu veröffentlichen (rot) und zu prüfen (orange) ist das Vier-Augen-Prinzip zu befolgen, d. h. mindestens zwei Personen der jeweiligen Behörde müssen zum gleichen Ergebnis kommen. Ist dem nicht so, muss der Fall über eine dritte Instanz, etwa die geplante Open-Data-Beratungsstelle, eskaliert werden.

Neben dieser Beratung erarbeitet die zu schaffende Open-Data-Beratungsstelle die Materialien, die den Bearbeitern als Hilfestellung dienen sollen. Dabei kann auch direkt auf Expertise der Bundesdatenschutzbeauftragten zurückgegriffen werden.
Über Datensätze, die eine hohe soziale Relevanz haben oder die intensiv angefragt werden, entscheidet ein externes Beratungsgremium, das sich aus Vertretern der jeweiligen Behörde, der Open-Data-Community und Datenschutzexperten zusammensetzt.

		Instrument	Chance / Vorteil	Risiko / Nachteil	Beispiele
1.1	KF	**Bewertungshilfe** **Werkzeugkasten für Open-Data-Veröffentlichung** (etwa Vorgehensmodell, Merkblätter, Ansprechpartner)	• Wenn anschaulich aufbereitet, gute Hilfestellung für Verwaltungsmitarbeiter, um adäquate Datensätze zu identifizieren • Wenn offen zur Verfügung gestellt bzw. sogar kollaborativ erarbeitet, erzeugt dies Vertrauen seitens der Community	• Wenn unvollständig, werden weitergehende Herausforderungen für den Datenschutz nicht berücksichtigt • Ständige Revision dringend erforderlich	Siehe „Open Data Release Toolkit“ der Stadt San Francisco: https://drive.google.com/file/d/0B0jc1tmJAlTcR0RMV01PM2NyNDA/view Siehe „Handreichung Datenschutz“ des Rats für Sozial- und WirtschaftDaten: https://www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf Siehe "Vorgehensmodell Open Government" des KDZ Zentrum für Verwaltungsforschung und der Stadt Wien: http://kdz.eu/de/open-government-vorgehensmodell
1.2	KF	**Bewertungshilfe** **Checklisten zur Bewertung des generellen Datenschutzrisikos, das von Datensätzen ausgeht** (idealerweise als Teil des Werkzeugkastens unter 1.1)	• Baut Unsicherheiten der Verwaltungsmitarbeiter ab • Gutes Material als Anregungen aus dem Ausland verfügbar	• Bewertungskriterien und Standards entwickeln und verändern sich mit der Zeit • Die Listen müssen daher (nicht oft, aber regelmäßig) auf Aktualität und Angemessenheit überprüft werden • Checklisten tragen evtl. dazu bei, dass Beurteiler automatisch nach Schema X vorgehen	Siehe „Open Data Release Form“ der Stadt San Francisco: https://docs.google.com/document/d/12uk04YOXqP10oqFy6EcJ-wRa0IrGx1B-BaCNUITP-EA/edit
1.3	MF	**Bewertungshilfe** **Ampelsystem zur Kategorisierung von Datensätzen nach potenziellem Datenschutzrisiko** **Grün** = Datensatz kann bedenkenlos als Rohdaten (nicht-anonymisiert) geöffnet werden **Rot** = Datensatz darf so auf keinen Fall geöffnet werden **Orange** = Datensätze müssen zunächst eine Prüfung durchlaufen und können ggf. unter Berücksichtigung bestimmter Schutzmaßnahmen geöffnet werden	• Deutliche Erleichterung für Datenbereitsteller • Gesteigerte Akzeptanz, wenn Bewertungskatalog für die Kategorisierung offengelegt oder gar kollaborativ mit der Zivilgesellschaft erarbeitet wird	• Risiken können (und werden) sich ändern • Muss daher ständig aktualisiert werden • Könnte evtl. dazu beitragen, Risiken zu verharmlosen	Siehe z. B. Vorschläge aus der Forschung von Zuiderveen Borgesius et al. (2015): https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2695005, oder den Ansatz des sechsstufigen „Datatag System“ von Sweeney et al. (2015): http://techscience.org/a/2015101601 Siehe „Open Data Release Toolkit“ der Stadt San Francisco: https://drive.google.com/file/d/0B0jc1tmJAlTcR0RMV01PM2NyNDA/view Siehe PSI-Bewertungsmodell der österreichischen Regierung: https://www.ref.gv.at/fileadmin/_migrated/content_uploads/psi-klassifikation_1-0-0_20150622.pdf

