

Barnickel N. & Klessmann, J. (2012). Open Data - Am Beispiel von Informationen des öffentlichen Sektors. In U. Herb (Hrsg.), *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft* (S. 127-158).

Saarbrücken: universaar.

URN: [urn:nbn:de:bsz:291-universaar-873](https://nbn-resolving.org/urn:nbn:de:bsz:291-universaar-873)

URL: <http://universaar.uni-saarland.de/monographien/volltexte/2012/87>



Dieses Werk bzw. dieser Inhalt steht unter einer Creative Commons Namensnennung 3.0 Deutschland Lizenz. <http://creativecommons.org/licenses/by/3.0/de/legalcode>

Nils Barnickel, Jens Klessmann

Open Data - Am Beispiel von Informationen des öffentlichen Sektors

1 Einführung

Daten und Informationen spielen in nahezu allen gesellschaftlichen Bereichen eine wichtige Rolle. Wissensbasierte Gesellschaften sind angewiesen auf Daten, die als der wesentliche Rohstoff im Informationszeitalter gesehen werden können. Über Faktoren wie Verfügbarkeit, Korrektheit, Aktualität, Preis dieser Daten haben Akteure wie Ersteller, Wiederverwerter oder Endnutzer von Daten häufig differierende Ansichten. Alle Gruppen haben jedoch ein hohes Interesse an Daten.

Werden Daten als eine Art Antriebsmittel unserer Gesellschaft verstanden, so kann die Bereitstellung oder das Fehlen von Daten bedeutende Auswirkungen auf Abläufe und Entscheidungen einer Gesellschaft und die Leistungsfähigkeit von Wirtschaft und öffentlichem Sektor haben.

Öffentliche Verwaltungen erzeugen und verarbeiten täglich große Mengen von Daten und Informationen. Einen Teil davon stellen sie für Dritte häufig in Form von Dokumenten wie Bescheiden, Verordnungen, Gesetzen oder Strategiepapieren zur Verfügung. Viele der in diesen Dokumenten enthaltenen Informationen sind bereits aggregiert, die zugrundeliegenden Rohdaten nur organisationsintern verfügbar.

Dies hat verschiedene Gründe, wie etwa, dass es in vielen Einrichtungen des öffentlichen Sektors als eine wesentliche Aufgabe aufgefasst wird, Informationen in entsprechend verdichteter Form zu erstellen und zu kommunizieren. In historischer Perspektive war es bis vor kurzem gar nicht oder nur mit hohem technischen und organisatorischen Aufwand möglich, einer Vielzahl von Personen die zugrundeliegenden Rohdaten von Verwaltungs- und Regierungsentscheidungen zur Verfügung zu stellen. Nicht zuletzt sind aktuelle Informationen und deren Besitz die Grundlage für Herrschaftswissen, mit dem die Steuerung einer Gesellschaft vermeintlich leichter möglich wird.

Technologische und ethisch-kulturelle Entwicklungen moderner Gesellschaften bewirken jedoch, wie aktuell in verschiedenen Kontexten, von z.B. Großbauvorhaben bis zu Demokratiebestrebungen, zu beobachten ist, eine Entwicklung hin zu einem Verlangen nach mehr Mitbestimmung und Transparenz staatlicher Prozesse und Entscheidungen. Die Bereitstellung von Verwaltungsdaten zur Nutzung durch Gruppen außerhalb des öffentlichen Sektors kann hier ein wesentlicher Baustein sein. Semantische Technologien bieten die Möglichkeit, diese Daten erstmalig in größerem Umfang und teilweise automatisiert weiterverarbeiten zu können. Dies erlaubt es vergleichsweise kleinen Organisationen wie NGOs oder mittelständischen Unternehmen leichter mit großen Datenmengen umzugehen und für ihre eigenen Ziele einzusetzen.

Auch die Repräsentanten öffentlicher Einrichtungen versprechen sich vom Open-Data-Ansatz zunehmend positive Effekte für ihre Organisationen. Zum einen dürfte eine leichtere Verfügbarkeit der eigenen Daten die Effizienz der eigenen Abläufe erhöhen. Zum anderen, so die Hoffnung, wird die Input- und Output-Legitimität des politisch-administrativen Systems verbessert. Transparenz und Mitbestimmung sollen besseren Input für politische Prozesse bewirken, was wiederum zu einem verbesserten Output in Form politischer Entscheidungen führt.

Im folgenden Text wird zu Beginn das Verständnis von Open Data dargestellt, um dann seine Relevanz im öffentlichen Sektor vorzustellen. Der aktuelle Stand in Deutschland wird umrissen und einige internationale Beispiele für erfolgreiche Open-Data-Portale werden benannt. Im darauf folgenden Abschnitt wird auf die Unterschiede von Open Data und „Linked Data“ eingegangen, um anschließend die Struktur einer Open-Data-Softwarearchitektur für Organisationen bzw. Gebietskörperschaften des öffentlichen Sektors zu skizzieren. Abschließend erfolgt ein Ausblick auf noch offene Herausforderungen.

2 Verständnis und Potentiale von Open Data und Open Government Data

Der öffentliche Sektor ist ein stark informations- und wissensbasierter Sektor. Prozesse und Entscheidungen von Verwaltung und Politik werden fast immer auf Grundlage von Informationen durchgeführt.

Zwei Beispiele hierfür sind:

- 1) Die Entwicklung einer Strategie zur Förderung der regionalen Wirtschaft erfordert die Zusammenführung und Auswertung ganz unterschiedlicher Datensätze.

- 2) Für die Durchführung eines Planfeststellungsverfahrens in der Stadt- und Regionalplanung werden Informationen über die betreffende bauliche Veränderung, ihre Umwelt, die Belange der Anwohner, etc. benötigt.

Öffentliche Verwaltungen handhaben täglich viele verschiedene Daten. Ein Großteil dieser Daten steht verwaltungsintern zur Verfügung. Dritte, außerhalb der jeweiligen Behörden, haben keinen oder nur eingeschränkten Zugriff darauf. In Fällen wie der Entwicklung einer Wirtschaftsförderungsstrategie werden ausgewählte Dokumente zugänglich gemacht, etwa das finale Dokument der Strategie. Dem Dokument zugrundeliegende, feingranulare Daten bleiben jedoch häufig allein in der Hand der Behörden bzw. der beauftragten Ersteller. In Vorgängen wie einem Planfeststellungsverfahren werden bereits vergleichsweise viele (Zwischen-) Informationen für die Öffentlichkeit zugänglich gemacht. Bei der Umsetzung eines Open-Government-Data-Ansatzes geht es auch genau darum: Wie viele Informationen werden in welcher Form und wem zur Verfügung gestellt? Im Folgenden wird dargestellt, was unter offenen Daten und speziell offenen Daten der Verwaltung zu verstehen ist und worin die Potentiale dieses Ansatzes gesehen werden.

2.1 Verständnis von Open Data und Open Government Data

Laut der Definition der Open Knowledge Foundation ist freies Wissen als ein Gegenstand oder Werk zu verstehen, mit dem Wissen transferiert wird und das verschiedene Kriterien erfüllt. Das Werk sollte u.a. im Ganzen zugänglich sein, einer diskriminierungsfreien Lizenz unterliegen, die eine Weiterverteilung und Wiederverwendung erlaubt, ohne dabei technischen Restriktionen zu unterliegen. Die Verpflichtung bei der Wiederverwendung die Urheber zu nennen und ein verändertes Werk als solches zu kennzeichnen kann Bestandteil dieser Lizenz sein. Weiter muss die Lizenz des Werkes mit diesem weiter verteilt werden, dabei Gültigkeit behalten und gleichzeitig nicht die Weiterverteilung anderer Werke behindern (Open Knowledge Foundation, 2010). Im Open-Data-Ansatz werden diese Punkte für Wissensbestände aus allen Bereichen aufgegriffen. Open Government Data ist einer dieser Bereiche, mit teilweise besonderen Anforderungen.

Bei der Bereitstellung von Open Government Data geht es um Daten der öffentlichen Verwaltung, welche Dritten zur Weiterverwendung zur Verfügung gestellt werden. Ob die bereitgestellten Daten als offen bezeichnet werden können, ist abhängig von verschiedenen Faktoren wie der Zugänglichkeit, den

Formaten und den rechtlichen Bedingungen unter denen die Daten genutzt werden dürfen (von Lucke & Geiger, 2010, S. 6). Obwohl bislang keine einheitliche Definition des Begriffs „Open Government Data“ existiert, besteht weitgehend Einigkeit über die Grundsätze, welche Einfluss auf den Grad der Offenheit haben. Hierzu werden die Prinzipien des Open Data -Ansatzes auf Daten des öffentlichen Sektors angewendet (siehe Abschnitt zu Open Data).

