

Open Data und Big Data

Jan Schallaböck, Referat 6

ULD-Sommerakademie
Kiel im August 2013



www.datenschutzzentrum.de

Open Data Prinzipien

1. Vollständigkeit

Alle öffentlichen Daten werden verfügbar gemacht. Als Öffentliche Daten werden hierbei Daten verstanden, die nicht berechtigten Datenschutz-, Sicherheits- oder Zugangsbeschränkungen unterliegen.

2. Primärquelle

Die Daten werden an ihrem Ursprung gesammelt. Dies geschieht mit dem höchstmöglichen Feinheitegrad, nicht in aggregierten oder sonstwie modifizierten Formaten.

3. Zeitnah

Daten werden so zügig, wie zur Werterhaltung, notwendig zur Verfügung gestellt.

4. Zugänglich

Daten werden so vielen Nutzern wie möglich für möglichst viele Verwendungszwecke bereit gestellt.

5. Maschinenlesbar

Daten werden zur automatisierten Verarbeitung strukturiert zur Verfügung gestellt.

6. Nicht diskriminierend

Daten sind für Alle verfügbar, ohne dass eine Registrierung notwendig ist.

7. Nicht proprietär

Daten werden in standardisierten Formaten bereit gestellt, über die keine juristische Person die alleinige Kontrolle hat.

8. Lizenzfrei

Daten unterliegen keinem Urheberrecht, Patenten, Markenzeichen oder Geschäftsgeheimnissen. Sinnvolle Datenschutz-, Sicherheits- und Zugangsbeschränkungen sind zulässig.

Big Data

Ermöglicht

- "neue Erkenntnisse über gesellschaftliche, ökonomische oder politische Sachverhalte" (Weichert)
- die Analyse menschlichen Verhaltens und die Gewinnung von Erkenntnis über menschliche Verhaltensweisen

"Denkbar ist so nicht nur die Kontrolle von einzelnen Individuen, sondern auch die umfassende Überwachung von Menschengruppen oder gar ganzer Gesellschaften." (Weichert)

- Besonders exklusives Wissen über menschliche Verhaltensweisen gefährdet die informationelle Selbstbestimmung

Informationelle Selbstbestimmung : Entfaltungschancen

(als Mantra zur Erinnerung)

"Wer unsicher ist, ob abweichende Verhaltensweisen jederzeit notiert und als Informationen dauerhaft gespeichert, verwendet oder **weitergegeben** werden, **wird versuchen, nicht durch solche Verhaltensweisen aufzufallen.** [...] Dies würde nicht nur die individuellen Entfaltungschancen des Einzelnen beeinträchtigen, sondern auch das Gemeinwohl, weil Selbstbestimmung eine elementare Funktionsbedingung eines auf Handlungs- und Mitwirkungsfähigkeit seiner Bürger begründeten freiheitlichen demokratischen Gemeinwesens ist."

Rechtliche Anforderungen an die Veröffentlichung personenbezogener Daten

- Ges. Grundlage vH (etwa öff. Register)

V absolute Anonymität (16 I Nr. 4 BStatG)

V faktische Anonymität (etwa BDSG, LDSGen)

- Insbes. "on-site" faktische Anonymität (praktische Umsetzung der Anforderung des § 16 VI BStatG)

V Interessensabwägung (etwa: Meinungsfreiheit)

Anonymisierungsverfahren

- Traditionell
 - Variablenunterdrückung
 - Objekt- oder Wertunterdrückung
 - Informationsreduktion
- Datenverändernde Verfahren
 - Mikroaggregation
 - Zufallsvertauschungen
 - Zufallsüberlagerungen
 - Simulationsverfahren (zB. Resampling, nicht behandelt)
 - Imputationsverfahren (nicht behandelt)

Variablenunterdrückung

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben	Geburtsland
22	W	10119	40.000	2.200	Ethyopien
25	M	10439	44.000	2.600	Syrien
30	M	10461	64.000	3.600	Sudan
35	W	10119	63.000	4.000	China
40	W	10436	136.000	7.000	Afghanistan
50	M	10117	56.000	2.400	Ethyopien

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

7

Variablenunterdrückung

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben	Geburtsland
22	W	10119	40.000	2.200	Ethyopien
25	M	10439	44.000	2.600	Syrien
30	M	10461	64.000	3.600	Sudan
35	W	10119	63.000	4.000	China
40	W	10436	136.000	7.000	Afghanistan
50	M	10117	56.000	2.400	Ethyopien

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

8

Objekt- oder Wertunterdrückung

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	40.000	2.200
25	M	10439	44.000	2.600
30	M	10461	64.000	3.600
35	W	10119	63.000	4.000
40	W	10436	136.000	7.000
50	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

9

Objekt- oder Wertunterdrückung

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	40.000	2.200
25	M	10439	44.000	2.600
30	M	10461	64.000	3.600
35	W	10119	63.000	4.000
40	W	10436	136.000	7.000
50	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