1.4	KF	**<u>Bewertungshilfe</u>** (gilt bezogen auf 1.1 – 1.3) **Erarbeitung maßgeschneiderter Lösungen für bestimmte Themenfelder und/oder Behörden**	• Adressiert die Herausforderung, dass Datenschutzrisiken in einzelnen Behörden unterschiedlich hoch sind, da einige Behörden über sensiblere Daten verfügen als andere (vgl. z. B. Gesundheitsdaten)	• Die Risiken einzelner Bereiche könnten vorab in Gänze nur schwer abzuschätzen sein; von daher Einteilung/Priorisierung vermutlich nur schwer möglich	
1.5	MF	**<u>Bewertungshilfe</u>** **Nutzung eines vereinfachten Privacy Impact Assessment für orange Daten** (idealerweise als Teil des Werkzeugkasten unter 1.1)	• Wenn anschaulich aufbereitet, sinnvolles – und in anderen Bereichen etabliertes – Verfahren, um schwierige Datensätze zu identifizieren und zu bewerten • Wenn offen zur Verfügung gestellt bzw. sogar kollaborativ erarbeitet, erzeugt dies Vertrauen seitens der Community	• Ggf. abschreckend, wenn nur in „Juristen-Deutsch“ oder als Bleiwüste verfügbar • Um entsprechende Instrumente erstellen zu können, müsste zunächst eine umfangreiche Bottom-up-Analyse bestehender Fälle (idealerweise auch aus dem Ausland) durchgeführt werden • Zeitliche Verzögerung der Datenveröffentlichung • Komplexität	Siehe als Grundlage das Material des britischen Datenschutzbeauftragten zu Privacy Impact Assessments: https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf Siehe Bieker et al. (2016). A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation, S.27: http://link.springer.com/chapter/10.1007/978-3-319-44760-5_2
1.6	KF	**<u>Institutionalisierung</u>** **Prüfung zu veröffentlichender Datensätze nach dem Vier-Augen-Prinzip (als Bestandteil von 1.1 - 1.4);** bei Uneinigkeit Klärung mit unabhängiger Stelle (z. B. Beratungsstelle des BMI); bei orangen Daten mindestens mit dem Datenschutzbeauftragten der Behörde	• Vermeidung von Irrtümern • (Möglichst) Kompetenzmischung		*(Eigentlich normales und bewährtes Verfahren der Verwaltung auch in anderen Bereichen.)*

1.7	KF/ MF	**Institutionalisierung** **Zentrale Beratungsstelle Open Data für alle Behörden, die bei Unsicherheiten in den zuständigen Behörden unterstützt** (z. B. bei Uneinigkeit bzgl. oranger Datensätze) (Aktuell im Rahmes des Open-Data-Gesetzes vorgesehen.)	• Klarer Ansprechpartner für Datenbereitsteller • Hilfreich, insb. wenn technische und Datenschutzexpertise in der Stelle gegeben ist	• Bindet (noch zu schaffende) Ressourcen	
1.8	LF	**Institutionalisierung** **Externes Beratungsgremium (vgl. Ethik-Kommissionen), das mit über Daten von großer gesellschaftlicher Bedeutung, aber tendenziell hohen Datenschutzrisiken entscheidet** (z. B. Daten zu Flüchtlingen, Gesundheit, Smart Cities)	• Erhöht Akzeptanz in der Gesellschaft • Stärkt Bindung mit der Community • Gut an Open-Government-Partnership-Aktivitäten anbindbar • Hilfreich, um sich wandelnden Zeitgeist abzubilden, etwa bezügl. Datensätze, deren Veröffentlichung auf einmal als besonders wichtig erachtet werden (wie etwa in den USA die Veröffentlichung von Polizeistatistiken, um Diskriminierung aufzudecken), oder aber der Klassifizierung von bislang unproblematischen Daten in risikoreiche Daten (wie etwa mit Datensätzen zu Flüchtlingsunterkünften geschehen)	• Akzeptanz bei Verwaltungsmitarbeitern ggf. nicht ausreichend gewährleistet • Legitimation kann von außen angezweifelt werden	Siehe als Inspiration die Open-Government-Diskussionsplattform des UK Open Government Network: http://www.opengovernment.org.uk/engage, sowie den zivilgesellschaftlichen Konsultationsprozess der britischen Regierung zu „Data Sharing“: http://www.datasharing.org.uk (allerdings nur einmalig), in Anlehnung an die Empfehlung 6 („Create sector transparency panel“) von O’Hara (2011): https://eprints.soton.ac.uk/272769/1/272769OHARA11.pdf