In einer deutschsprachigen Übersetzung des Begriffs Open Government Data wird dieser Begriff als „offene Verwaltungsdaten“ bezeichnet und wie folgt definiert: „Offene Verwaltungsdaten sind jene Datenbestände des öffentlichen Sektors, die von Staat und Verwaltung im Interesse der Allgemeinheit ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden“ (von Lucke & Geiger, 2010, S. 6).

Ob Daten, speziell Verwaltungsdaten, tatsächlich offen sind lässt sich also an der Erfüllung verschiedener Kriterien festmachen (Sebastopol Group, 2007; Sunlight Foundation, 2010):

- **Vollständigkeit:** Datensätze öffentlicher Einrichtungen sollten so vollständig wie möglich veröffentlicht werden, soweit dies die Regelungen zum Datenschutz zulassen. Die Daten sollen dabei möglichst im Rohformat inklusive beschreibender Metadaten veröffentlicht werden.
- **Primärquellen:** Daten, welche als offene Verwaltungsdaten bereitgestellt werden, sollten direkt aus den ursprünglichen Quellen veröffentlicht und mit Informationen zum Ablauf der Sammlung und Erstellung der Daten angereichert werden. Dies erlaubt es Dritten nachzuvollziehen, wie die Datenerstellung bewerkstelligt wurde.
- **Zeitliche Nähe:** Geschieht die Veröffentlichung der Daten möglichst zeitnah zur Entstehung oder Aktualisierung des Datensatzes? Werden Daten erst mit großem Abstand zu ihrer Kreierung zur Nutzung durch externe Personen bereitgestellt, so können sie unter Umständen nur noch eingeschränkt Verwendung finden, da in der Zwischenzeit bspw. bereits aktualisierte Daten zur internen Verwendung bereitstehen. Welcher Zeitraum als zeitnah zu verstehen ist, ist dabei abhängig vom Einzelfall, statistische Daten sind hier anders einzuordnen als Daten zur Auslastung von Verkehrssystemen. Die zeitnahe Bereitstellung kann große Aufwände bei den datenerhebenden und –bereitstellenden Einrichtungen erzeugen.
- **Leichter Zugang:** Welcher Aufwand ist notwendig, um Zugang zu einem oder mehreren Datensätzen zu erhalten? Hindernisse hierbei reichen von

technischen Problemen, wie unterschiedlichen Formaten oder fehlenden Metadaten über organisatorische Barrieren (z.B. Pflichtregistrierung) hin zu rechtlichen Vorgaben.

- **Maschinenlesbarkeit:** Erst durch die Möglichkeit zur automatisierten Verarbeitung lassen sich viele Potentiale offener Daten ausschöpfen. Maschinenlesbarkeit eines Datensatzes erlaubt die Einbindung in Softwareanwendungen. Große Mengen an Daten lassen sich dementsprechend in neuen Applikationen nutzbar machen. Dies setzt u.a. eine solide Auszeichnung mit Metadaten, die Verwendung einheitlicher Formate und Terminologien voraus.
- **Diskriminierungsfreiheit:** Ein Zugriff auf Daten ohne Ansehen der Person, zeitlichen Restriktionen, der Anforderung die eigene Identität nachzuweisen und eine Begründung für den Zugriff vorzulegen. Beispiele für eine Diskriminierung beim Zugriff auf Daten wären eine notwendige Registrierung, Mitgliedschaft oder Einschränkung auf ausgewählte Softwareanwendungen.
- **Verwendung offener Standards:** Mit dem Einsatz offener Standards wird die Verwendung von Formaten angestrebt, welche nicht nur von ausgewählten Programmen gelesen und verarbeitet werden können. Offene Standards gewährleisten die Freiheit mit verschiedenen Programmen auf die Daten zugreifen zu können, ohne dass dafür Lizenzkosten an einzelne Hersteller abgeführt werden müssen.
- **Lizenzierung:** Eine restriktive Lizenzierung stellt ein Hindernis für die Nutzung von Daten dar. Offene Verwaltungsdaten sollten eindeutig sichtbar mit einer Lizenz versehen und gemeinfrei nutzbar sein.
- **Dauerhaftigkeit:** Daten, die im Internet veröffentlicht werden, können einfach verändert werden, ohne dass dies für Dritte transparent wird. Offene Verwaltungsdaten sollten permanent verfügbar sein und Änderungen, Aktualisierungen und Löschungen durch Versionskontrolle und Archivierung nachvollziehbar gestaltet werden.
- **Nutzungskosten:** Die Bereitstellung von Verwaltungsdaten zu Kosten behindert die Weiterverwendung dieser Daten. Ein Großteil der Daten der öffentlichen Hand wird unabhängig von späteren Nutzungsgebühren gesammelt und erstellt. Die Erhebung von Gebühren beschränkt generell die Gruppe der Nutzer, darunter auch den Einsatz zu wirtschaftlichen Zwecken, was wiederum einen Effekt auf das Steuereinkommen haben kann.

2.2 Open Government Data, Public Sector Information und Informationsfreiheit

Neben dem noch relativ neuen Ansatz der offenen Verwaltungsdaten existieren, teilweise schon seit Jahrzehnten, Aktivitäten für mehr Transparenz und leichtere Weiterverwendung von Verwaltungsdaten. Diese werden kurz vorgestellt, um die Ähnlichkeiten und Unterschiede der verschiedenen Begriffe und Konzepte zu umreißen (siehe hierzu auch Tabelle 1).

Bereits deutlich älter als der Open-Data-Ansatz sind die Diskussionen um die Weiterverwendung von Informationen des öffentlichen Sektors sowie die Veröffentlichung administrativer Unterlagen im Zusammenhang mit Anfragen auf Grundlage von Informationsfreiheitsgesetzen. Zu beiden Bereichen existieren in vielen Ländern und auch auf EU-Ebene gesetzliche Regelungen.

Beim Ansatz der Informationsfreiheit wird davon ausgegangen, dass Bürger eines Landes einen Anspruch auf Informationen des öffentlichen Sektors haben. Mit der regulativen Umsetzung dieses Anspruchs in Form rechtlicher Normen zur Informationsfreiheit wird Bürgern die Möglichkeit eingeräumt, Anfragen an öffentliche Einrichtungen zu stellen, in denen sie um die Bereitstellung spezifischer Informationen bitten (Kubicek, 2008, S. 9). Solange die gewünschten Informationen nicht unter verschiedene Ausnahmetatbestände fallen, ist die jeweilige Einrichtung gezwungen, die entsprechenden Unterlagen zur Verfügung zu stellen.

Mit der Weiterverwendung von Informationen öffentlicher Einrichtungen werden bislang vor allem Ziele der wirtschaftlichen Entwicklung verfolgt, dies geschieht in der Annahme, ausgewählte Verwaltungsdaten besitzen einen Wert, der in der primären Verwendung, aber vor allem in der erneuten Verwendung durch Dritte zum Tragen kommt. Verschiedene Untersuchungen legen diese volkswirtschaftliche Bedeutung nahe (Dekkers, Polman, te Velde & de Vries, 2006; Fornefeld, Boele-Keimer, Recher & Fanning, 2009; Pollock, 2009). So werden in den Untersuchungen durch die Weiterverwendung teilweise geschätzte Effekte in Höhe von bis zu zweistelligen Milliardenbeträgen angeführt.

	Open Govern- ment Data (Offene Verwal- tungsdaten)	Public Sector Information	Informations- freiheit
Ziele (in absteigender Bedeutung)	Transparenz, zivilgesell- schaftliches Engagement, Wirtschafts- förderung, Effizienz	Wirtschafts- förderung, Transparenz	Transparenz
Bereitstellung	Proaktiv	Proaktiv	Auf Nachfrage
Kosten für die Nutzer ¹	Nach Möglich- keit ohne Kosten	Marginalkosten – Profitmaximie- rung	Üblicherweise Marginalkosten
Umfang der Datensätze	Große Mengen	Große Mengen	Eher kleine bis mittlere Mengen
Digital/analog	Digital	Digital	beides
Weiterverwen- dung durch Dritte	Ja	Ja	Derzeit häufig nicht vorgesehen
Art der Daten	Überwiegend quantitativ	Überwiegend quantitativ	Überwiegend qualitativ

Tabelle 1: Abgrenzung der Begriffe Open Government Data (Offene Verwaltungsdaten), Public Sector Information, Informationsfreiheit anhand verschiedener Kategorien. Eigene Darstellung

2.3 Potentiale für den öffentlichen Sektor

Open Data bietet gerade im und um den öffentlichen Sektor enorme Potentiale. Durch die Bereitstellung von Daten kann eine größere Transparenz für Außenstehende über die Abläufe und Entscheidungen in Einrichtungen der öffentlichen Verwaltung erreicht werden. In Ergänzung finanzieller Subventionen

¹ Vergleiche Newbery PSI Provision 2008, S. 11 f.

für die eigene Wirtschaft kann der öffentliche Sektor Unterstützung in Form relevanter, aktueller Daten bieten. Diese können dazu beitragen, vorhandene Geschäftsmodelle zu verfeinern oder neue zu entwickeln. Eine strukturierte Veröffentlichung von Daten und Informationen kann auch die Arbeit von Einrichtungen des öffentlichen Sektors positiv befördern. Die Möglichkeit nach Arbeitsergebnissen anderer Organisationseinheiten zu suchen und darauf zuzugreifen, dies eventuell sogar organisations- und ebenenübergreifend, kann helfen Doppelarbeiten zu vermeiden und die Ergebnisse der eigenen Arbeiten zu verbessern.

Im Folgenden werden die drei genannten Dimensionen, Transparenz, Wirtschaftsförderung und Produktivitätssteigerung kurz im Einzelnen betrachtet.

2.3.1. Transparenz

Transparenz wird meist als die Möglichkeit verstanden, für Dritte außerhalb von Einrichtungen Einsicht in Prozesse und Entscheidungen zu nehmen, um sich so selbst eine Meinung dazu bilden zu können (Florini, 2007, S. 5; Piotrowski & Van Ryzin, 2007, S. 308). Diese recht breit angelegte Definition von Transparenz lässt sich auch für das Verständnis von Transparenz im öffentlichen Sektor übertragen.

Mit der Bereitstellung von Daten und Informationen durch öffentliche Einrichtungen erhalten Außenstehende die Möglichkeit, ein genaueres Bild über bestimmte Vorgänge in Behörden und Gebietskörperschaften zu bekommen. Werden diese Daten so zur Verfügung gestellt, dass sie einfach weiterverarbeitet werden können, vereinfacht dies die Aufbereitung der Daten für unterschiedliche Fragestellungen, die Organisationen oder interessierte Bürger haben.

Verschiedene Untersuchungen deuten auf einen gewissen positiven Zusammenhang der Durchsichtigkeit der Handlungen und Ergebnisse öffentlicher Organisationen und der allgemeinen Qualität politisch-administrativer Steuerung hin (Relly & Sabharwal, 2009, S. 149). Insofern kann davon ausgegangen werden, dass die proaktive Veröffentlichung staatlicher Daten auch einen positiven Einfluss auf die Handlungen der Akteure in Politik und Verwaltung haben dürfte.

Ebenso kann die größere Informiertheit zivilgesellschaftlicher Akteure dazu beitragen, dass diese sich stärker in politisch-administrative Vorgänge einbringen, ein aus demokratie-theoretischer Sicht sicher erwünschter Effekt (Schmidt, 2008, S. 236). Gleichzeitig ist durch die bloße Bereitstellung von Daten und Informationen nicht automatisch eine enorme allgemeine Steige-

zung der Bürgerpartizipation zu erwarten. Dies lässt sich eindeutig aus diversen Untersuchungen zu teilweise aufwendig organisierten, partizipativen Angeboten in der öffentlichen Verwaltung ableiten (Holtkamp, 2006, S. 201; Kubicsek, 2007, S. 4 f.). Unter Informationsbereitstellung als solcher ist auch noch keine Bürgerbeteiligung zu verstehen, offene Verwaltungsdaten können die Grundlage für eine aktive Teilhabe am politischen Geschehen bilden.

2.3.2. Wirtschaftsförderung

Eine übliche Methode zur Unterstützung von Wirtschaftsunternehmen ist die Bereitstellung finanzieller Mittel in Form von Subventionen. Neben diversen Einschränkungen dieser Werkzeuge bedeutet die direkte finanzielle Unterstützung eine enorme Belastung der öffentlichen Haushalte in Zeiten knapper Mittel. Als eine ergänzende Möglichkeit zur Beförderung des Wirtschaftswachstums kann die Bereitstellung von Verwaltungsdaten verstanden werden.

Mit Hilfe offener Daten der Verwaltung können Unternehmen bestehende Geschäftsmodelle verfeinern oder neue entwickeln. Für eine Firma, im Bereich der Immobilienvermittlung etwa, können präzise und aktuelle Daten zur öffentlichen Sozialinfrastruktur von Interesse sein, da diese Informationen eine weitere Verfeinerung der Einschätzung des Werts einer Immobilie erlauben.

Untersuchungen deuten darauf hin, dass die Bereitstellung offener Verwaltungsdaten zu Grenzkosten für eine Volkswirtschaft insgesamt vorteilhaft ist (Newbery, Bently & Pollock, 2008, S. 2). Im Falle einer digitalen Bereitstellung der Daten tendieren die Grenzkosten gegen Null. Als Argumente für eine beinahe kostenlose bis kostenfreie Bereitstellung von Verwaltungsdaten wird u.a. angeführt: Um die Durchschnittskosten der Datenerstellung tatsächlich zu decken, wäre ein deutlicher Preisaufschlag notwendig, der zu einer geringeren Verwendung der Daten und damit einer niedrigeren Wirkung auf die Gesamtwirtschaft führt (Newbery u. a., 2008, S. 123 f.). Zumindest sollten Rohdaten nicht preismaximierend, sondern höchstens kostendeckend angeboten werden (Office of Fair Trading, 2006, S. 134).

Am Beispiel des Bereichs der Geodaten lässt sich bereits heute ablesen, welchen Effekt eine Bereitstellung von Daten zu Grenzkostenpreisen haben kann. Das Österreichische Bundesamt für Eich- und Vermessungswesen hat im Jahr 2006 von einer Durchschnittskostenbepreisung auf die Bereitstellung zu Grenzkostenpreisen umgestellt. Dadurch sanken die Preise für ausgewählte Datentypen um bis zu 97%. Gleichzeitig wuchs der Absatz dieser Daten um teilweise bis zu 7000%. Im Ergebnis konnte das BEV so die Weiterverwendung seiner Daten bei gleichbleibendem Umsatz steigern. Wesentlichen Anteil an

der gestiegenen Nachfrage hatten kleine und mittlere Unternehmen (Fornefeld u. a., 2009, S. 30).

2.3.3. Größere Produktivität in der Verwaltung

Wesentliche Teile der Aufgaben der öffentlichen Verwaltung lassen sich als wissensintensive Dienstleistungen beschreiben. Die Erarbeitung strategischer Dokumente zur wirtschaftlichen Entwicklung einer Metropolregion oder die Forschungsförderung eines Bundeslandes, aber auch die Koordinierung verschiedener Aktivitäten auf sogenannte Sozialräume erfordert die Zusammenführung, Aufbereitung und Interpretation von Daten und Informationen aus unterschiedlichen Quellen. Wie generell bei großen Organisationen gilt auch hier die Annahme, dass die einzelnen Einheiten der Organisation meist nicht ausreichend informiert sind, welche Daten bereits von anderen Abteilungen erhoben wurden und somit der Organisation prinzipiell zur Verfügung stehen. Bereits eine Übersicht der in einer Organisation und ihren Teilen verfügbaren Daten und Informationen kann insofern helfen, die Arbeit der Mitarbeiter der Einrichtung zu erleichtern und zu besseren Arbeitsergebnissen führen (Glock & Broens, 2008, S. 274). Ein Portal mit offenen Verwaltungsdaten kann somit ein Baustein eines organisationsweiten Informationsmanagements sein. Eine Auszeichnung der in einem Datenportal verzeichneten oder bereitgehaltenen Daten mit Metainformationen unterstützt das zielgerichtete Auffinden und Verwenden der eigentlichen Daten. Bei Rückfragen können die Datenbereitsteller einfach identifiziert und kontaktiert werden.