10

Informationsreduktion

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	40.000	2.200
25	M	10439	44.000	2.600
30	M	10461	64.000	3.600
35	W	10119	63.000	4.000
40	W	10436	136.000	7.000
50	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

11

Informationsreduktion

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	niedrig	niedrig
25	M	10439	Niedrig	niedrig
30	M	10461	mittel	mittel
35	W	10119	mittel	Mittel
40	W	10436	hoch	hoch
50	M	10117	mittel	niedrig

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

12

Mikroaggregation

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	40.000	2.200
25	M	10439	44.000	2.600
30	M	10461	64.000	3.600
35	W	10119	63.000	4.000
40	W	10436	136.000	7.000
50	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

13

Mikroaggregation

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	46.666	2.200
25	M	10439	46.666	2.600
30	M	10461	87.666	3.600
35	W	10119	87.666	4.000
40	W	10436	87.666	7.000
50	M	10117	46.666	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

14

Zufallsvertauschungen

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
20-24	W	10119	40.000	2.200
25-29	M	10439	44.000	2.600
30-34	M	10461	64.000	3.600
35-39	W	10119	63.000	4.000
40-44	W	10436	136.000	7.000
50-54	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

15

Zufallsvertauschungen

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
34-39	W	10119	40.000	2.200
25-29	M	10439	44.000	2.600
30-34	M	10461	64.000	3.600
20-24	W	10119	63.000	4.000
40-44	W	10436	136.000	7.000
50-54	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

16

Zufallsüberlagerungen

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	40.000	2.200
25	M	10439	44.000	2.600
30	M	10461	64.000	3.600
35	W	10119	63.000	4.000
40	W	10436	136.000	7.000
50	M	10117	56.000	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

17

Zufallsüberlagerungen

Alter	Geschlecht	PLZ	Einkommen	Monatl. Ausgaben
22	W	10119	39.828	2.200
25	M	10439	45.862	2.600
30	M	10461	64.960	3.600
35	W	10119	60.547	4.000
40	W	10436	134.951	7.000
50	M	10117	55.692	2.400

Tabelle nach ICO UK (Hrsg.):
Anonymisation: managing data
protection risk, code of practise,

18

Fazit Anonymisierung

- Absolute Anonymität ist schwer sicher zu stellen
- Zunehmend fehlt das notwendige Wissen mögliche Angriffe zu erkennen
- Mit größerer Verbreitung von möglichen Verknüpfungstabellen steigt das Risiko von Einzel- und Fischzugangriffen weiter.
- Bei guter Sicherheit sinkt die Qualität der Daten zumeist erheblich – grade für unbestimmte Auswertungen wie dies bei der Veröffentlichung von Open Data der Fall ist

Informationelle Selbstbestimmung: Gemeinwohl

(als Mantra zur Erinnerung)

“Wer unsicher ist, ob abweichende Verhaltensweisen jederzeit notiert und als Informationen dauerhaft gespeichert, verwendet oder weitergegeben werden, wird versuchen, nicht durch solche Verhaltensweisen aufzufallen. [...] Dies würde nicht nur die individuellen Entfaltungschancen des Einzelnen beeinträchtigen, **sondern auch das Gemeinwohl, weil Selbstbestimmung eine elementare Funktionsbedingung eines auf Handlungs- und Mitwirkungsfähigkeit seiner Bürger begründeten freiheitlichen demokratischen Gemeinwesens ist.**”

Anonyme personenbezogene Daten Ein Paradoxon?

- “Das Gegenteil von personenbezogenen Daten sind anonyme Daten”
(<http://www.berliner-datenschutzrunde.de/2-einige-grundbegriffe>)
- Ansatzpunkt muss die Tatsache sein, dass Informationen über Menschen gesammelt werden. Auf den Grad der “Bestimmbarkeit” einer Person kommt es gar nicht an.

Open Data Nur ein Thema für die Verwaltung?

- Es ist fraglich, ob und wann Erkenntnisse die auf der Beobachtung von Menschen beruhen einem umfassenden Schutz als Geschäftsgeheimnisse unterfallen können.
- Gesellschaftlich ist dies jedenfalls in eine definierte Balance zu bringen, die derzeit noch nicht hergestellt ist.

Thesen

- Exklusives Wissen über menschliche Verhaltensweisen- basierend auf "big data" - ist zumindest in Teilen – einer gesellschaftlichen Regulation zugänglich zu machen
- Unabhängig vom Personenbezug sollten entsprechende Analysen nur nach behördlicher Genehmigung zulässig sein wobei die eingesetzten Verfahren transparent zu machen sind.
- Eine reine "Einwilligungslösung" sollte nicht möglich sein.
- Die Analyse menschlicher Verhaltensweisen sollte andererseits in der Regel zulässig sein, wenn die Daten öffentlich verfügbar sind (etwa in Form von open data)
- Die Veröffentlichung von anonymisierten personenbezogenen Daten ist nur bei absoluter Anonymität zulässig (oder in gesetzlichen Ausnahmefällen, etwa bei öffentlichen Registern)

Vielen Dank