2. Bei der Veröffentlichung: Maßnahmen im Zuge der Veröffentlichung von Daten

Mögliche Risiken, die bei der Veröffentlichung entstehen können:

- Schlechte oder unzureichende Anonymisierung
- Möglichkeit der Reidentifizierung trotz Anonymisierung

Beispiele aus dem Ausland, die diese Risiken verdeutlichen:

- Nach der Veröffentlichung anonymisierter Taxi-Daten in New York konnten Hacker das Gehalt von Taxifahrern, die Bewegungsmuster von Prominenten sowie die Wohnorte einzelner Fahrgäste identifizieren. https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset
- Das ÖPNV-Unternehmen Transport for London publizierte Daten über die Nutzung öffentlicher Fahrräder, die deanonymisiert werden konnten und die Erstellung von Bewegungsmustern einzelner Radfahrer erlaubte. http://qz.com/199209/londons-bike-share-program-unwittingly-revealed-its-cyclists-movements-for-the-world-to-see
- Auf einem australischen Open-Data-Portal wurden – anonymisierte – Daten u. a. über Rezeptverschreibungen veröffentlicht. Die Universität Melbourne zeigte, dass eine Verknüpfung dieser Daten mit anderen Datensätzen Rückschlüsse auf einzelne Ärzte zuließ.
- Die britische Regierung stellte sensible Gesundheitsdaten ihrer Bürger auf der Plattform care.data bereit, ohne diese hinreichend zu informieren oder zu Beginn um ihre Einwilligung zu bitten. Der Datenzugang war zwar nur mit Registrierung möglich und die Daten pseudonymisiert, doch die Privatsphäre der Bürger war de facto nur marginal geschützt, zumal intransparent blieb, wer alles Zugriff auf die Daten erhalten sollte. Der Widerstand in der Bevölkerung war groß und führte zur Einstellung des Projekts. https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details

Prozessvorschlag

Im vorgeschlagenen Ampelsystem als orange eingestufte Datensätze können eventuell nach einer Anonymisierung veröffentlicht werden. Damit dies qualitativ hochwertig und zügig gelingt, müssen technische Instrumente eingesetzt werden, um eine hohe Anonymisierungsqualität zu erreichen. Darüber hinaus müssen entsprechende Fortbildungen für die Mitarbeiter angeboten werden.
Eingesetzte Anonymisierungsverfahren sind vorab von Experten zu evaluieren. Außerdem werden sie bei Veröffentlichung der Daten als Metadaten dokumentiert. Zu prüfen ist, inwiefern sich Privacy-by-design-Lösungen in Datenplattformen integrieren lassen. Generell sollte die Entwicklung und Implementierung nutzerfreundlicher technischer Maßnahmen für eine qualitativ hochwertige Anonymisierung gefördert werden.

		Instrument	Chance / Vorteil	Risiko / Nachteil	Beispiele
2.1	KF	**Kapazitätsaufbau** **Leitfäden für die Aggregierung und die Anonymisierung von Daten**	• Orientierung und Arbeitserleichterung, wenn anschaulich aufbereitet • Kompetenzenaufbau für technischen Datenschutz	• Der Nutzen aus Daten sinkt normalerweise umgekehrt proportional zum Aggregierungsgrad • Kann Fehler nicht ausschließen • Gute Anonymisierung ist schwierig, daher fraglich, ob Leitfäden reichen	Siehe hier den „Code of Practice“ für Anonymisierung des britischen Datenschutzbeauftragten: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf
2.2	KF	**Kapazitätsaufbau** **Trainings für Datenbereitsteller im Hinblick auf Anonymisierung** (über Kurse, Online-Tools, Blended-Learning-Angebote)	• Steigert erforderliche Kompetenzen für technischen Datenschutz	• Bindet Ressourcen und braucht Zeit	Siehe als erste Anregung etwa die Materialien des UK Anonymisation Networks, z. B. das „Anonymisation Decision-Making Framework“: http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf, oder die „Case Studies“: http://ukanon.net/ukan-resources/case-studies
2.3	MF	**Technischer Datenschutz** **Nutzung technischer Anwendungen, die hochwertige Anonymisierung erlauben** (Stichwort Privacy Enhancing Technology)	• Höheres Datenschutzniveau • Empowerment von Datenbereitstellern	• Verlass auf die Technik lässt Fehler möglicherweise schneller übersehen, der Kapazitätsaufbau innerhalb der Stellen muss deswegen vorab stattfinden • Oft sind Anwendungen nicht besonders nutzerfreundlich	Siehe die Softwarelösungen von ARX – Data Anonymization Tool: http://arx.deidentifier.org, oder von Aircloak: https://www.aircloak.com
2.4	KF	**Institutionalisierung** **Verzeichnung der Art der Anonymisierung als Metadaten**	• Erleichtert Fehlererkennung und Fehlervermeidung • Entsprechende technische Verfahren (2.3) könnten dies automatisieren • Erleichtert die Risikoprüfung nach 3.5 bzw. 3.6	• Erhöht ggf. das Risiko gezielter Deanonymisierung	