Mit den richtigen Mechanismen bietet sich auch die Möglichkeit, über ein Datenportal die Qualität der Informationen zu verbessern. Eingebaute Feedbackmöglichkeiten erlauben den Datennutzern Rückmeldungen zu Problemen mit einzelnen Datensätzen zu geben oder sogar selbst die Fehler in den bereitgestellten Informationen zu beheben oder die Daten zu ergänzen.

3 Situation offener Verwaltungsdaten in Deutschland und im Ausland

Von der Entwicklung des Konzepts und den damit verbundenen Prinzipien offener Verwaltungsdaten bis zur praktischen Umsetzung in öffentlichen Einrichtungen und Gebietskörperschaften unterschiedlicher Länder ist es ein weiter Weg. Im Ausland wurde der Ansatz teilweise wesentlich zügiger aufgegriffen

und entsprechend ist die Realisierung weiter fortgeschritten. Neben dem zeitlichen Faktor lassen sich sicher auch kulturelle und verwaltungsorganisatorische Gründe für ein schnelleres Aufgreifen des Ansatzes benennen. Transparenz hat in einigen Nationen eine lange Tradition, auf entsprechend weniger Barrieren stößt hier die Umsetzung von Transparenz im öffentlichen Sektor mithilfe von Technologien des 21. Jahrhunderts. Daneben spielt sicher auch die Ausgestaltung des staatlichen Aufbaus eine Rolle. In stärker zentralistisch orientierten Ländern, wie etwa Großbritannien, können neue transformative Konzepte im öffentlichen Sektor schneller implementiert werden als in eher föderalistisch ausgeprägten Nationen.

Im Folgenden wird übersichtsartig die gegenwärtige Situation in Deutschland im Hinblick auf die Bereitstellung von Daten durch die öffentliche Verwaltung dargestellt und beispielhaft entsprechende internationale Ansätze skizziert. Dabei erfolgt eine Konzentration auf Aktivitäten, welche bewusst dem Ansatz offener Verwaltungsdaten folgen bzw. diesem zugeordnet werden können. Die bereits vielfach existierenden Initiativen und Normen zur Weiterverwendung öffentlicher Daten und Regelung von Informationsfreiheit werden hier nicht dargestellt.

3.1 Situation in Deutschland

In Deutschland setzen sich neben zivilgesellschaftlichen Organisationen wie dem Open Data Network² oder dem Government 2.0 Netzwerk³, erste Personen und Organisationseinheiten in Politik und Verwaltung mit der Bereitstellung von Open Government Data auseinander. Open Data hat als Pilotvorhaben Eingang in die Strategie des Bundes zur Verwaltungsmodernisierung gefunden. Im Wahlkampf 2010 zur aktuellen Landesregierung NRW wurden Open-Data-Positionen geäußert und fanden Eingang in das Regierungsprogramm. Auf kommunaler Ebene entwickeln einige Großstädte erste Aktivitäten.

3.1.1 Strategie des Bundes

Verschiedene Bundesbehörden und ihre nachgeordneten Einrichtungen stellen schon jetzt Verwaltungsdaten online bereit. Zum Beispiel sind vom Bundesamt für Statistik⁴ bereits heute vielfältige Informationen zu ökonomischen,

2 <http://opendata-network.org/>

3 <http://www.gov20.de/>

4 <http://www.destatis.de/>

gesellschaftlichen und weiteren Entwicklungen zum Download erhältlich. Im Umweltbereich bietet das PortalU⁵ des Bundesumweltamtes Zugang zu umfangreichen Datenbanken mit Umweltinformationen. Das Portal ist Ergebnis einer Bund-Länder-Kooperation und wurde mit dem Ziel eines einheitlichen Einstiegspunkts für Umweltdaten entwickelt. Nutzer können über die Suchmaschine in den angeschlossenen Datenbeständen mit Unterstützung eines Fachwörterbuches, dem Semantik-Network-Service, recherchieren (Schenk, Kruse & Klenke, 2010).

Ein über die Umweltdaten hinausgehender einheitlicher Zugriff auf offene Verwaltungsdaten über einen gemeinsamen Einstiegspunkt, eine durchgängig abgestimmte Auszeichnung mit Metadaten sowie die Bereitstellung als Linked Data steht jedoch vielfach noch aus. Die Entwicklung einer organisationsübergreifenden Strategie für offene Verwaltungsdaten ist in diesem Zusammenhang mittel- bis langfristig sicher unerlässlich.

Im Herbst 2010 hat die Bundesregierung ihr aktuelles Programm zur Verwaltungsmodernisierung „Vernetzte und transparente Verwaltung“ veröffentlicht. Als eines der wesentlichen Ziele, die bei der Reform der öffentlichen Verwaltung berücksichtigt werden sollen, wird hier die stärkere Kooperation mit Akteuren aus Zivilgesellschaft und Wirtschaft sowie eine generell höhere Transparenz der eigenen Handlungen und Entscheidungen festgestellt.

Im Regierungsprogramm werden 19 verschiedene Modernisierungsprojekte mit kurzen Steckbriefen aufgeführt. Im Projekt „Open Government“ soll - „wo es rechtlich möglich und zweckmäßig ist – ein freier Zugang zu Informationen der Verwaltung geschaffen werden“ (Bundesministerium des Innern, 2010, S. 57 f.). Ziel ist es, Pilotvorhaben für die Bereitstellung von Verwaltungsdaten zu realisieren und bis zum Jahr 2013 eine mit den Ländern abgestimmte Open-Government-Strategie entwickelt zu haben, um dann an die Umsetzung zu gehen.

3.1.2 Aktivitäten auf regionaler und lokaler Ebene

Auf regionaler Ebene in Deutschland findet sich zum Beispiel in Nordrhein-Westfalen die Ankündigung, sich dem Thema Open Government und damit auch offenen Verwaltungsdaten widmen zu wollen. In der Koalitionsvereinbarung der aktuellen Regierungsparteien in NRW, wird eine stärkere Öffnung von Verwaltung angekündigt. Mit der beschriebenen Initiative soll eine „kosten- und diskriminierungsfreie Bereitstellung öffentlicher Daten und Informa-

5 <http://www.portalu.de/>

tionen“ erfolgen. Weiterhin sollen regelmäßig Wettbewerbe durchgeführt werden, bei denen „Entwickler unter Verwendung öffentlich bereitgestellter Daten Applikation erstellen, die zu Transparenz, Partizipation und Zusammenarbeit beitragen“ (NRW SPD, 2010, S. 81).

Auf lokaler Ebene sind in Deutschland bislang vor allem Aktivitäten in einigen Großstädten feststellbar. Sowohl Berlin als auch München haben jeweils einen Anwendungs- und Ideenwettbewerb durchgeführt^{6,7}. Dabei ging es in beiden Fällen auch darum, Anwendungen oder Vorschläge zu entwickeln, wie Daten des öffentlichen Sektors kreativ zur Lösung städtischer Probleme, Verbesserung existierender Geschäftsmodelle oder mehr Transparenz in Entscheidungen und Prozessen der öffentlichen Verwaltung eingesetzt werden können. Der Berliner Senat ist darüber hinaus mit dem Aufbau von Strukturen für die Online-Bereitstellung seiner Daten beschäftigt. Gemeinsam mit den Städten Amsterdam, Barcelona, Helsinki und Paris ist Berlin am EU-Projekt „Open Innovation Mechanisms in Smart Cities“ (Open Cities) beteiligt. Eines der wesentlichen Ziele dieses Projektes ist die Entwicklung und Umsetzung einer prototypischen Open-Data-Plattform. Im Anschluss an die Implementierung der Plattform in den beteiligten Städten werden Softwareentwicklungswettbewerbe durchgeführt, um die Bekanntheit und Nutzung der offenen Verwaltungsdaten zu erhöhen (Almirall, 2010).

Im Gesetz des Bundeslands Bremen zur Informationsfreiheit sind Regelungen zur pro-aktiven Bereitstellung einzelner Informationen des öffentlichen Sektors und die Erstellung eines entsprechenden Informationsregisters festgelegt. Hierin unterscheidet sich das bremische IFG von vergleichbaren Normen in anderen Bundesländern und auf Bundesebene. Um eine strategische Herangehensweise an offene Verwaltungsdaten zu entwickeln, wurde von lokalen Akteuren aus Verwaltung, Politik und Wissenschaft eine „Bremer Empfehlung zu Open Government Data“ verfasst und veröffentlicht (Linnert, Kubicek & Sommer, 2011).

In dieser Empfehlung werden zum einen ausgewählte Open-Data-Prinzipien für die lokalen Gegebenheiten interpretiert und zum anderen eigene Anliegen formuliert. In dem Dokument wird die Notwendigkeit einer an den Bedürfnissen verschiedener Nutzergruppen orientierten Erschließung als wichtig angesehen. Technische Standardisierung wird verstanden als wesentliche Voraussetzung für die großflächige Umsetzung offener Verwaltungsdaten, diese soll kombiniert werden mit Arbeiten, die ohnehin in der Verwaltung notwendig

6 Apps4Berlin: <http://www.berlin.de/projektzukunft/wettbewerbe/apps-contest/>

7 Mogdy Programmierwettbewerb: <http://www.muenchen.de/Rathaus/dir/limux/mogdy/Programmierwettbewerb/463049/index.html>

werden, um etwa den Einsatz von Dokumentenmanagementsystemen sinnvoll zu gestalten. Ebenso behandelt werden die Themen Kosten und Maschinenlesbarkeit. Für ersteres wird eine differenzierte Betrachtung empfohlen, abhängig von den späteren Nutzergruppen. Letzteres wird als grundsätzlich richtig erachtet, jedoch sollte die Umwandlung in maschinenlesbare Formate möglichst keine Mehrkosten auf Seiten der Verwaltung erfordern.

Die Bremer Empfehlung enthält mehrere neue Ergänzungen zu den existierenden Open-Data-Prinzipien. Es wird die Schaffung organisatorischer Verfahren und Strukturen in der Verwaltung gefordert, um offene Zugänglichkeit von Verwaltungsdaten nachhaltig zu ermöglichen. Zur Realisierung einer föderalen Plattform für offene Verwaltungsdaten wird ein verteiltes Redaktionsmodell empfohlen, bei dem von Beginn an auf Qualität der Inhalte und Beschreibungen geachtet werden sollte. Abschließend wird der Aufbau einer kooperativen Dateninfrastruktur von den Urhebern der Empfehlung als notwendig erachtet.

3.2 Internationale Beispiele für die Bereitstellung offener Verwaltungsdaten

Datenportale, über die der Zugang zu offenen Verwaltungsdaten realisiert wird, funktionieren wie Kristallisationspunkte für die verschiedenen Bemühungen eine stärkere Öffnung staatlicher Datenbestände zu bewirken. Daher werden hier Beispiele für Portalinitiativen auf nationaler und lokaler Ebene, sowie für entsprechende Fachportale kurz vorgestellt.

Das Datenportal der britischen Regierung data.gov.uk wurde im Frühjahr 2010 für die Öffentlichkeit freigeschaltet. Vorausgegangen war ein Forschungsprojekt und eine Planungsphase. In dem Forschungsprojekt AKTive PSI wurde die allgemeine Tauglichkeit von Semantic Web-Ansätzen für die Verwendung im öffentlichen Sektor geprüft (Shadbolt, O'Hara, Salvadores & Alani, 2011, S. 866). Aufbauend auf den Erkenntnissen wurde die Idee des einheitlichen Einstiegspunkts für offene Verwaltungsdaten in Großbritannien weiter verfolgt.

Heute werden über das Portal circa 5700 Datensätze bereitgestellt. Dem Ansatz folgend, erst Rohdaten online zur Verfügung zu stellen und sich anschließend um die Aufbereitung in das bestmögliche Format zu kümmern, stehen viele dieser Daten in Formaten wie XML, CSV, XLS, PDF oder DOC zur Verfügung, wobei einige Formate besser für die Weiterverarbeitung geeignet sind als andere, wie zum Beispiel XML und CSV. Der geringere Prozentsatz der Datensätze ist in Form von aufbereiteten Linked Data verfügbar.

Das Portal bietet semantische und nicht-semantische Funktionen (Shadbolt u. a., 2011, S. 876):

- Semantic Web Funktionen
- SPARQL-Endpunkte
- Datensätze im Linked-Data-Format
- Richtlinien zur Erzeugung und Bereitstellung von Linked Data

Nicht-semantische Funktionen

- Diskussionsforum und ein Wiki zur Unterstützung der Zusammenarbeit der Datennutzer und -bereitsteller
- Funktionen zu Suche und Browsen von Datensätzen. Verzeichnis von Ideen und Anwendungen, die auf offenen Verwaltungsdaten aus data.gov.uk aufbauen

Beispiele für lokale Aktivitäten gibt es mittlerweile zahlreich. So wurde etwa das Datenportal „DataSF“⁸ der kalifornischen Stadt San Francisco im Jahr 2009 gestartet. Es umfasst Anfang 2011 etwa 200 Datensätze in unterschiedlichen Formaten. Vorhandene Datensätze und Anwendungen, welche auf den Verwaltungsdaten aufbauen, können von Nutzern in ihrer Qualität mit einem einfachen Sternesystem bewertet werden.

Ein Beispiel für ein Portal für offene Verwaltungsdaten einer Fachbehörde bietet die Weltbank. Diese Einrichtung stellt seit Mitte des Jahres 2010 große Mengen an Daten über ein spezielles Portal zur Verfügung. Unter data.worldbank.org macht die Weltbank ihre statistischen Indikatoren verfügbar, mit denen die Mitarbeiter der Einrichtung arbeiten um die Entwicklung einzelner Länder oder ganzer Regionen zu beurteilen und miteinander zu vergleichen. Gegenwärtig sind über die Programmierschnittstelle der Weltbank ungefähr 2000 Datensätze in Formaten wie XML, CSV, KML als RSS oder JSON verfügbar.

4 Umsetzung von Open Government Data mit Open Linked Data Ansätzen

Wie die vorherigen Abschnitte zeigen, bedarf es politischer und organisatorischer Maßnahmen, um Open Government Data-Initiativen in der Praxis umzusetzen. Auch auf technologischer Ebene gilt es einige Herausforderungen zu

8 <http://datasf.org/>

meistern. Neben der öffentlichen Bereitstellung ist insbesondere die Förderung der Wiederverwendung oder Nachnutzung der Daten von zentralem Interesse. Um aus Daten verwertbare Informationen für Bürger und Unternehmen zu gewinnen, d.h. nutzbare Antworten auf eine konkrete Fragestellung zu bekommen, müssen Daten in Beziehung zueinander gestellt werden können und möglichst flexibel in verschiedenste z.B. kartenbasierte oder mobile Anwendungen integrierbar sein. Bedenkt man die Vielzahl der möglichen Wiederverwendungsszenarien und die dafür eingesetzten webbasierten Technologien und Endgeräte, zeigt sich die hohe Anforderung an die Interoperabilität bei der Datenbereitstellung.

Um im Web neben Internetseiten auch Daten interoperabel austauschen und wiederverwenden zu können, hat sich in den letzten Jahren zunehmend der Linked-Data-Ansatz durchgesetzt. Der folgende Abschnitt beschreibt daher näher wie Open Government Data mit dem Linked-Data-Ansatz in der Praxis umgesetzt werden kann.

4.1.1 Grundprinzipien von Linked-Data

Der Linked-Data-Ansatz beschreibt eine Methode und zugehörige Technologien zur Veröffentlichung und Verknüpfung von strukturierten Daten über das World Wide Web (Heath, 2010). Er ermöglicht es Maschinen, also Web-Applikationen, wie z.B. Mashups oder Smartphone-Apps, das Web wie eine strukturierte Datenbank zu nutzen und darin gezielt zu browsen oder spezifische Anfragen zu stellen.

Historisch gesehen kann der Linked-Data-Ansatz als pragmatische Umsetzung der Semantic Web-Vision angesehen werden. Sein maßgebender Schöpfer Sir Tim Berners Lee, der Erfinder des World Wide Web, beschrieb bereits 2001 in seinem viel zitierten Artikel im *Scientific American* (Berners-Lee, Hender & Lassila, 2001) eine Erweiterung des bestehenden Internets. Neben Dokumenten zur Darstellung von Webseiten solle das Internet ebenfalls für den Austausch von Informationen genutzt werden, die von Anwendungen intelligent verarbeitet werden können. Der Linked-Data-Ansatz fokussiert dabei pragmatisch auf diejenigen Konzepte und Standards des Semantic Web, die heute mit wenig Aufwand praxistauglich einsetzbar sind. Weitergehende Ansätze des Semantic Web wie z.B. das Entwerfen von komplexen Ontologien oder aufwendige Algorithmen für logische Schlussfolgerungen (Reasoning) stehen dabei weniger im Vordergrund.

Der Linked-Data-Ansatz zeichnet sich durch sein flexibles und plattformunabhängiges Datenmodell aus. Im Gegensatz zum relationalen Datenmodell

(gebunden an Spalten und Zeilen in Datenbanktabellen) oder dem hierarchischen Datenmodell (gebunden an eine Baumstruktur z.B. in einem XML-Dokument) lassen sich mit dem Linked-Data-Ansatz Daten in Graphstrukturen abbilden. Damit können insbesondere Beziehungen zwischen Daten flexibler ausgedrückt werden, so dass eine Vernetzung von Daten vereinfacht wird. D.h. neue Beziehungen zwischen Daten können einfach hinzugefügt oder geändert werden, ohne an fixe Tabellen- oder Baumstrukturen gebunden zu sein. Das folgende Beispiel illustriert den Linked-Data-Ansatz:

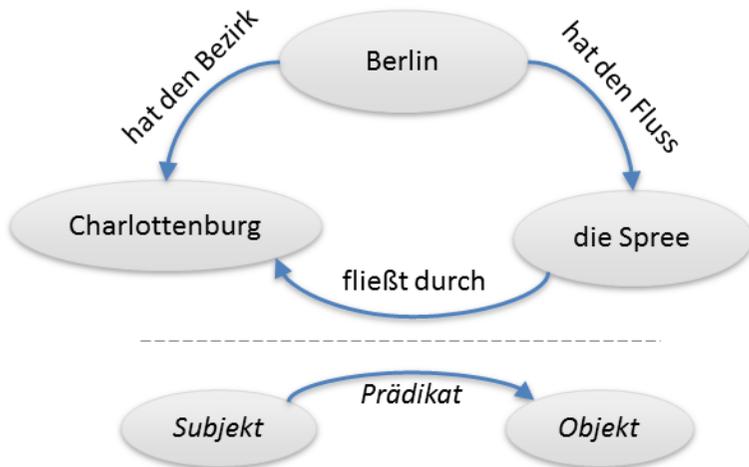


Abbildung 1: graph-basiertes Datenmodell des Linked-Data-Ansatzes

Das graph-basierte Datenmodell setzt sich dabei aus sog. Tripeln zusammen, die jeweils aus einem Subjekt, einem Prädikat und einem Objekt bestehen. Knotenpunkte können als Subjekt oder Objekt fungieren, je nachdem in welcher Beziehung sie zu anderen Knotenpunkten im Datengraphen stehen. Um die Daten über das Web austauschbar und untereinander verknüpfbar zu machen, erhält jedes Subjekt, Prädikat und Objekt einen eindeutigen im Web abrufbaren Identifier. Ist dieser öffentlich zugänglich spricht man von Open Linked Data. Damit wird es möglich, Datensätze aus unterschiedlichen Quellen und IT-Systemen über das Internet im Sinne eines Webs der Daten miteinander zu verknüpfen und kombinierbar abrufbar zu machen.

Durch die explizit gemachten Beziehungen zwischen den Daten, wird ebenfalls der Kontext eines Datensets beschrieben und abrufbar gemacht. Nutzer oder von ihnen verwendete Anwendungen nutzen diese Kontextbeschreibungen, um die Verarbeitung und Einordnung der Daten zu ermöglichen und

Eine wesentliche Rolle für diesen Erfolg spielte die Standardisierung der zugrunde liegenden Technologien des Linked-Data-Ansatzes. Zum einen wurde dadurch die Interoperabilität der einzelnen Initiativen sichergestellt, zum anderen konnten Synergien bei der Werkzeugentwicklung und deren Anwendung genutzt werden. Im folgenden Abschnitt werden daher die grundlegenden Technologien und Standards des Linked-Data-Ansatzes kurz in einem Überblick vorgestellt.

4.1.2 Linked Data Technologien und Standards

Die Schlüsseltechnologien des Linked Data-Ansatzes bauen zum einen auf klassischen Internetstandards der Internet Engineering Task Force (IETF) auf, als auch auf Standards der Semantic Web Initiativen, die durch das World Wide Web Konsortium (W3C) etabliert wurden. Folgende fünf Standards können als die Schlüsseltechnologien des Linked-Data-Ansatzes bezeichnet werden:

- **Uniform Resource Identifier (URI):** URIs sind einheitliche Bezeichner für Ressourcen, die eine Dateneinheit darstellen. Sie ermöglichen es jedem Subjekt, Prädikat und Objekt in einem Datengraphen einen eindeutigen Identifier zuzuordnen, so dass andere auf diese Bezug nehmen können. Ein Beispiel für so einen eindeutigen Bezeichner für die Dateneinheit Berlin in unserem obigen Datengraphen (vgl. Abb. 1) wäre die URI <http://www.beispiel.net/Berlin>. Eine sinnvolle Nutzung von URIs ist ebenfalls deren Verwendung als Webadresse. Beim Aufruf eines URI im Webbrowser erhält man dann direkt die betreffende Dateneinheit oder deren Beschreibung inkl. ihrer Beziehungen im Datengraphen. Analog zu Links in Webseiten wird so ein Browsen durch den Datengraphen und seinen Verknüpfungen ermöglicht.
- **Hypertext Transfer Protocol (HTTP):** Um den Aufruf von URIs und den Transport der dahinter stehenden Dateneinheiten über das Internet zu ermöglichen, wird das Kommunikationsprotokoll HTTP analog wie beim Aufruf von Webseiten über einen Webbrowser verwendet. Möchte man z.B. Informationen über Berlin aus dem obigen Beispiel eines Datengraphen erhalten, so verwendet man den HTTP-Befehl GET und die Webadresse <http://www.beispiel.net/Berlin>. Die Anfrage wird dadurch über das Internet an den Server www.beispiel.net gerichtet, der die Ressource Berlin z.B. mit einer Beschreibung und den Beziehungen zu weiteren Daten zurückliefert.

- **Resource Description Framework (RDF):** Die Beschreibung der Ressourcen erfolgt anhand des graph-basierten Datenmodells, das durch das Resource Description Framework definiert wird. Wesentlich ist dabei der Grundgedanke, Aussagen über Ressourcen anhand von Tripeln zu definieren. Jedes Subjekt, Prädikat und Objekt ist dabei durch eine eindeutige URI bezeichnet.

Unsere exemplarischer Datengraph (vgl. Abb. 1) würde sich demnach durch folgende Tripel ausdrücken lassen:

Subjekt	Prädikat	Objekt
http://www.beispiel.net/Berlin	http://www.beispiel.net/hat_den_Fluss	http://www.beispiel.net/Spree
http://www.beispiel.net/Berlin	http://www.beispiel.net/hat_den_Bezirk	http://www.beispiel.net/Charlottenburg
http://www.beispiel.net/Spree	http://www.beispiel.net/fließt_durch	http://www.beispiel.net/Charlottenburg

Tabelle 2: Beispiel eines in Tripeln modellierten Datengraphen

- **RDF Schema (RDFS), Web Ontology Language (OWL):** Im RDF-Datenmodell können einem Subjekt über beliebige Prädikate beliebige Objekte zugeordnet werden. Um diese Beliebigkeit einzuschränken, kann anhand von Ontologie-Sprachen wie RDFS oder OWL festgelegt werden, welche Klassen von Dateneinheiten im Kontext der jeweiligen Anwendung als sinnvoll erachtet werden und welche Beziehungen zwischen ihnen herrschen können. So macht es z.B. Sinn im obigen Beispiel (vgl. Abb. 1) das Prädikat „hat den Fluss“ nur Objekten zuzuordnen, die aus einer zu definierenden Klasse „Fluss“ stammen. So definierte Ontologien lassen sich dann innerhalb von Anwendungen nutzen, um z.B. verschiedene Flüsse miteinander zu vergleichen oder aggregiert darstellen zu können. Ontologien werden also für bestimmte Zwecke von Fachleuten eines Gebietes geschaffen, das die jeweilige Ontologie strukturieren soll.
- **SPARQL Protocol and RDF Query Language:** Die expliziten Beziehungen zwischen Dateneinheiten erlauben es im Linked-Data-Netz zu browsen. Um darüber hinaus auch die Möglichkeit der Recherche zu bieten, wird die Abfragesprache SPARQL verwendet. SPARQL ermöglicht es, komplexe Abfragen über Linked-Data-Quellen zu machen analog einer Abfrage in einer Datenbank anhand von Variablen und einschränkenden Attributen. Für unseren exemplarischen Datengraphen (vgl. Abb 1) könnte

man z.B. die Abfrage nach allen Flüssen in Berlin wie folgt formulieren:

```
PREFIX b: <
http://www.beispiel.net/>
SELECT ?fluss
WHERE {
  b:Berlin b:hat_den_Fluss ?fluss.
}
```

Tabelle 3: Beispielhafte SPARQL-Abfrage

Auf Basis solcher Abfragen können Entwickler Anwendungen aufbauen, wobei die direkte Nutzung von SPARQL von Endanwendern weniger im Vordergrund steht.

4.1.3 Vorteile des Linked-Data-Ansatzes für Open Government Data

Zusammenfassend lassen sich folgende Vorteile des Linked-Data-Ansatzes für die Bereitstellung von öffentlichen Daten herausstellen:

- Linked Data können durch ihr maschinenlesbares Format leicht weiterverarbeitet und durch externe Akteure nachgenutzt werden.
- Das generische graph-basierte Datenmodell kann als Mediator zwischen den verschiedensten heterogenen Datenquellen des öffentlichen Sektors agieren und Interoperabilität sicherstellen.
- Der Linked Data-Ansatz bietet mächtige Abfragemechanismen, die Recherchemöglichkeiten über öffentlichen Daten optimieren.
- Die zugrundeliegenden etablierten Standards erleichterten es, unabhängige Datensätze über organisatorische Grenzen hinweg miteinander in Beziehung zu setzen und somit neue Synergien zu ermöglichen.

Dabei ist es wichtig herauszustellen, dass Open-Data-Initiativen sich schrittweise dem Linked-Data-Ansatz nähern können. Der initiale Schritt ist daher auch nicht die Technologiewahl, sondern die grundsätzliche Strategie öffentliche Daten möglichst frei zur Nachnutzung zur Verfügung zu stellen. In diesem Zusammenhang schlägt Tim Berners Lee, der maßgeblich an der Konzeption

und Umsetzung der Open-Data-Plattform Großbritanniens data.gov.uk mitgewirkt hat, ein 5-Sterne Bewertungssystem vor, um Orientierung für eine stufenbasierte Umsetzungsstrategie zu geben. Es beschreibt aufeinander aufbauende Kriterien für die Bereitstellung von öffentlichen Daten ausgehend von der Veröffentlichung von Rohdaten bis hin zur Nutzung der Vorteile des Linked-Data-Ansatzes (Berners-Lee, 2009):

	Daten sind über das Web verfügbar (unabhängig vom Format) mit einer offenen Nutzungslizenz.
	Daten sind strukturiert und maschinenlesbar verfügbar.
	Daten sind in nicht proprietären Formaten verfügbar (z.B. CSV anstatt Excel).
	Für die Datenbereitstellung werden offene Standards des W3C (RDF und SPARQL) verwendet.
	Daten werden miteinander in Beziehung gesetzt, um deren Kontext explizit verfügbar zu machen.

Tabelle 4: Fünf-Sterne Bewertungssystem für Open Data

Darüber hinaus wird die Bereitstellung von übergeordneten Beschreibungen von Daten als weiteres Qualitätskriterium angeführt. Diese sogenannten Metadaten beschreiben z.B. den Ursprung der Datenquellen oder deren zeitlichen Bezug. Sie sollten über ein öffentliches Verzeichnis wie z.B. ckan.net¹² oder nationale bzw. regionale Open-Data-Portale auffindbar sein.

5 Modell und Architektur einer Open-Data-Plattform nach dem Linked-Data-Ansatz

Im folgenden Abschnitt soll diskutiert werden, wie eine Open-Data-Plattform nach dem Linked-Data-Ansatz konzipiert werden kann. Zunächst wird ein generelles Modell einer Open-Data-Plattform und der relevanten Akteure vorgestellt. Anschließend wird eine Architektur aufgezeigt, deren Komponenten nach dem Linked-Data-Ansatz aufgebaut sind.

¹² Open Knowledge Foundation, Comprehensive Knowledge Archive Network (CKAN)

5.1 1 Modell einer Open-Data-Plattform

Der grundlegende Aufbau einer Open-Data-Plattform kann im Wesentlichen mit einem Vier-Schichten-Modell beschrieben werden. Die Nutzerschicht der Open-Data-Plattform umfasst den Zugang für Bürger, Unternehmen und öffentliche Einrichtungen. Diese agieren zum einen als Kunden oder Konsumenten der Open-Data-Plattform, als auch im interaktiven Sinne als Bereitsteller oder Lieferanten der Daten. Sie agieren als sogenannte Prosumenten (Leitl, 2008), denen zugleich die Rolle des Produzenten und des Konsumenten inneohnt. Der Zugang zur Open-Data-Plattform erfolgt dabei über eine Anwendungsschicht, in der öffentliche und kommerzielle Applikationen nutzungsspezifisch zur Verfügung stehen. Der Fokus wird hierbei auf die Innovationskraft des Marktes und der Gesellschaft gesetzt, die, unterstützt durch die öffentliche Hand, Mehrwertdienste und Applikation auf Basis des Datenbestandes der Open-Data-Plattform hervorbringt. Die somit vorrangig extern bereitgestellten Anwendungen kommunizieren über eine Diensteschicht mit der Open-Data-Plattform. Für den Anwendungsentwickler spielt es dabei keine Rolle aus welchem IT-System die unterschiedlichen Daten ursprünglich kommen. Er nutzt einen einheitlichen Zugangspunkt und bezieht die Daten im Sinne von „Data-as-a-Service“. Die Basis der Open-Data-Plattform bildet eine Infrastrukturschicht, welche die Integration und Bereitstellung von umfassenden öffentlichen und ggf. auch kommerziellen Daten ermöglicht. Das folgende Schaubild veranschaulicht diesen Ansatz:

Modell „City Data Cloud Berlin“ - Leitsätze

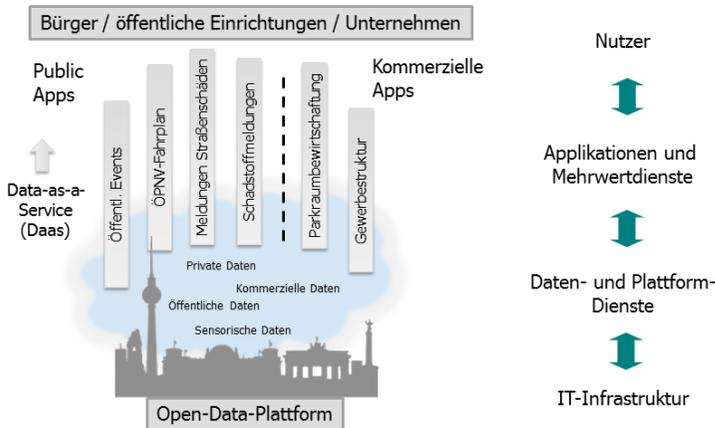


Abbildung 3: Modell einer Open-Linked-Data-Plattform

Den Kern der Open-Data-Plattform bilden vorwiegend öffentliche Daten. Darüber hinaus kann zusätzlich die Möglichkeit gegeben werden, auch kommerzielle Daten (wie z.B. aus dem Immobilien- oder Marktanalysenbereich) und Daten privaten Ursprungs¹³ bereitzustellen. Diese können miteinander Mehrwert-orientiert verknüpft und dann über verschiedene Anwendungskanäle nutzbar gemacht werden.

5.1 2 Basiskomponenten und Grobarchitektur einer Open-Data-Plattform

Im Folgenden wird die Grobarchitektur einer Open-Data-Plattform skizziert. Diese besteht wie in Abbildung 4 dargestellt aus fünf wesentlichen Basiskomponenten:

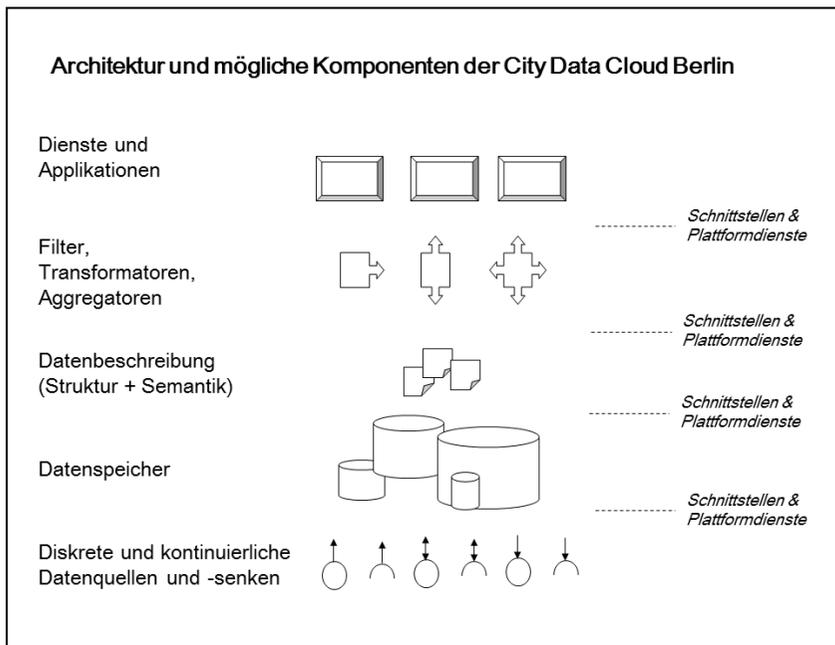


Abbildung 4: Basiskomponenten einer Open-Data-Plattform

13 z.B. bzgl. persönlicher Bewertungen von Dienstleistungen oder Beiträge im Rahmen von Beteiligungsverfahren, jedoch keine unter Datenschutz fallenden personenbezogenen Daten

Um die Grobarchitektur anschaulich zu beschreiben, wird der Weg der Daten von der Übergabe an die Open-Data-Plattform bis zu dessen Nutzung über entsprechende Applikationen beschrieben. Die technischen Eckpfeiler für eine Realisierung werden dabei jeweils kurz diskutiert:

— Diskrete und kontinuierliche Datenquellen- und Senken

Ein wichtiger Ansatz für den Erfolg einer Open-Data-Plattform ist es, die bestehenden Systeme weder technisch noch organisatorisch zu beeinträchtigen. D.h. es müssen möglichst flexibel - angepasst an den jeweiligen Ursprung der Daten - verschiedene Mechanismen bereitgestellt werden, um die Daten für die Open-Data-Plattform nutzbar zu machen.

Das umfasst zum einen die Möglichkeit, Daten komplett als Kopie (z.B. als Datenbankabzug oder auch semi-strukturiert in Form von Spreadsheets) zu beziehen, um diese dann innerhalb der Open-Data-Plattform weiter aufzubereiten und zu verarbeiten (diskrete Datenquelle). Zum anderen muss ebenfalls die Möglichkeit gegeben sein, an eine Organisation gebundene Daten in ihren bisherigen Systemen zu belassen und nur einen Zugriff auf die Daten durch die Open-Data-Plattform bereitzustellen (kontinuierliche Datenquellen). Um dies zu ermöglichen müssen wohldefinierte Schnittstellen und Plattformdienste geschaffen werden, die möglichst auf Web-Service-Standards (W3C, 2004) beruhen und einem service-orientierten Architekturansatz folgen, um eine hohe Interoperabilität zu gewährleisten. Die Schnittstellen und Plattformdienste sollten dabei echtzeit-basierte Infrastrukturen unterstützen, um ebenfalls sensorische Daten kontinuierlich in die Open-Data-Plattform einspeisen zu können. Dies kann z.B. bei Echtzeitinformationen zu Verbindungen oder aktuellen Störungen des öffentlichen Nahverkehrs relevant sein.

Neben den Datenquellen können analog auch Datensinken bedient werden, die kontinuierlich oder diskret einen direkten Zugriff auf die Daten benötigen.

— Datenspeicher und Datenbeschreibung

Die Daten liegen nach der Übergabe oder durch Zugriff zunächst noch in ihrer Rohfassung vor. In den Datenspeichern der Open-Data-Plattform lassen sie sich dann möglichst automatisiert aufbereiten und strukturieren, so dass deren weitere Verarbeitungsfähigkeit erhöht wird. In diesem Kontext kann auch die explizite Datenpflege zur Verbesserung der Datenqualität als weiterer Wertschöpfungsschritt gesehen werden.

Damit die Daten flexibel von möglichst vielen potentiellen Applikationen verarbeitet werden können, sollten Aspekte wie Nutzung offener Standards,

Modularisierbarkeit und Skalierbarkeit sichergestellt sein (Berners-Lee, 2009). Um diese Ziele zu gewährleisten, empfiehlt es sich daher dem Linked-Data-Ansatz zu folgen, der vorsieht Daten in einer ausdrucksstarken Graphstruktur zu beschreiben (vgl. Abschnitt 4.1.1). Somit ist sichergestellt, dass sämtliche heterogene Datenformate der Quellen auf beliebige Datenformate der Applikation abbildet werden können, wobei das generische RDF-Modell als flexibler Mittler gesehen werden kann. Sogenannte Graphdatenbanken oder RDF-Triple-Stores dienen dabei als hoch-performante persistente Datenspeicher (Rodriguez, 2009).

Modularität wird durch die Linked-Data-Technologie insbesondere dadurch sichergestellt, dass unterschiedliche Vokabulare für die Datenbeschreibung und Strukturierung verwendet werden können, ohne dass ein umfassendes Datenschema über den gesamten Datenbestand der Open-Data-Plattform vorab aufwendig abgestimmt und festgelegt werden muss. Dort wo notwendig lassen sich dann Interoperabilitätsmechanismen durch Abbildungen zwischen verschiedenen Vokabularen integrieren.

Darüber hinaus wird die Skalierbarkeit bzw. Evolution der Datenbeschreibung durch die Fähigkeit der Linked-Data-Technologie gewährleistet. Sie erlaubt ändernde oder neu erforderliche Terme und Definitionen flexibel zu ergänzen, ohne die bisherigen Datenstrukturen aufbrechen zu müssen oder neu aufzusetzen (Berners-Lee, 2009).

— **Filter, Transformatoren, Aggregatoren**

Um die Daten möglichst flexible und effizient für die darauf aufsetzenden Dienste und Applikation verarbeitbar und auswertbar zu machen, werden umfangreich konfigurierbare Filter, Transformatoren und Aggregatoren für Datenfusion bereitgestellt. Diese basieren auf geeigneten Abfrage- bzw. Datenmanipulationssprachen wie z.B. SPARQL (W3C, 2008). Damit lassen sich anhand von frei auswählbaren Eigenschaften und Mustern der Datenbeschreibungen die jeweils relevanten Datensätze aus der Open-Data-Plattform herausfiltern. Das Rückgabeformat der Abfragen lässt sich dabei beliebig transformieren, so dass die Ergebnisse auch in Datenformaten bereitgestellt werden, wie sie für die Entwicklung von Web-Applikationen verwendet werden (z.B. strukturiertes XML oder JSON). Analog lassen sich somit auch flexible Aggregatoren konfigurieren, die unterschiedliche Datensätze zusammengeführt bereitstellen - auch solche, die mit heterogenen Vokabularen beschrieben sind. Die Filter, Transformatoren und Aggregatoren können jeweils in Echtzeit über Dienste, als auch anwendungsspezifisch vorverarbeitet (logisch aggregiert) oder als persistierte Replikate (physisch aggregiert) genutzt werden.

— Dienste und Applikationen

Dienste und Applikation werden, wie bereits oben erläutert, vorrangig extern entwickelt und bereitgestellt. Diese umfassen klassische Web-basierte Anwendungen, die die bereitgestellten Daten der Open-Data-Plattform in ihre Portale integrieren und ggf. mit weiteren Informationen anreichern und verknüpfen. Einen weiteren Schwerpunkt bilden sowohl mobile Applikationen des sich stark entwickelnden mobilen Internets, als auch Dienste und Applikationen für spezifische eingebettete Systeme.

Insgesamt lässt sich feststellen, dass die notwendigen technischen Infrastrukturkomponenten bereits vorhanden und weitgehend zudem als Open-Source-Software verfügbar sind. Die Herausforderung besteht demnach insbesondere in der Auswahl, