

Matthias Stürmer

Governance bei der Freigabe von Open Data

Die Herausforderungen offener Daten und was sonst noch gegen Open Data spricht

Open data includes data release by authorities, companies and other organisations. In the article, the features of open data are being elucidated, as well as what must be taken into account when approving open data and how potential data bases can be valuated for logical reasons by means of a governance approach. Both relevant challenges and possible bogus arguments against open data reveal difficulties that can be met frequently when releasing data. (ah)

Category: Articles

Region: Switzerland

Field of law: E-Democracy, E-Government

Citation: Matthias Stürmer, Governance bei der Freigabe von Open Data, in: Jusletter IT 25 May 2016

Inhaltsübersicht

1. Einleitung
2. Definition von «Open Data»
3. Herausforderungen bei Open Data
 - 3.1. Strategie
 - 3.2. Datenschutz
 - 3.3. Sicherheit
 - 3.4. Datenqualität
 - 3.5. Externes Interesse
 - 3.6. Aufwand
4. Governance durch Bewertung der Datenfreigabe
5. Scheinargumente gegen Open Data
6. Schlussfolgerungen

1. Einleitung

[Rz 1] Die Freigabe von nicht personenbezogenen und nicht sicherheitskritischen Daten als Open Data macht aus vielen Gründen Sinn und lässt auf grossen Nutzen hoffen. Die Europäische Kommission spricht in ihrem Bericht «Creating Value through Open Data»¹ von einem bis zu 286 Mrd. Euro grossen Marktvolumen für Open Data in den EU- und EFTA-Ländern (d.h. inklusive der Schweiz) bis ins Jahr 2020. Gleichzeitig sollen bis dahin an die 100'000 neue Stellen geschaffen werden, die in direktem Zusammenhang mit Open Data stehen. Und ausserdem, so der Bericht, werden dank Open Data die aufaddierten Kosteneinsparungen in der öffentlichen Verwaltung bis 2020 rund 1.7 Mrd. Euro betragen.

[Rz 2] Bei all diesen euphorischen Meldungen über die positiven Auswirkungen von Open Data stellt sich die berechtigte Frage, welches die negativen Seiten bei Datenfreigaben sind und wie diese angegangen werden können. Denn kritische Aspekte bei Open Data gibt es durchaus, auch wenn deren Relevanz von Fall zu Fall unterschiedlich einzustufen ist. Die folgenden Erläuterungen definieren den Begriff «Open Data», zeigen Herausforderungen bezüglich deren Freigabe auf und beschreiben, wie diese Problemstellungen in der Praxis angegangen werden können.

2. Definition von «Open Data»

[Rz 3] Bei der Diskussion über Open Data ist es entscheidend, dass Klarheit über die Begrifflichkeiten herrscht. Nur zu oft schmücken sich Behörden und Firmen mit dem Label «Open Data» und halten sich dabei nicht im Geringsten an die Anforderungen. Zugegebenermassen sind die Open Data Kriterien sehr anspruchsvoll, wenn sie alle vollumfänglich eingehalten werden sollen. Dennoch machen alle zehn der sogenannten Open Data Prinzipien Sinn:²

1. **Vollständigkeit:** Daten werden so detailliert wie möglich inklusive Metadaten, Formeln und anderen Erläuterungen veröffentlicht.

¹ WENDY CARRARA/WAE SAN CHAN/SANDER FISCHER/EVA VAN STEENBERGEN (2015), Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources. European Commission. http://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf (alle Internetquellen zuletzt besucht am 17. Mai 2016).

² JOHN WONDERLICH (2010), Ten Principles for Opening Up Government Information, Sunlight Foundation, 11. August 2010. <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>.

2. **Primärquelle:** Daten werden soweit wie möglich in ihrer ursprünglichen Form d.h. nicht aggregiert oder anderweitig modifiziert veröffentlicht.
3. **Zeitnähe:** Daten werden möglichst rasch veröffentlicht, idealerweise in Echtzeit zugänglich über eine Programmierschnittstelle (Application Programming Interface API).
4. **Zugänglichkeit:** Daten werden ohne organisatorische oder technische Hürden so veröffentlicht, dass sie möglichst einfach und gut auffindbar zugänglich sind.
5. **Maschinenlesbarkeit:** Daten werden in strukturierten Formaten freigegeben, die ohne Einschränkungen durch Software gelesen und weiterverarbeitet werden können.
6. **Einschränkungsfreiheit:** Daten werden ohne Registrierung und ohne die Angabe von Rechtfertigungsgründen freigegeben.
7. **Verwendung offener Standards:** Daten werden in Formaten abgespeichert, die ohne proprietäre Software gelesen werden können.
8. **Freie Lizenz:** Daten werden unter einer offenen Lizenz freigegeben, sodass sie uneingeschränkt genutzt, verändert und auch kommerziell weiterverwendet werden können (bspw. Creative Commons Namensnennung CC-BY oder Public Domain CC0)
9. **Dauerhaftigkeit:** Daten werden permanent online zur Verfügung gestellt und Aktualisierungen werden versioniert veröffentlicht.
10. **Kostenlos:** Daten stehen ohne Gebühr zur Verfügung.

[Rz 4] Wesentlich für das Verständnis des Begriffs ist ausserdem, dass Open Data grundsätzlich keine personenbezogenen und sicherheitsrelevanten Daten umfasst. Damit ist gewährleistet, dass weder Datenschutz noch Sicherheit durch Open Data Freigaben gefährdet werden.

[Rz 5] Die Umsetzung der aufgeführten Anforderungen ist jedoch nicht immer so einfach, wie die nachfolgenden Erläuterungen zeigen.

3. Herausforderungen bei Open Data

[Rz 6] Wenngleich der gesamtwirtschaftliche Nutzen von Open Data wie eingangs erwähnt hoch sein mag, so stellt die konkrete Freigabe von Daten die jeweiligen Organisationen meist vor schwerwiegende individuelle Fragestellungen. Verlieren wir nicht die Kontrolle über unsere Daten, wenn wir sie freigeben? Gefährden wir die Sicherheit unserer Organisation? Oder: Verletzen wir möglicherweise das Datenschutzgesetz, wenn wir diese Daten freigeben?

[Rz 7] Nachfolgend werden sechs unterschiedliche strategische, rechtliche, technische, finanzielle und organisatorische Aspekte aufgeführt, die je nach Situation der Datenfreigabe mehr oder weniger relevant sind und beachtet werden sollten. Nicht relevant sind weitere Argumente gegen die Freigabe von Open Data, die zwar in den Diskussionen oftmals genannt werden, aber letztlich als Scheinargumente dazu dienen, aus Gründen der Bequemlichkeit oder aufgrund von Ängsten die Datenfreigabe zu verhindern. Auch diese Argumente werden der Vollständigkeit halber aufgeführt, allerdings erst als letztes Kapitel.

3.1. Strategie

[Rz 8] Daten sind sowohl im öffentlichen wie auch im privaten Sektor oftmals sehr wertvoll. Einerseits kann es einen sehr hohen Aufwand bedeuten, die Daten überhaupt zu generieren, dann zu bereinigen und schliesslich aufzubereiten. Andererseits können Daten einen grossen finan-

ziellen Nutzen stiften, wenn sie Firmen als Basis ihres Geschäftsmodells dienen wie bspw. Produktepläne oder Absatzzahlen. Oder Behörden müssen in gewissen Fällen wie bei Geodaten oder Meteorodaten aufgrund gesetzlicher Vorschriften zurzeit Gebühren verlangen.

[Rz 9] Derart wertvolle Daten, die als direkter Wettbewerbsvorteil oder auf andere Weise von strategischer Bedeutung sein können, werden von den jeweiligen Organisationen verständlicherweise gar nicht oder nur teilweise als Open Data freigegeben. Gleichzeitig kann es für Unternehmen und öffentliche Institutionen auch eine Chance sein, bislang wettbewerbsdifferenzierende bzw. gebührenpflichtige Daten freizugeben um neue Konkurrenzvorteile bzw. öffentlichen Nutzen zu schaffen. In jedem Fall hängt es von den spezifischen Kosten-Nutzen Überlegungen eines Datensatzes ab, ob dessen Open Data Freigabe die übergeordnete Organisations-Strategie unterstützt ist oder nicht.