2.5	MF	**Institutionalisierung** **Externe Prüfung von Anonymisierungsverfahren** Einbindung von Experten über Konsultationsprozess, um Anonymisierungsqualität zu prüfen; erneute Prüfung in regelmäßigen Abständen	• Stellt hohe und zeitgemäße Standards der Anonymisierung sicher • Dynamische Anpassung an technische Entwicklungen • Stärkt Akzeptanz und Vertrauen	• Zusätzlicher Schritt, der Ressourcen bindet und Zeit kostet • Sorgt evtl. für Verzögerung von weiterer Datenöffnung	Siehe den Vorschlag australischer Wissenschaftler gewählte Anonymisierungverfahren der Behörden von Experten vorab prüfen zu lassen: https://pursuit.unimelb.edu.au/articles/crime-and-privacy-in-open-data
2.6	LF	**Technischer Datenschutz** **SafeAnswer-Anwendungen für sensible Daten,** technische Anwendungen, die Anfragen an Datensätze, aber kein Zugang zu Rohdaten zulassen	• Deutlich gesteigerter Datenschutz • Ermöglicht auch die Nutzung von sensiblen Daten • Förderung und Forschung in derartige Systeme könnte langfristig Datenschutzproblematik lösen	• Noch im Forschungsstadium • Entspricht nicht der Open-Definition	Siehe die Anwendung der SafeAnswer-Technologie bei OpenPDS: http://openpds.media.mit.edu Siehe den Ansatz der Differential Privacy: https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy
2.7	LF	**Technischer Datenschutz** **Privacy by design für Fachanwendungen und Datenplattformen**, z. B. automatisierte Anonymisierung	• Höheres Datenschutzniveau • Empowerment von Datenbereitstellern	• Verlass auf die Technik lässt Fehler möglicherweise schneller übersehen • Anonymisierung begrenzt bereits Nutzungspotenziale, sinnvolle und wirksame Anonymisierung hängt stets stark von späterer Nutzung ab	Siehe z. B. den Ansatz einer Datennutzungskontrolle, wie er im Projekt IND²UCE des Fraunhofer IESE verfolgt wird: https://www.iese.fraunhofer.de/de/competencies/security/ind2uce-framework.html; vgl. dazu auch 3.2 (Datenzugangskontrolle)

3. Nach der Veröffentlichung: Steuerung der Nutzung von geöffneten Daten

Risiken, die durch einmal geöffnete Daten entstehen können:

- Durch die beschriebenen Grenzen der Anonymisierung könnten Daten mit anderen öffentlich zugänglichen Daten in Verbindung gebracht werden, wodurch ggf. eine Deanonymisierung möglich wird.
- Offene Daten werden Teil des größeren „Datenkosmos“ und können so ohne das Wissen der Betroffenen von Datenhändlern für die Profilbildung genutzt werden. Damit könnten sie, ebenso wie andere Daten, etwa von Versicherungs- oder Kreditanbietern für die Tarifbildung verwendet werden.

