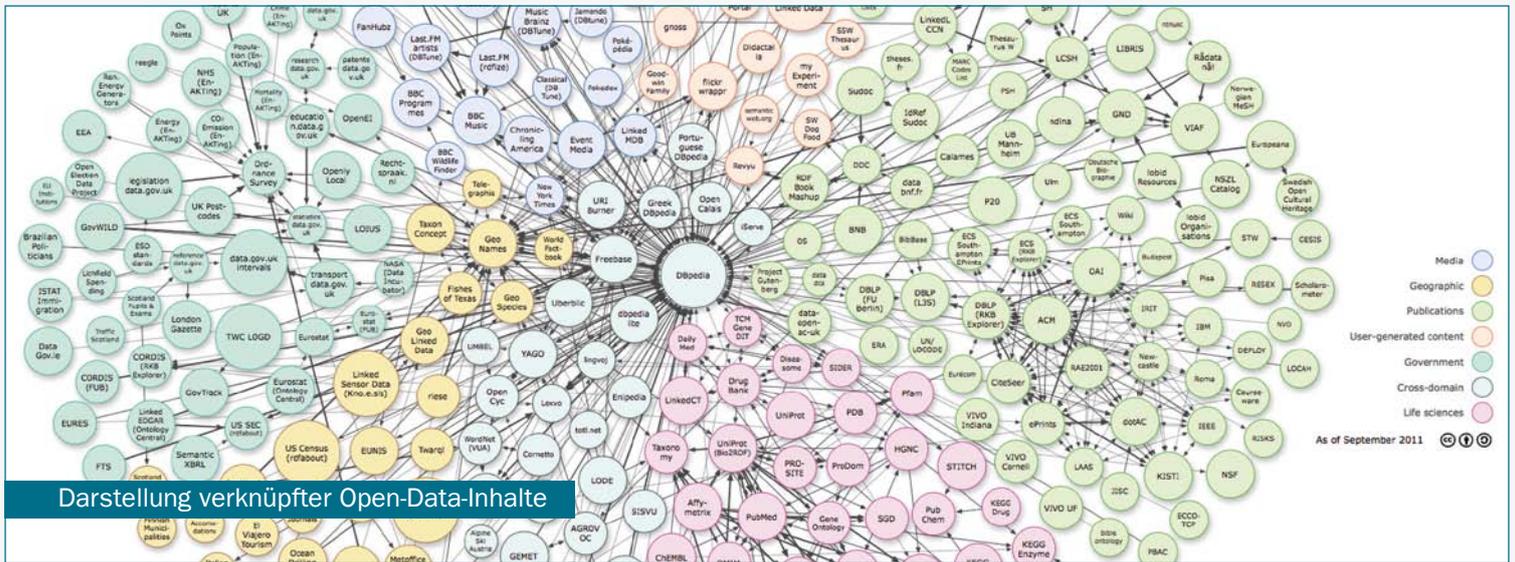


IM BLICKPUNKT: Open Data



Wie entsteht Open Data?
Partizipation durch Open Data
Open Data auf der Agenda
Open Data nutzen
Visualisieren & Verknüpfen
Data Mining & Daten-Journalismus
Wirtschaftliches Potenzial
Barrieren überwinden
(Urheber-)Rechte
Format, Auffindbarkeit, Kosten





Darstellung verknüpfter Open-Data-Inhalte

Open Data

Früher füllten Aktenordner meterweise die Regale, heute werden Texte und Zahlen einfach auf Festplatten, USB-Sticks und SD-Karten gespeichert. Dieser Wandel vom Aktenordner zum Datenträger hat bewirkt, dass die meisten (Verwaltungs-)Daten in digitaler Form vorliegen. Das bedeutet auch, dass sie sich durch IT-Technik bequem weitergeben und verarbeiten lassen.