3.2. Datenschutz

[Rz 10] Open Data sind per Definition Daten, die keinen Rückschluss auf einzelne Personen zulassen, um deren Privatsphäre nicht zu verletzen. Ausnahmen bilden in gewissen Situationen öffentliche Personen aus Politik, Verwaltung und Wirtschaft. In jedem Fall muss bei der Freigabe von Daten grosser Wert auf den Datenschutz gelegt werden, um persönliche Daten zu schützen und bestehende Gesetze oder andere Vorgaben wie Allgemeine Geschäftsbedingungen nicht zu verletzen. Insbesondere muss darauf geachtet werden, dass auch durch Verlinken mit anderen anonymen Daten keine Individuen identifiziert werden können und dadurch schützenswerte Informationen veröffentlicht würden. Trotzdem sollte der Datenschutz aber auch nicht vorschnell als Pauschalargument für die Verhinderung von Open Data eingesetzt werden, denn oftmals besteht die Möglichkeit, Personendaten zu anonymisieren oder auf andere Weise freizugeben ohne dass der Datenschutz verletzt wird.

3.3. Sicherheit

[Rz 11] Analog zum Grundsatz, dass Open Data keine personenbezogenen Daten umfassen soll, sind gemäss Definition bei Open Data auch sicherheitsrelevante Daten ausgeschlossen: Informationen über die Geheimdienste oder Militäranlagen können die Staatssicherheit gefährden, Details über kritische Infrastrukturen wie Elektrizitätswerke oder Staudämme können terroristische Anschläge erleichtern. Ähnlich wie beim Thema Datenschutz sollte aber auch das Argument «Sicherheit» nicht als Generalentschuldigung gegen die Datenfreigabe angewendet werden. Es bedarf stets einer differenzierten Analyse, ob gewisse Datenbestandteile als Open Data ein Sicherheitsrisiko darstellen könnten oder ob kriminelle Kräfte die betroffenen Daten bspw. nicht auch auf andere Weise ebenso einfach beschaffen könnten. Wenn bestimmte Datenfelder tatsächlich sicherheitskritische Informationen enthalten sollten, könnten diese bei einer Open Data Freigabe ausgelassen werden, um dennoch den restlichen Datensatz veröffentlichen zu können.

3.4. Datenqualität

[Rz 12] Geben Firmen, öffentliche Stellen oder Nichtregierungsorganisationen Daten als Open Data frei, so tun sie gut daran, eine hohe Datenqualität sicherzustellen. Aufgrund der öffentlich

zugänglichen und ausserdem maschinenlesbaren Daten können Fehler relativ einfach aufgedeckt und über soziale und traditionelle Medien rasch verbreitet werden. Die Daten sind deshalb gegenüber den bekannten Datenqualitäts-Kriterien Vollständigkeit, Genauigkeit, Aktualität, Konsistenz und Fehlerfreiheit zu prüfen. Nur wenn die Ansprüche an die Datenqualität ausreichend erfüllt sind, sollte eine Freigabe als Open Data überhaupt in Betracht gezogen werden. Falls die Datenqualität als unzureichend eingestuft werden muss, macht deren Verbesserung sowohl für die Freigabe als Open Data als auch für den eigenen Gebrauch der Daten Sinn.

3.5. Externes Interesse

[Rz 13] Open Data Portale beklagen heute oftmals, dass die aufwändig publizierten Datensätze zu wenig genutzt werden. Das ist nicht weiter erstaunlich, denn das Interesse externer Zielgruppen an bestimmten Daten ist letztlich abhängig davon, wie hoch der Nutzen oder die Erkenntnisse daraus eingeschätzt werden. Wenn bspw. Dienstleister anhand von Open Data der SBB³ künftig ihre Mobile Apps mit neuen Informationen wie bspw. Ausrüstung der Haltestellen anreichern können, dann lässt sich für diese Software-Anbieter eine Wertschöpfung der Daten realisieren, was den Nutzen und gleichzeitig die Nutzung der Daten erhöht. Oder Journalisten haben oftmals grosses Interesse an Subventionszahlungen und fragen diese über das Öffentlichkeitsgesetz gezielt ab. Es ist deshalb sinnvoll, dass die Datenbesitzer die Freigabe derjenigen Datensätze priorisieren, bei denen sie von einem grossen externen Interesse und deshalb einer hohen Nutzung ausgehen können.

