

Einwände gegen die Veröffentlichung von Open Data und bewährte Antworten darauf

(oder: „Wie man andere davon überzeugt, ihre Daten zu teilen“)

Version 1.0, veröffentlicht am 23. Mai 2013

Inhalt

Unsere Daten könnten missinterpretiert werden	2
Die Daten sind bereits auf unserer Webseite verfügbar.....	2
Wir werden Unmengen an Rückfragen bekommen	2
Wir wollen die Daten in einem Forschungspapier verwenden	2
Die Daten sind zu umfangreich.....	3
Dann werden wir ja jede Menge Spam-E-mails erhalten!.....	3
Die Daten sind nicht sonderlich interessant	3
Terroristen könnten die Daten verwenden	4
Wir machen uns Sorgen um den Datenschutz.....	4
Es gibt keine API für das Quellsystem	5
Wir sind uns nicht sicher, ob uns die Daten gehören	5
Ich hätte nichts gegen eine Veröffentlichung, fürchte aber, dass jemand anderer dies beeinsprucht	6
Unsere Datenqualität ist zu schlecht	6
Wir haben viel zu tun und die Priorität hierfür ist nicht hoch genug	6
Unser Anwälte wollen, dass wir eine eigene Lizenz verwenden	7
Die Daten ändern sich zu schnell	7
Wir führen schon ein ähnliches Projekt durch	7
Teile der Daten sind vertraulich	7
Mir gehören die Daten nicht, deshalb kann ich sie auch nicht freigeben	8
Wir haben diese Daten nicht.....	8
Weil ein Teil der Freigabe nicht möglich ist, können wir die kompletten Daten nicht anbieten	9
Was passiert wenn die Gültigkeit der offenen Daten plötzlich nicht mehr gegeben ist?	9
Wir sehen hier keinen Nutzen	9
Aber was ist wenn wir die Daten verkaufen möchten?	9
Wenn wir die Daten veröffentlichen, können wir verklagt werden	10
Wir möchten die Kontrolle über die Daten behalten und Nutzer überprüfen	10
Die Daten unterliegen dem Amtsgeheimnis.....	10

Ergänzungen/Kommentare sind unter <http://open3.at/einwaende> willkommen!

Unsere Daten könnten missinterpretiert werden

- Die Angabe von Informationen wie die Interpretation laut Datenerheber erfolgen kann (z.B. durch Details über Erhebungsmethoden), kann dem entgegenwirken
- Man könnte sich darauf vorbereiten, Personen zu helfen, die lt. Datenbereitsteller die Daten falsch interpretieren – vor allem jene, welchen dies irrtümlich passiert ist, werden dankbar für die Hilfe sein.
- Eine Veröffentlichung kann unter manchen Umständen ganz im Gegenteil dazu beitragen, Missinterpretationen im Vorfeld zu entgegnen (z.B. durch Auszüge von Daten die im Rahmen eines Informationsfreiheitsgesetzes zugänglich gemacht wurden) – dadurch kann schnell auf die kompletten Rohdaten im Web verwiesen werden, um falsche Interpretationen zu widerlegen.

Die Daten sind bereits auf unserer Webseite verfügbar

- Ja, aber auch in maschinenlesbarer Form und nicht nur z.B. als PDF?
- Hilfreich kann es sein, bestehende Daten zu scrapen und einen Showcase zu erstellen, bei dem diese Daten wieder maschinenlesbar angeboten werden

Wir werden Unmengen an Rückfragen bekommen

- Dies ist generell ein Einwand von Personen, die sich überarbeitet fühlen und diese Tätigkeit nicht als Aufgabe ihres Jobs verstehen. Hier kann die Eskalation an eine höhere Management-Ebene dabei helfen festzustellen, ob dieser Einwand gerechtfertigt ist.

Wir wollen die Daten in einem Forschungspapier verwenden

- Eine Option hierfür wäre ein automatische oder optionale Vereinbarung/Verpflichtung, welche die Bereitstellung der verwendeten Daten X Monate nach Veröffentlichung des Forschungspapieres vorsieht.

Auch die Option, bei Bedarf eine Verlängerung dieser Datensperfrist vorzusehen, könnte angedacht werden, um bei geänderten Rahmenbedingungen weiterhin die Daten vorerst nicht zugänglich zu machen.

Die Daten sind zu umfangreich

- Die Größe ist mit hoher Wahrscheinlichkeit kein echtes Problem (außer man arbeitet z.B. beim Teilchenbeschleuniger CERN, wo 75 Petabytes – was 75 Millionen Gigabyte entspricht – in den letzten 3 Jahren verarbeitet wurden).
- Für die Daten verantwortliche Personen sind oft stolz auf die Größe der von ihnen verwalteten Daten – die Herstellung einer Relation zu größeren Datenmengen sollte daher behutsam erfolgen.
- Zielführend kann es sein zu fragen, ob einige Experimente durchgeführt werden könnten. Die Ergebnisse könnten dann verwendet werden um zu zeigen, dass die Datenmenge praktikabel verarbeitet werden könnte.

Dann werden wir ja jede Menge Spam-Emails erhalten!

- Email-Adressen müssen nicht in den Daten inkludiert werden
- Oft sind Email-Adressen bereits auf der Webseite veröffentlicht
- Spam-Filter schaffen Abhilfe
- Falls doch Email-Adressen in den Daten enthalten sein sollen, könnten die betroffenen Empfänger per opt-in oder opt-out entscheiden, ob ihre Email-Adresse sichtbar sein soll

Die Daten sind nicht sonderlich interessant

- Schon alleine die Tatsache, dass man gerade Interesse an den Daten bekundet zeigt, dass diese interessant sind. Eine Aufklärung warum genau man sich für diese Daten interessiert, kann hier hilfreich sein.
- Ebenfalls hilfreich sind potentielle Anwendungsfälle/Usecases, die aufzeigen, wie die Integration mit anderen Datensätzen zu einem Mehrwert führen könnte. Z.B. könnte die Daten zur Frage „Welche Universität bietet welche Kurse an?“ auf einer Karte dargestellt werden und zusätzlich auch mit den Gebäuden verknüpft werden, in denen diese Kurse stattfinden – was va. für neue Studenten eine große Erleichterung darstellen würde.
- Die Bewertung der Nützlichkeit von Datensätzen sollte auch von externen Personen erfolgen – auch für Nischen-Datensätze gibt es interessierte Personen und über das Internet/soziale Medien sind diese leicht ausfindig zu machen.

Hinsichtlich Nischendatensätze ist es aber empfehlenswert, zuerst die Veröffentlichung von „low-hanging-fruits-Datensätzen“ zu initiieren, damit auch rascher Erfolge für den Datenbereitsteller sichtbar werden.

Terroristen könnten die Daten verwenden

- Dieser Einwand kommt gewöhnlich von Personen, die keinen anderen Grund finden und nicht aktiv werden wollen
- Für gewisse Datensätze ist dieser Einwand natürlich gültig (z.B. bei Infos über Kernenergie, tlw. Chemie/Biologie oder Waffen) oder aber auch bei Informationen, die Personen zur Zielscheibe machen könnten, die in einem kontroversiellen Gebiet wie z.B. Tierversuche arbeiten. Das sind legitime Einwände.
- Die Daten könnten bereits jetzt mit etwas mehr Aufwand beschafft werden. Als Beispiel seien hier Sammelpunkte im Falle eines Feuers genannt. Diese Information ist bereits jetzt meist auf den Wänden der betreffenden Gebäude als Aushang ersichtlich. Terroristen könnten also schon jetzt in diese Gebäude gehen und die Pläne kopieren – oder mit anderen Worten ausgedrückt: eine motivierte Person kann diese Information bereits jetzt ausfindig machen – und es gibt kein bzw. maximal ein geringes Risiko, diese Daten auch in einer nützlicheren Form (z.B. durch maschinenlesbare Bereitstellung) verfügbar zu machen.
- Falls dies nicht zutreffen sollte - gibt es eventuell eine Untermenge dieser Daten die stattdessen veröffentlicht werden können, um das Risiko zu reduzieren oder ganz zu eliminieren?
- Eine Analyse von ein paar betroffenen Datensätze könnte helfen, hier eine Lösung zu finden

Wir machen uns Sorgen um den Datenschutz

- Der Datenschutz umfasst (in Österreich) natürliche und als europäischer Sonderfall auch juristische Personen. Open Data bedeutet an sich ist per Definition die Bereitstellung von nicht-personenbezogenen Informationen – alternative Informationen sind hiervon nicht erfasst.
- Das Aufzeigen, was bereits jetzt in nicht-maschinenlesbarer Form veröffentlicht wird, kann helfen, Bedenken auszuräumen.

- Sollten personenbezogene Informationen in den Daten vorhanden sein, können diese gestrichen, aggregiert oder anonymisiert werden.

Es gibt keine API für das Quellsystem

- Es braucht nicht zwingend eine öffentliche API/Schnittstelle (zumindest für statische Daten). Bsp. SQL-Datenbanken: es könnte auch ausreichend sein, die Daten als Flatfiles zu exportieren und diese im Web statisch zur Verfügung zu stellen, wo diese dann von Anwendern weiter transformiert werden könnten.
- Ein automatisiertes Abgreifen der Daten über die Webseite (Scraping) kann ev. auch jetzt schon möglich sein. Durch eine API könnte auch die Last auf den Webserver verringert werden.
- Eine Einbeziehung des Anbieters/Entwicklers des Quellsystems kann hilfreich sein. Die meisten Anbieter sind daran interessiert, ihre Systeme für die neuesten technischen Entwicklungen und Trends fit zu machen und z.B. eine Open Data-API in ihre Standardprodukte zu integrieren.

Wir sind uns nicht sicher, ob uns die Daten gehören

- Dies ist ein Thema für Datenbereitsteller wie z.B. Bibliothekskataloge, bei denen z.B. die Metadaten oft aus kostenpflichtigen Systemen stammen, welche ein offenes Wiederveröffentlichen nicht erlauben. Oft es ist nicht klar, von wo genau ein Datensatz stammt und daher ist der gesamte betroffene Datensatz „unrein“.
- Gibt es Datensätze, wo sich der Datenbereitsteller sicher ist? Könnten diese veröffentlicht werden?
- Wenn die Datensätze nicht dem Bereitsteller gehören, ist dann überhaupt bekannt, wem sie gehören oder handelt es sich eventuell um verwaiste Datensätze?
- Gibt es weitere Datensätze, von denen der Bereitsteller definitiv weiß, dass er bei diesen die vollen Nutzungsrechte besitzt?

Ich hätte nichts gegen eine Veröffentlichung, fürchte aber, dass jemand anderer dies beeinträchtigt

- Hierbei handelt es sich meist um einen vorgeschobenen Einwand als um ein wirkliches Problem.
- Der Einwand deutet darauf hin, dass die Person fürchtet, für Fehler im Rahmen der Veröffentlichung persönlich verantwortlich gemacht zu werden.
- Dem kann z.B. entgegengewirkt werden, indem z.B. vom Management die Zusage eingeholt wird, dass diese Person keine Probleme durch eine Veröffentlichung bekommen wird.
- Die Konzentration auf weniger kontroversielle Datensätze könnte hier auch helfen.

Unsere Datenqualität ist zu schlecht

- Mehr Datennutzer tragen durch Feedback dazu bei, dass die Qualität verbessert werden kann
- Unterstützungsleistungen für die Verbesserung der Datenqualität sind hilfreich – Systeme zur Messung der Datenqualität können dies ebenfalls unterstützen.
- Die Veröffentlichung von Daten auf Webseiten ist ein guter Weg, um auch technisch-unversierteren Personen die Möglichkeit zu geben, an der Verbesserung der Daten mitzuarbeiten. Solange die Daten jedoch nicht öffentlich einsehbar sind, werden z.B. Feedback-E-mails weniger beachtet und die Datenqualität oft nicht verbessert. Sind diese jedoch online verfügbar, hat die datenbereitstellende Organisation ein eigenes Interesse daran, Fehler in publizierten Daten zeitnah zu korrigieren.

Wir haben viel zu tun und die Priorität hierfür ist nicht hoch genug

- Womit genau ist die Organisation beschäftigt? Eventuell gibt es Anknüpfungspunkte bei denen geholfen werden kann – z.B. dass durch die Bereitstellung von Daten könnten Karten/Tools/Apps/Visualisierungen durch die Community erstellt werden, welche wiederum für die Organisation selbst hilfreich sein könnten.

Unser Anwälte wollen, dass wir eine eigene Lizenz verwenden

- Dieser Einwand kommt oft, wenn den Anwälten nicht im Detail erklärt wurde, was genau mit der Veröffentlichung von Daten geplant ist.
- Anwälten fehlt oft das Wissen über freie Lizenzen wie Creative Commons CC-BY und ein direkter Kontakt mit den zuständigen Personen.
- Präzedenzfälle können helfen – va. wenn aufgezeigt werden kann, dass bereits andere ähnliche Organisationen Daten unter einer offenen Lizenz veröffentlicht haben.

Die Daten ändern sich zu schnell

- Echtzeit-APIs bieten sich hier als Lösung an
- Alternativ könnten ev. nur jene Daten veröffentlicht werden, die sich nicht so schnell ändern
- Ebenso denkbar wäre die Veröffentlichung von Metadaten mit Links zu einer maschinenlesbaren Version der Daten (oder einer API), die weniger häufig zu pflegen sind

Wir führen schon ein ähnliches Projekt durch

- Derartige Projekte haben oft eine Projektlaufzeit von mehreren Monaten oder sogar Jahren – vielleicht gibt es einen Ansatzpunkt, etwas jetzt mit geringem Aufwand umzusetzen.
- Der Output des bestehenden Projekts ist meist z.B. ein Tool oder eine mobile App... – das ist nicht dasselbe wie die Rohdaten zur Verfügung zu stellen.
- Hilfreich kann sein herauszuarbeiten, wie die Extraktion der Daten aus dem Projekt als Open Data das Projekt selbst unterstützen und späteren Projekt-Aufwand reduzieren könnte.

Teile der Daten sind vertraulich

- Welche Teile? Können diese entfernt werden, ohne die Nützlichkeit der übrigen Daten zu stark einzuschränken?
- Sind die Daten wirklich vertraulich? Sind sie ev. schon in anderer Form (z.B. im Web oder offline) verfügbar?

Mir gehören die Daten nicht, deshalb kann ich sie auch nicht freigeben

- Falls Eigentümer nicht bekannt, ist hilfreich rauszufinden, wem die Daten überhaupt gehören
- Wenn eine Person die Aufgaben einer anderen übernommen hat, ist oft unbekannt, dass die neue verantwortliche Person auch meist Dateneigentümer ist.
- Ist das Eigentum an den Daten unklar, kann ev. das Management bei der Klärung helfen.
- Für verwaiste Daten empfiehlt es sich, einen erfahrenen (Daten)manager hinzuzuziehen, der Entscheidungen über deren Freigabe treffen kann.

Wir haben diese Daten nicht

- Manche Organisationen haben die Daten, wissen es aber nicht – nach dem Motto: „Oh, das sind nicht Daten – das ist nur eine Exceltabelle“.
- Hat die Organisation eventuell andere interessante Daten?

Diese Daten werden bereits von externer Organisation XY veröffentlicht

- V.a. externe Organisationen & Datenaggregatoren veröffentlichen Daten selbst nicht unter offenen Lizenzen und schränken dadurch den potentiellen Nutzen ein. Denkbar wäre auch eine Wiederveröffentlichung durch den sekundären Nutzer, sofern dies möglich ist.
- Indem Organisationen Daten selbst veröffentlichen, haben sie mehr Kontrolle darüber und weniger Abhängigkeiten zur externen Organisation.
- Ungenauigkeiten im publizierten Daten könnten dann in der Organisation selbst (in-house) ausgebessert werden, was wiederum zu genaueren Daten führen würde.
- Können die Daten von der externen Organisation besorgt werden und mit anderen eigenen Daten kombiniert werden? Falls nicht, ist dies im Regelfall ein (nicht erkanntes?) Problem für die Organisation.
- Können dieselben Datensätze, welche an die externe Organisation gesendet werden, zusätzlich auch selbst im Web veröffentlicht werden?

Weil ein Teil der Freigabe nicht möglich ist, können wir die kompletten Daten nicht anbieten

- Erfahrungen zeigen, dass die Freigabe kompletter Datensätze manchmal abgelehnt wird, weil ein kleiner Teil der Daten nicht verfügbar oder zu kontroversiell erscheint.
- Die Prüfung, ob eine Freigabe auch ohne dieser Teile möglich ist, kann hilfreich sein

Was passiert wenn die Gültigkeit der offenen Daten plötzlich nicht mehr gegeben ist?

- Vorausschauende Planung kann dies verhindern (z.B. durch Monitoring von Timestamps)
- Haftungsausschlüsse könnten zu den Daten hinzugefügt werden
- Wenn Open Data in die internen Prozesse integriert wird, bedeutet dies, dass ungültige Datensätze schneller erkannt werden könnten.

Wir sehen hier keinen Nutzen

- Beispiele ähnlicher Organisationen bieten hier Ansatzpunkte, um einen Nutzen für den Datenbereitsteller aufzuzeigen.

Aber was ist wenn wir die Daten verkaufen möchten?

- Wenn der Verkauf der Daten Kern des Geschäftsmodells der Organisation ist, ist dies ein gültiger Einwand.
- Um eine Analogie zu verwenden: ein Pizza-Zustelldienst würde sehr schnell bankrott sein, wenn er entscheiden würde, sein Essen gratis abzugeben – aber auch, wenn er seine Speisekarte nicht kostenlos zur Verfügung stellen würde. Bei seinen Rezepten kann es sich um wertvolle Geheimnisse handeln, allerdings sind diese auch von großem Interesse für Leute mit Nahrungsunverträglichkeiten und könnten zu einem höheren Umsatz führen, weil diese genau wissen wollen/müssen, welche Zutaten für das Essen verwendet wurde.
- Hilfreich ist es auch zu identifizieren, ob dieser Einwand realistisch ist oder ob dieser nur im Zusammenhang mit einer generellen Risiko-Aversion geäußert wurde
- Organisation könnten dabei unterstützt werden, praktische Vorteile durch das Öffnen von Daten für sie zu erarbeiten.

Wenn wir die Daten veröffentlichen, können wir verklagt werden

- Wenn die Sorge Informationen über Personen betrifft, ist die Erstellung ein Einhaltung einer Richtlinie empfehlenswert (z.B. „wir korrigieren Fehler innerhalb von 7 Tagen nach Bekanntwerden“).
- Wenn die Sorge Fehler betrifft, welche eine Auswirkung auf Unternehmen haben könnten, könnte Präzedenzfälle hilfreich sein (allerdings gibt es hier noch nicht allzu viele).
- Indem der Veröffentlichungsstatus von Daten auch im Nachhinein nachvollzogen werden kann, können Organisationen besser mit nachträglichen Rechtsansprüche umgehen (Bsp. wenn jemand behaupten würde, er hätte das nächstgelegene Spital nicht gefunden weil es nicht auf der Karte eingezeichnet war und deswegen ist jemand gestorben).
- Wenn die Daten bereits auf der Webseite oder über eine eigene App abrufbar sind: wird durch die zusätzliche maschinenlesbare Bereitstellung der Daten das Risiko verklagt zu werden wirklich erhöht?

Wir möchten die Kontrolle über die Daten behalten und Nutzer überprüfen

- Dieser Einwand ist ähnlich wie „Unsere Daten könnten missinterpretiert werden“.
- Könnten als Minimalversion zumindest die Metadaten oder ein Auszug der Daten zur Verfügung gestellt werden?

Die Daten unterliegen dem Amtsgeheimnis

- Österreichische Besonderheit in Verfassungsrang, welche nach Stand 23. Mai 2013 und Intention der aktuellen Regierungsparteienⁱ spätestens 2015 durch ein Informationsfreiheitsgesetz ersetzt werden und ein Recht auf Informationen als Standard festsetzen soll. Aktuelle Infos hierzu unter <http://blog.transparenzgesetz.at>
- Mittelfristig ist durch die Umsetzung der überarbeiteten PSI-Richtlinie der EU in nationales Recht ebenso zu erwarten, dass Bürger ein gestärktes Recht auf den Zugang zu Informationen erhalten. Details unter <http://www.open3.at/2013/04/eu-richtlinie-zu-open-government-data-vor-beschluss-im-eu-parlament>.

ⁱ <http://derstandard.at/1363711872457/Koalition-bei-Informationsfreiheit-weitgehend-einig>,
besucht am 23.5.2013