Beispiele aus dem Ausland, die diese Risiken verdeutlichen:

- In Minneapolis wurden die Daten von Kfz-Kennzeichenlesern nach der Veröffentlichung von Datenhändlern weiterverarbeitet. Dies führte zu massiver öffentlicher Empörung. http://openscholar.mit.edu/sites/default/files/dept/files/modernopendataprivacy.pdf
- In Seattle wurden offene Regierungsdaten von Datenhändlern genutzt, um in Kombination mit anderen Daten Profile von Bürgern zu erstellen und diese beispielsweise an Werbetreibende weiterzuverkaufen. http://btlj.org/data/articles2015/vol30/30_3/1899-1966%20Whittington.pdf

Prozessvorschlag

Über die Grenzen von Anonymisierungsverfahren und empfehlenswerte Vorgehensweisen wird auf Open-Data-Plattformen proaktiv informiert. Sensible Datensätze werden lediglich mit einem restriktiven Zugang zugänglich gemacht. Welche Datensätze als sensibel einzustufen sind, wird in Absprache mit einem externen Beratungsgremium beschlossen. Um gegen Deanonymisierung vorzugehen, wird die Verbreitung und Nutzung deanonymisierter Datensätze in Anlehnung an die statistikrechtliche Praxis sanktioniert. In jeder veröffentlichenden Stelle werden Prozesse aufgesetzt, wie mit potenziell gefährdeten Datensätzen umgegangen werden kann (Nicht-Veröffentlichung, restriktiver Zugang, verbesserte Anonymisierung etc.). Fälle werden gesammelt und ausgewertet, um Fehler zukünftig zu vermeiden und ein Frühwarnsystem zu entwickeln. In regelmäßigen Abständen werden die bereitgestellten Datensätze auf Risiken der Deanonymisierung geprüft, was durch interne Bearbeiter oder externe Experten erfolgen kann.

		Instrument	Chance / Vorteil	Risiko / Nachteil	Beispiele
3.1	KF	**Kapazitätsaufbau** **Aufklärung über Anonymisierungsverfahren und deren Grenzen auf Open-Data-Plattformen und über andere Medien,** ggf. inkl. Zertifizierung von Plattformen bzgl. Privacy-Berücksichtigung	• Proaktive Hinweise auf Herausforderungen können vertrauensbildend wirken • Kann auch als Referenz genutzt werden für den Fall, dass etwas schief geht	• Wirkt nur aufklärend, aber minimiert nicht das unmittelbare Risiko	Siehe Erläuterung der Anonymisierungsverfahren auf Webseite von UK Police Data: https://data.police.uk/about/#anonymisation
3.2	MF	**Regulierungsansatz** **Restriktiver Zugang für registrierte oder sogar zertifizierte Nutzer oder auf spezielle Anfrage,** z. B. durch Forscher	• Höherer Sicherheitsstandard durch Überprüfbarkeit von Datennutzern • Ermöglicht den Zugang zu relevanten, jedoch risikoreicheren Datensätzen	• Entspricht nicht der Open-Definition • Bürokratischer Mehraufwand	Siehe den Ansatz aus San Francisco, nach dessen Bewertungsschema einige Datensätze nur restriktiv zur Verfügung gestellt werden („Open Data Release Toolkit“, S. 19): https://drive.google.com/file/d/0B0jc1tmJAlTcR0RMV01PM2NyNDA/view Siehe z. B. den Ansatz einer Datennutzungskontrolle, wie er im Projekt IND²UCE des Fraunhofer IESE verfolgt wird, vgl. dazu auch Punkt 2.8
3.3	LF	**Regulierungsansatz** **Sanktionierung der Verbreitung und Nutzung deanonymisierter Daten**	• Kann bei hohen Strafen Weitergabe von deanonymisierten Daten einschränken	• Internationaler Datenfluss macht nationale Regulierung nur bedingt wirksam • Entspricht nicht der Open-Definition • Abhängig von wirksamer Rechtsdurchsetzung	Siehe die Empfehlung australischer Wissenschaftler (gegen die Entscheidung der australischen Regierung): https://pursuit.unimelb.edu.au/articles/crime-and-privacy-in-open-data Datenschutzgesetze schränken bereits die weitere Verwendung öffentlicher Daten insofern ein, als berechtigte Interessen evtl. betroffener Personen nicht unverhältnismäßig beeinträchtigt werden dürfen. Siehe auch Ausführungen im Statistikgesetz zur Untersagung vorsätzlicher Reidentifizierung, siehe § 21 des Bundesstatistikgesetzes zum Verbot der Reidentifizierung.
3.4	MF	**Kapazitätsaufbau** **Etablierung von Prozessen, um mit Deanonymisierung umzugehen,** z. B. Anleitung zur schnellen Entfernung von Daten (inkl. Reporting)	• Effiziente Prozesse essenziell, um größere Risiken zu verhindern • Reporting unterstützt Akzeptanz • Gutes Fehlermanagement hilft, ähnliche Fehler künftig zu vermeiden	• Wirkt lediglich als Schadensbegrenzung und nicht gegen die tatsächlichen Risiken	Siehe hier die Empfehlung aus der Studie des BMI „Open Government Data Deutschland“, „organisatorische Prozesse für die Reanonymisierung zu etablieren“: https://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile

3.5	MF	**Institutionalisierung** **Jährliche generelle Risikoprüfung der Open-Data-Aktivitäten auf Datenschutzrisiken**	• Dynamische Anpassbarkeit • Notwendige Zusammenarbeit mit Wissenschaft fördert Community-Building	• Erfordert Ressourcen • Begrenzt lediglich den Schaden	Siehe die „Open Data Policy“ von Seattle: http://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf
3.6	MF	**Institutionalisierung** **Stetig forcierte Prüfung auf mögliche Reidentifizierungsrisiken von außen** (ggf. Bescheinigung der Prüfung auf Open-Data-Plattformen ähnlich einer Zertifizierung, siehe 3.1)	• Gut, um mit der Community in Kontakt zu kommen und an Privacy-Standards zu arbeiten • Dynamische Anpassung an technische Entwicklungen möglich	• Erfordert Ressourcen • Begrenzt lediglich den Schaden • Setzt Vertrauen in die und seitens der Community voraus	In Anlehnung an etablierte Verfahren aus der IT-Sicherheit: Auf Basis definierter Codes of Ethics werden sog. „Bug-Bounty-Programme“ (Angriffe durch externe Hacker zur Identifizierung von Systemschwachstellen) durchgeführt: https://en.wikipedia.org/wiki/Bug_bounty_program Siehe etwa das Angebot von Penetrationstests des Chaos Computer Clubs Oder über Wissenschaft: In Seattle hat ein Forschungsteam Daten auf Reidentifizierbarkeit geprüft
3.7	LF	**Institutionalisierung** **Aufbau eines Fehlerkatalogs** Fälle von Privacy-Verletzung sind meldepflichtig, werden (international) gesammelt und analysiert, um dann ein Frühwarnsystem daraus zu entwickeln	• Wirksames Instrument an sich und um Datenschutzexpertise in den Behörden aufzubauen • Hilft in Verknüpfung mit 2.4 Fehler besser zu verstehen und zukünftig zu vermeiden	• Dauert und muss kontinuierlich gepflegt werden • Hat keinen unmittelbaren Effekt	Siehe Vorgabe der EU-DSGVO, die zur Meldung von Data-Breaches verpflichtet, siehe „Notification of a personal data breach to the supervisory authority“ Art. 33, EU-DSGVO.
3.8	MF	**Regulierungsansatz** **Allgemeine Beschränkung der Datennutzung in Nutzungsbedingungen / Lizenzen**	• Setzt beim allgemeinen aktuellen Datendiskurs an • Verringert z. B. das Risiko, dass Daten diskriminierend genutzt werden	• Internationaler Datenfluss macht nationale Regulierung nur bedingt wirksam • Entspricht nicht der Open-Definition	Siehe hier das Archivrecht, nach dem die Nutzung von personenbezogenen Informationsbeständen nur zur Forschung zugelassen wird, sofern diese nicht veröffentlicht werden

Stiftung Neue Verantwortung
März 2017
Toolbox

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 nationa tragfähig ist und die internationale Entwicklung auf diesem Feld voran bringt.

So erreichen Sie die Autoren:

Julia Manske
jmanske@stiftung-nv.de
Twitter: @juka_ma

Tobias Knobloch
tknobloch@stiftung-nv.de
Twitter: @tobiasknobloch

Stiftung Neue Verantwortung
März 2017
Toolbox

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
Twitter: @SNV_berlin

Design:

Make Studio
www.make-studio.net

Layout:

Franziska Wiese
Kostenloser Download:
www.stiftung-nv.de

Dieser Beitrag unterliegt einer CreativeCommons-Lizenz (CC BY-SA). Die Vervielfältigung, Verbreitung und Veröffentlichung, Veränderung oder Übersetzung von Inhalten der stiftung neue verantwortung, die mit der Lizenz „CC BY-SA“ gekennzeichnet sind, sowie die Erstellung daraus abgeleiteter Produkte sind unter den Bedingungen „Namensnennung” und „Weiterverwendung unter gleicher Lizenz“ gestattet. Ausführliche Informationen zu den Lizenzbedingungen finden Sie hier:
http://creativecommons.org/licenses/by-sa/4.0/