3.6. Aufwand

[Rz 14] Meistens ist der Aufwand, Rohdaten erstmals zu erheben und aktuell zu halten, wesentlich höher als deren Publikation auf einem Open Data Portal. Dennoch muss bis zur Freigabe der Daten möglicherweise nochmals einige Arbeit in deren Bereinigung von Inkonsistenzen und Fehlern, Transformation in passende Datenformate und in die Dokumentation und Bekanntmachung investiert werden. Insbesondere die technischen Anforderungen an Open Data können für Organisationen eine erhebliche Herausforderung darstellen, wenn sie bisher ihre Daten bspw. bloss als PDF-Datei, also nicht in einem maschinenlesbaren Format, veröffentlicht haben. Dabei stellt sich letztlich die Frage, wie viel Aufwand in die Datenaufbereitung investiert werden soll, um einen möglichst optimalen Kosten-Nutzen Effekt zu realisieren. Ein weiterer Aspekt betrifft die Organisationskultur. Besteht Unterstützung aus der Leitungsebene und Praxiserfahrung mit Open Data, ist der Aufwand der Datenfreigabe wesentlich niedriger, als wenn gegen Widerstände aus der Führungsebene angetreten werden muss oder wenn die Organisation sich erstmalig an Open Data freigibt.

³ <https://opendata.swiss/de/organization/schweizerische-bundesbahnen-sbb>.

4. Governance durch Bewertung der Datenfreigabe

[Rz 15] Diese Erläuterungen möglicher Problemfelder bei der Freigabe von Daten erlauben einerseits die systematische Benennung der Ursachen, was einen ersten Schritt in Richtung Problemlösung darstellt. Andererseits bieten diese Herausforderungen auch die Chance, als Grundlage für eine Hilfestellung bezüglich Governance von Open Data zu dienen. So betrachtet stellen die erwähnten Herausforderungen nämlich eine Art Checkliste dar, welche Fragen bei der Freigabe von Open Data behandelt werden müssen.

[Rz 16] Basierend auf dem Open Government Vorgehensmodell der KDZ⁴ zeigt das folgende Bewertungsschema auf, welche Aspekte eine Organisation bei der Open Data Freigabe betrachten sollte. Wie in der Anleitung der KDZ zum Datenmonitoring beschrieben, aber abgeändert nach leicht angepassten und zusammengefassten Kriterien, lassen sich daraus Bewertungen von potentiell zu publizierenden Datensätzen ableiten. So stellt jede Herausforderung ein Kriterium dar, anhand dessen ein Datensatz bezüglich der Praktikabilität und Sinnhaftigkeit bewertet werden kann. Je höher die Bewertung pro Herausforderung und letztlich in der Summe ausfällt, desto sinnvoller, sicherer und günstiger ist die Freigabe des Datensatzes als Open Data:

| >p0.2 | >p0.2 | >p0.6 |

Herausforderung | **Erläuterung** | **Bewertung**

1. Strategie Welches ist die strategische Bedeutung der Datenfreigabe als Open Data? 0: Einzigartige wettbewerbsdifferenzierende Daten bzw. gesetzliche Gebührenpflicht

- 1: Wesentliche strategische Nachteile
- 2: Unwesentliche strategische Nachteile
- 3: Keine strategischen Nachteile
- 4: Keine strategischen Nachteile und strategische Vorteile möglich
- 5: Keine strategischen Nachteile und sicher strategische Vorteile

2. Datenschutz Handelt es sich um personenbezogene Daten bzw. wird der Datenschutz verletzt? 0: Personenbezogene Daten

- 1: Nicht anonymisierbare Daten, fehlende Zustimmung kaum einholbar
- 2: Nicht anonymisierbare Daten, fehlende Zustimmung einholbar
- 3: Zustimmung zur Veröffentlichung vorhanden
- 4: Anonymisierbare Daten
- 5: Kein Rückschluss auf Personen oder Unternehmen ableitbar, bzw. keine Verletzung schutzwürdiger Geheimhaltungsinteressen

3. Sicherheit Unterliegen die Daten Geheimhaltungspflichten oder tangieren sie die Sicherheit auf andere Weise? 0: Geheimhaltungspflicht gegeben

- 1: Einschränkungen vorhanden, kaum änderbar (z.B.: EU-Vorgaben)
- 2: Einschränkungen vorhanden, änderbar (z.B.: Gesetz)
- 3: Einschränkungen vorhanden, leicht änderbar (z.B.: Verordnung)
- 4: Einschränkungen vorhanden, sehr leicht änderbar (z.B.: interne Richtlinien, Verwaltungskultur)

⁴ BERNHARD KRABINA/THOMAS PROROK/BRIGITTE LUTZ (2016), Open-Government-Vorgehensmodell 3.0 – Umsetzung von Open Government. KDZ Zentrum für Verwaltungsforschung.

5: keine Einschränkungen

4. Datenqualität Wie hoch wird die Datenqualität eingeschätzt? 0: Datenqualität nicht vertretbar

- 1: Die Datenqualität ist sehr gering
- 2: Die Datenqualität ist gering
- 3: Die Datenqualität ist durchschnittlich
- 4: Die Datenqualität ist hoch
- 5: Die Datenqualität ist sehr hoch

5. Externes Interesse Wie hoch wird das Interesse bei den Zielgruppen eingeschätzt? 1: Das externe Interesse ist sehr gering

- 2: Das externe Interesse ist gering
- 3: Das externe Interesse ist durchschnittlich
- 4: Das externe Interesse ist hoch
- 5: Das externe Interesse ist sehr hoch

6. Aufwand Wie hoch ist der Aufwand für die Veröffentlichung? 0: Aufwand nicht vertretbar

- 1: Der Aufwand ist sehr hoch
 - 2: Der Aufwand ist hoch
 - 3: Der Aufwand ist durchschnittlich
 - 4: Der Aufwand ist gering
 - 5: Der Aufwand ist sehr gering
-

[Rz 17] Mit diesem Bewertungsschema lassen sich die sechs relevanten Herausforderungen von Open Data Freigaben evaluieren um systematisch zu bestimmen, ob eine Datenfreigabe überhaupt möglich ist und wenn ja, ob mit geringem oder hohem Aufwand gerechnet werden muss. Mit dem Wert «0» wird ein Ausschlussgrund für einen Datensatz aufgezeigt, sodass er in dieser Form nicht freigegeben werden kann.

[Rz 18] Gleichzeitig bietet dieses Verfahren auch an, eine Priorisierung der vorhandenen, noch nicht veröffentlichten Datenbestände vorzunehmen. So lassen sich nach der 80:20 Regel (Pareto-Prinzip) zunächst diejenigen Datensätze identifizieren, deren Freigabe mit relativ wenig Aufwand und wenig Risiko vorgenommen werden können.

5. Scheinargumente gegen Open Data

[Rz 19] Während die genannten sechs Herausforderungen tatsächliche Problemfelder der Datenfreigabe behandeln, die in jedem Fall ernst genommen werden sollten, werden in Entscheidungsprozessen und Diskussionen aus Unwissenheit oder aus Angst vor Transparenz oftmals auch fadenscheinige Argumente gegen Open Data ins Feld geführt.

[Rz 20] Ein typisches derartiges Scheinargument betrifft die Haftung. Organisationen befürchten, dass sie für Verluste, Verletzungen oder Schäden bei der Verwendung von veröffentlichten Daten haftbar gemacht werden könnten. Dies ist prinzipiell nicht möglich, denn letztlich trägt der Datennutzer die Verantwortung für die Verwendung der Daten. Eine Behörde oder eine Firma kann nicht für falsche Zahlen haftbar gemacht werden, denn es ist stets der Datennutzer, der die Daten in einem bestimmten Kontext interpretiert und daraus seine Schlüsse zieht. Auf diesen

Umstand kann mit entsprechenden Disclaimern auf den jeweiligen Datenportalen ausdrücklich hingewiesen werden.

[Rz 21] Ein weiteres, oft gehörtes Argument betrifft die Möglichkeit von Fehlinterpretationen von Rohdaten. Dieses Scheinargument impliziert, dass die Bevölkerung zu wenig von den Daten verstehe und deshalb falsche Schlüsse ziehen könnte. Dieser Befürchtung liegt zu Grunde, dass die Organisation durch die Abgabe der Deutungshoheit über die betroffenen Daten Falschinformationen ermöglicht und letztlich Reputationsschäden befürchten muss. In der Tat ist es so, dass für die Interpretation von Datensätzen oftmals ein gewisses Kontextwissen notwendig ist. Dieses kann jedoch durchaus mit den entsprechenden Hinweisen, Dokumentationen, Begleitinformationen, Leitfäden und FAQs hergestellt werden, sodass interessierte Personen rasch ein fundiertes Verständnis der Daten erhalten. Tatsächlich hat die Praxis bei Datenfreigaben gezeigt, dass kaum jemand bewusst oder fahrlässig falsche Schlüsse zieht, wenn umfassende Datensätze zugänglich gemacht werden und diese ausreichend mit Beschreibungen und Hintergrundinformationen versehen sind und gut kommuniziert werden. Gerade die Medien sind mit ihren spezialisierten Datenjournalisten heute mehr denn je darauf bedacht, zugänglich gemachte Daten sorgfältig und nach bestem Wissen und Gewissen zu analysieren und zu interpretieren.

[Rz 22] Und letztlich wird manchmal beanstandet, dass aufgrund frei zugänglicher Daten bspw. gewisse Bevölkerungsgruppen oder Wohnbezirke plötzlich diskriminiert und dadurch noch weiter stigmatisiert werden könnten. Diesem Aspekt negativer Folgen von mehr Transparenz muss durchaus Beachtung geschenkt werden. Allerdings zeigt gerade diese Angst vor der Veröffentlichung fundierter Daten auf, dass hier wohl sehr wichtige Anhaltspunkte berührt werden, an denen ein hohes gesellschaftliches Interesse besteht. So kann umgekehrt argumentiert werden, dass besonders bei solchen heiklen Daten auch ein hoher Nutzen daraus gezogen werden kann, wenn daraus Verbesserungsmassnahmen resultieren und Missstände behoben werden. Wird bspw. in einem Wohnquartier eine besonders hohe Kriminalitätsrate aufgrund der Veröffentlichung von Polizeirapporten festgestellt, können gezielte Sicherheitsmassnahmen und andere Verbesserungen ergriffen werden.

6. Schlussfolgerungen

[Rz 23] Die genannten Punkte bei der Datenfreigabe konnten aufzeigen, welche Argumente tatsächlich relevante Herausforderungen darstellen und welche typischerweise als Ausrede verwendet werden. Beiden Seiten sollte Aufmerksamkeit verliehen werden, denn auch mit Scheinargumenten lassen sich Open Data Freigaben erfolgreich verhindern.

[Rz 24] Des Weiteren dient das erläuterte Bewertungsschema als Governance-Hilfsmittel um noch nicht veröffentlichte Datensätze bewerten zu können. Dies schafft Anhaltspunkte über Nutzen, Aufwand und Risiko bei Open Data Freigaben. Mit diesen Grundlagen lassen sich Entscheidungsprozesse versachlichen was der bewussten Freigabe und letztlich der Förderung von Open Data dient.

MATTHIAS STÜRMER (Dr. sc. ETH Zürich, lic.rer.pol.) ist Leiter der Forschungsstelle Digitale Nachhaltigkeit am Institut für Wirtschaftsinformatik der Universität Bern und ist dort als Dozent, Forscher und Berater zu Open Data, Open Government, Open Source Software und ICT-Beschaffungen tätig. Er ist ausserdem Geschäftsleiter der Parlamentarischen Gruppe Digitale

Nachhaltigkeit, Präsident von tcbe.ch – ICT Cluster Bern, Vorstandsmitglied der Vereine CH Open und Opendata.ch sowie Stadtrat von Bern.