So ist die Idee von Open Data (den offenen Daten) populär geworden: Gesammelte Daten sollen von öffentlichen Stellen freigegeben werden und zu jeglichen Zwecken weiterverwendbar sein – so die Forderung vieler Open-Data-Initiativen und Bürger, die sich selbst informieren wollen, um Gesellschaft und Politik mitzugestalten.

Mittlerweile steht Open Data auf der politischen Agenda, denn vor allem der öffentliche Sektor verfügt über viele Daten, die offen sein können. Noch haben die politischen und gesellschaftlichen Initiativen nicht zum möglichen Erfolg von Open Data in Deutschland geführt, aber der Durchbruch soll in den kommenden Jahren gelingen. Zudem wird es durch den technischen Fortschritt möglich, immer mehr Daten zu erheben, zu speichern und zu analysieren. So rechnet die Europäische Kommission mit einem jährlichen weltweiten Wachstum dieser Informationen um 40 Prozent.

Den aktuellen Stand beschreibt IM BLICKPUNKT: Open Data, erläutert werden die verschiedenen Nutzungsmöglichkeiten und auch die Barrieren, die noch zu überwinden sind. Zuletzt wird ein Blick in die Zukunft und auf die Konsequenzen von funktionierender Open Data geworfen.

Wie entsteht Open Data?

Open Data meint die (nahezu) uneingeschränkte Verfügbarkeit von Informationen, die in digitaler Form vorliegen. Es handelt sich vor allem um Datensätze, nicht um Textmaterial wie Dokumente und Akten. Sie können von jedem Menschen und allen Unternehmen genutzt und weiterverwendet werden – unabhängig, ob es sich um private oder kommerzielle Zwecke handelt. Voraus-

setzung ist ein grundsätzlich freier Zugang zu den Daten, für den die amerikanische Bildungsorganisation Sunlight Foundation einige Prinzipien definiert hat:

- **Vollständig:** Es werden alle Daten, die keinen Beschränkungen unterliegen, freigegeben.
- **Primär:** Es handelt sich um Rohdaten, die nicht zusammengefasst oder bearbeitet sind.
- **Zeitnah:** Die Daten werden so schnell wie möglich freigegeben, um aktuell zu sein.
- **Zugänglich und nicht diskriminierend:** Möglichst viele Menschen und Unternehmen sollen Zugriff haben, dafür müssen die freigegebenen Daten leicht zu finden, ohne Registrierung zu nutzen und kostenlos sein.
- **Maschinenlesbar und nicht proprietär:** Es werden Dateiformate verwendet, die Computer automatisch verarbeiten können und die nicht von einzelnen Unternehmen (zum Beispiel Microsoft oder Apple) kontrolliert werden.
- **Lizenzfrei:** Die Daten unterliegen keinen Urheberrechten oder anderen Nutzungsbeschränkungen, so dass Modifikationen und Derivate erlaubt sind.

Diese Prinzipien gelten als Parameter für den jeweiligen Entwicklungsstand von Open Data, denn nur unter diesen Voraussetzungen können Daten weitergegeben und weiterverwendet werden. Somit sind Informationen, die im Internet veröffentlicht sind, nicht automatisch offene Daten.

Ausgenommen von Open Data sind sicherheitsrelevante und personenbezogene Daten, so dass der Datenschutz nicht berührt wird. Auch Geschäftsgeheimnisse und ähnliches bleiben geschützt. Anders als Plattformen wie „Wikileaks“ zielt Open Data nicht auf die anonyme Veröffentlichung von vertraulichen Informationen; die Quelle und Legalität sollten nicht fraglich sein, schreibt etwa Christiane Schulzki-Haddouti im Open-Data-Dossier der Bundeszentrale für politische Bildung.

Wegen der Gesetze zum Datenschutz und Bestimmungen zur staatlichen Sicherheit sowie geschäftlichen Geheimnissen werden auch zukünftig nicht alle Informationen frei zugänglich sein.

Open Data stützt sich hauptsächlich auf vorhandene Datensätze. Jedoch ist es auch möglich, Informationen selbst zu generieren, beispielsweise durch (mobiles) Crowdsourcing. Diese Daten können sehr interessant sein, entsprechen aber nicht statistischen Standards und sind von der freiwilligen Beteiligung anderer Internetnutzer abhängig.

Die meisten Datensätze, die zu Open Data werden können, befinden sich im Politik- und Verwaltungsbereich, da Behörden und Ämter für ihre Arbeit Daten sammeln, erstellen und erwerben. Diese reichen von der Verwendung von Steuergeldern und Haushaltsetats über Umwelt- und Medizindaten bis zu Verkehrsstatistiken und Kartenmaterial. Zudem verfügen wissenschaftliche Einrichtungen, Nichtregierungs-Organisationen und Unternehmen über große für Open Data geeignete Datenmengen. Im Fokus der politischen Agenda stehen indessen die Daten des öffentlichen Sektors.

Partizipation durch Open Data

Die Freigabe und Weiterverwendung von Daten ermöglicht, dass Bürger sich besser über das politische Geschehen informieren und sich im nächsten Schritt auch stärker beteiligen können. Open Data erlaubt es, politische Vorgänge zu kontrollieren, zu interpretieren und zu bewerten. Damit wird einerseits Demokratie gefördert, andererseits Korruption erschwert.

Durch Open Data lassen sich viele Prozesse nachvollziehen: Wofür werden Steuergelder ausgegeben? Wer trifft welche Entscheidungen? So kann Open Data zu einer transparenteren Politik führen und Bürger bei ihren politischen Entscheidungen unterstützen.

Noch seien die demokratischen Effekte von Open Data nicht belegt, schreiben Noor Huijboom und Tijs van den Broek (European Journal for ePractice). Mögliche Gründe sind, dass es sich um eine recht junge Idee handelt und noch einige Barrieren zu überwinden sind, damit Open Data wirken kann. Behörden und Ämter, die ihre Daten freigeben, können sich ein offenes und effizientes Image verschaffen. Schließlich sorgen die Standards für Open Data auch für einen einfachen Austausch innerhalb des öffentlichen Sektors. Dann können Daten aktuell und an verschiedenen Stellen

der Verwaltung genutzt werden, ohne dass eine rechtliche oder technische Abstimmung nötig ist.

Open Data auf der Agenda

Einen ersten Schritt zu Open Data hat die Europäische Union mit der Richtlinie über die Weiterverwendung von Informationen des öffentlichen Sektors (PSI-Richtlinie) im Jahr 2003 getan. Sie vereinbart ein Minimum an Bedingungen, um einen fairen Wettbewerb für Open Data in Europa zu schaffen. Jedoch sind die Vorgaben zu Formaten, Lizenzen und Gebühren so formuliert, dass die PSI-Richtlinie in den Mitgliedsstaaten nicht einheitlich umgesetzt wurde und sich Unterschiede in der Praxis von Open Data ergeben. Im Jahr 2011 stellte die EU eine Open Data-Strategie vor: Dazu gehörte ein neuer Vorschlag für die PSI-Richtlinie, um rechtliche, technische und finanzielle Beschränkungen weiter zu verringern. Es soll vorausgesetzt sein, dass Daten des öffentlichen Sektors offen und weiterverwendbar sind, Einrichtungen wie Bibliotheken, Museen und Archive werden eingeschlossen.

In Deutschland wurde erst 2006 das Gesetz zur Informationsfreiheit (IFG) verabschiedet. Es beschreibt das Recht, Zugriff auf Daten und Informationen zu beantragen, ohne dass eine Begründung nötig ist: Die Behörden sind verpflichtet, jeden Antrag zu beantworten. Das IFG beschränkt sich nicht auf digitalisierte Daten und regelt auch nicht die Weiterverwendung der erhaltenen Daten und Informationen. Zusätzlich existieren das Verbraucher- und das Umweltinformationsgesetz, die den Zugriff auf entsprechende Daten ermöglichen. Auch die rot-grüne Regierung in NRW bekennt sich im Koalitionsvertrag von 2012 zu den Prinzipien von Open Data, die auch für den Hochschul- und Wissenschaftsbereich umgesetzt werden sollen.

Open Data nutzen

Open Data ist zunächst ein Rohstoff, denn die Zahlen, Tabellen und Listen sind für viele Nutzer wenig aufschlussreich: Die freigegebenen Daten werden verdelt, um neue und hilfreiche Informationen zu liefern. Die heutige IT-Technik kann die riesigen, für Menschen unübersichtlichen Datenmengen automatisch verarbeiten und analysieren. So können Bürger, Journalisten, Politik und Wirtschaft von Open Data profitieren.

Einzelpläne

Strukturieren Einnahmen und Ausgaben grundsätzlich nach Bundesministerien. Dazu kommen besondere Einzelpläne, wie die Bundesschuld.





Visualisieren & Verknüpfen

Daten sind attraktiver und nachvollziehbarer, wenn sie als Diagramme oder Infografiken visualisiert werden. Weiter geht die Verknüpfung von Open Data zu neuen Anwendungen, dazu werden Datensätze aus verschiedenen Quellen kombiniert. So können neue nützliche Zusammenhänge auf einen Blick deutlich werden.

Es entstehen sogenannte Mash-ups (Re-Kombinationen) im Internet oder als App (als mobile Anwendung), die Open Data nutzen und abhängig von Aspekten wie Ort und Zeit visualisieren. Oft basieren die Anwendungen auf Kartenmaterial, so dass Daten geografisch verortet werden: Auf welchen Straßen ereignen sich die meisten Unfälle? Wann passieren die meisten Einbrüche in einer Stadt? Wie häufig sind Politiker auf Dienstreise? Solche Fragen lassen sich mit visualisierter und verknüpfter Open Data beantworten, ohne dass Nutzer sich durch die ursprünglichen Zahlen und Tabellen arbeiten müssen, die durch die Mash-ups automatisch ausgewertet werden.

Damit die aufgearbeiteten Daten nicht veralten, ist es am besten, wenn freigegebene Daten über sogenannte offene Schnittstellen (API = application programming interface) abgerufen werden. So beziehen sich die genutzten Daten immer auf die Originalquelle: Wenn diese aktualisiert wird, werden es die Datensätze auch.

Data Mining & Daten-Journalismus

Zur gezielten Analyse von großen, strukturierten Datenmengen wird Data Mining angewendet. Mithilfe statistischer Methoden werden interessante Zusammenhänge und neues Wissen aus den Daten generiert. Beim Data Mining wird eine ausgewählte Datenmenge automatisch nach möglichen Mustern, Regeln, Gesetzmäßigkeiten oder auffälligen Abweichungen durchsucht, die Menschen nur schwer erkennen können.

Open Data und Data Mining werden von den Medien genutzt: Eine zentrale Rolle beim Daten-Journalismus spielen Datensätze, die zunächst Recherchegrundlage und schließlich ein zentraler Bestandteil der jour-

nalistischen Geschichte sind. Die Aufgabe von Daten-Journalisten ist es, relevante Daten zu sammeln, mit technischer Hilfe auszuwerten und für das jeweilige Medium aufzubereiten.

Meist dürfte es sich um unbekannte oder negative Geschichten handeln, die Journalisten sonst nicht erfahren hätten. Denn Open Data bedeutet, dass sich Medien nicht mehr auf offizielle Studien und die Deutung durch die Quelle verlassen müssen. Mit Open Data und seriösen Recherche- sowie Statistik-Methoden können Journalisten zu eigenen Ergebnissen kommen.

Das Resultat von Daten-Journalismus sind nicht nur Infografiken, sondern auch klassische Artikel oder multimedial und interaktiv aufbereitete Websites. Als optimal definiert das Open Data-Dossier der Bundeszentrale für politische Bildung, wenn sich Nutzer selbst mit den Daten auseinandersetzen oder diese durch Parameter auswählen können, und die genutzten Datensätze im Sinne von Open Data zur Weiterverwendung angeboten werden. So kann auch eine verfälschende oder subjektive Interpretation des Materials verhindert werden. Somit sind Data Mining und Daten-Journalismus zwei sich überschneidende Methoden, um Datensätzen verborgene Informatio-

Links

- Das Projekt „Open Gov / Data Deutschland“. www.daten-deutschland.de
- „Open Data in Kommunen“ – Eine Initiative der Digitalen Gesellschaft e.V. mit Informationen, Argumenten und Kriterien zu Open Data. <http://opendata-kommunen.de>
- Website mit Links zu Quellen, Informationen und Anwendungen zu Open Data. <http://opendata-tools.org>
- „European Public Sector Information Plattform“ der EU mit Nachrichten, Meinungen und Veranstaltungen zu Open Data. www.epsiplatform.eu

nen zu entlocken und die Öffentlichkeit darüber zu informieren.

Wirtschaftliches Potenzial

Verwendet wird Open Data auch von der Wirtschaft, weil die Nutzung und Weiterverwendung auch kommerzielle Zwecke einschließt. So werden neue Geschäftsmodelle, Produkte, Dienstleistungen und Anwendungen entwickelt, die sich auf Open Data stützen. Beispiele sind Apps und Dienste, die Daten auf mobile Geräte liefern, um sie dort abzurufen, wo sie benötigt werden. Die wirtschaftliche Nutzung von Open Data sorgt für wachsende Märkte, zusätzliche Arbeitsplätze und letztendlich auch mehr gezahlte Steuern. Studien der EU-Kommission haben ergeben, dass Open Data einen wirtschaftlichen Wert von bis zu 68 Milliarden Euro in allen Mitgliedsstaaten ausmacht.

Barrieren überwinden

Damit Open Data tatsächlich zu politischer Partizipation, wirtschaftlichem Wachstum und unabhängigerem Journalismus führt, müssen noch einige Schwierigkeiten bewältigt werden. Bisher behindern unklare

oder unterschiedliche Konditionen die umfassende Nutzung. Im föderalen Deutschland und in den EU-Mitgliedsstaaten fehlt es an Gesetzen mit einheitlichen rechtlichen, technischen und finanziellen Regeln.

(Urheber-)Rechte

Wenn Open Data weiterverwendet und verknüpft werden soll, müssen alle Quellen ihre Daten unter einheitlichen oder zu vereinbarenden Nutzungsrechten und Lizenzen freigeben. Sonst wird Open Data zu einem rechtlichen Problemfeld für die Nutzer, das sie in Folge meiden könnten.

Datensätze unterliegen meistens nicht dem Urheberrecht und können im Sinne von Open Data freigegeben werden. Auch Daten, die dem Urheberrecht unterliegen, sind nicht ausgenommen, denn sie können von den Rechteinhabern durch eine offene Lizenz für Open Data zugelassen werden. Trotzdem stellen einige Quellenanbieter ihre Daten unter exklusive Lizenzen, die die Nutzung einschränken und auf den ersten Blick mehr Einnahmen als Open Data generieren.

Der Wandel von rechtlich stark geschützten Daten zu kompatiblen und offenen Lizenzen für Open Data muss noch stattfinden. Empfehlenswert ist eine (international) einheitliche Lizenz, die die Nennung der Quelle und/oder des Urhebers erfordert sowie die private oder kommerzielle Weiterverwendung erlaubt. Hierfür bieten sich die etablierten Creative Commons- und GNU-Lizenzen an, die sich entsprechend modifizieren lassen.

Format

Eine technische Barriere für Open Data sind die unterschiedlichen Formate, in denen die Datensätzen je nach Quelle vorliegen, die auf die interne IT-Technik zugeschnitten sind. Das führt dazu, dass Datensätze nicht gleichzeitig von Software verarbeitet werden können. Noch mehr Schwierigkeiten bereiten Formate, die nicht maschinenlesbar sind – beispielsweise eingescannte Dokumente, Formate für Fotos und Grafiken oder PDF-Dateien. Geeignet ist das offene Format csv, das Daten durch Kommata strukturiert und unabhängig von kommerzieller Software ist. Aktuell wird befürwortet, dass Open Data im vorhandenen Format freigegeben und schrittweise in offene Formate umgewandelt wird. Dann können Daten einfach genutzt, verknüpft und verarbeitet werden.

Auffindbarkeit

Das Auffinden bereits vorhandener Open Data bereitet Bürgern und Unternehmen noch Probleme, hat die EU-Kommission festgestellt. Freigegebene Daten sind oft im Internet verstreut und auf Websites versteckt. Es fehlt an zentralen Zugangspunkten, an denen Nutzer auf Daten zugreifen und nach ihnen suchen können.

Die Studie zum Stand von Open Data in Europa hat ergeben, dass noch keines der neun untersuchten Länder über ein nationales Open Data-Portal verfügt. Beim fünften Nationalen IT-Gipfel in Deutschland wurde vereinbart, ein gemeinsames Portal von Bund und Ländern bis 2013 aufzubauen. Auch die

Beispiele

■ **Open Data-Portale**

USA: www.data.gov

UK: www.data.gov.uk

Westfälische Wilhelms-Universität Münster.

<http://data.uni-muenster.de> (Beta-Version)

■ Internationaler Open Data-Katalog.

<http://datacatalogs.org>

■ Katalog für Open Data in Deutschland, betrieben von der Open Knowledge Foundation Deutschland e.V.

<http://offenedaten.de>

■ Open Data Projekte aus Deutschland, Europa und der Welt.

<http://opendata-showroom.org>

■ **Visualisierung**

des Bundeshaushaltes seit 2006:

<http://bund.offenerhaushalt.de>

des internationalen Flugverkehrs:

www.radarvirtuel.com

der deutschen Fernverkehrszüge mit Verspätungen:

<http://zugmonitor.sueddeutsche.de>

■ **Anwendungen zur Visualisierung**

von Daten auf Kartenmaterial:

www.datamaps.eu

www.openheatmap.com

<http://issuemap.org>

■ „Datenjournalismus: Chance für den Journalismus von morgen“ Aufsatz von Annette Leßmöllmann im Journalismik Journal 1/2012.

Kurzlink: www.grimme-institut.de/d/895687

■ Data Blog von Zeit Online.

<http://blog.zeit.de/open-data>

■ Open Source Software zur Visualisierung, entwickelt für Journalisten.

<http://datawrapperr.de>

Quellen

- Dossier der Bundeszentrale für politische Bildung.
Kurzlink: www.grimme-institut.de/d/158395
- „Unbekannte Gewässer – zum Stand von Open Data in Europa“ – Studie von 2011.
Kurzlink: www.grimme-institut.de/d/778057
- „Digital Agenda for Europe“ der Europäischen Kommission (englisch).
Kurzlink: www.grimme-institut.de/d/195294
- Koalitionsvertrag 2012 - 2017 der Regierung in NRW.
Kurzlink: www.grimme-institut.de/d/973756
- Regierungsprogramm „Vernetzte und transparente Verwaltung“ von 2010.
Kurzlink: www.grimme-institut.de/d/823178
- „Digital Agenda: Commission's Open Data Strategy, Questions & Answers“. Memo der EU.
Kurzlink: www.grimme-institut.de/d/326641
- „Open data: an international comparison of strategies“ Aufsatz von Noor Huijboom und Tijs van den Broek im European Journal of ePractice, März/April 2011 (englisch).
Kurzlink: www.grimme-institut.de/d/814809

EU plant noch 2012 einen zentralen Zugangspunkt zu Open Data aus ihren Institutionen zu schaffen, und anschließend soll ein pan-europäisches Portal folgen, das Open Data aus den Mitgliedsstaaten vereint. Hinsichtlich der Auffindbarkeit gelten die USA und Großbritannien als Vorbilder, da beide schon Open Data-Portale etabliert haben.

Auch für zentrale Zugangspunkte ist der Zugriff auf Open Data über offene Schnittstellen bedeutend, denn so bleiben die Datensätze auf dem Server der Quelle. Das heißt, die Behörden und Ämtern müssen ihre Daten nicht ganz aus der Hand geben und können sie pflegen und aktualisieren.

Kosten

Das Prinzip der Open Data umfasst den kostenlosen Zugriff – denn die Daten wurden mit öffentlichen Geldern wie Steuern und Gebühren finanziert. Dagegen strebt die Verwaltung häufig die Deckung der eigenen Kosten an, statt die gesamten wirtschaftlichen Effekte von Open Data zu berücksichtigen, kritisiert die EU. Es wird auch argumentiert, dass die Freigabe von Daten zusätzliche Kosten verursacht. So muss IT-Infrastruktur vorhanden sein und gewartet werden; die Datensätze selbst müssen den Open Data-Prinzipien angepasst werden.

Einen Kompromiss enthält der Vorschlag zur PSI-Richtlinie: Die Nutzer sollen die sogenannten Grenzkosten zahlen, aber keine Gebühren, um die Kosten für das Erstellen und Pflegen der Daten decken. Die Preise sollen so gering wie möglich und transparent sein, damit viele Nutzer auf Open Data zugreifen können.

Ausblick

Schrittweise werden in Deutschland die notwendigen Standards geschaffen, damit Open Data erfolgreich genutzt werden kann. Für Rückenwind soll die vorgeschlagene EU-Richtlinie sorgen, die einheitliche Regeln zu wichtigen Open Data-Prinzipien vorsieht. Somit können einige Barrieren verringert werden. Die treibenden

Kräfte für Open Data wirken vorwiegend von außen auf den öffentlichen Sektor. Behörden und Ämter sehen Open Data traditionell (noch) nicht als vorrangige Aufgabe an. Verwaltung in Deutschland ist eher geschlossenen Strukturen verhaftet – das Amtsgeheimnis (Datenschutz) hat eine längere Tradition als Open Data.

Tatsächlich lassen sich die Auswirkungen von Open Data nicht vollständig voraussagen: So ist es möglich, freigegebene Daten zu fälschen oder zu manipulieren, etwa durch Auswahl der Datenmenge. Im entsprechenden Dossier der Bundeszentrale für politische Bildung vermutet Christiane Schulzki-Haddouti, dass das parlamentarische System durch die Transparenz von Open Data auch in eine Glaubwürdigkeitskrise geraten könne: „Unerwartete Konsequenzen bzw. umwälzende Entwicklungen seien aber auch in anderen Bereichen zu erwarten, in denen neue Veröffentlichungsformate und Auswertungstechniken neue Darstellungen und Erkenntnisse ermöglichen.“

Impressum

Die Erstellung dieser Broschüre wurde von der Ministerin für Bundesangelegenheiten, Europa und Medien des Landes Nordrhein-Westfalen gefördert. Sie kann kostenlos unter www.grimme-institut.de/imblickpunkt heruntergeladen werden.

Grimme-Institut
Gesellschaft für Medien, Bildung und Kultur mbH
Eduard-Weitsch-Weg 25 • D-45768 Marl
Tel: +49 (0) 2365 9189-0 • Fax: +49 (0) 2365 9189-89
E-Mail: info@grimme-institut.de
Internet: www.grimme-institut.de

Text: Christina Quast
Redaktion: Annette Schneider
Gestaltung und Layout: Georg Jorczyk
Bildquellen: Spectral-Design / Fotolia.com (S. 1), Wikipedia (S. 1 u. 2), bundeshaushalt-info.de (S. 1 u. 3), sueddeutsche.de (S. 1 u. 4)

Redaktionsschluss: Oktober 2012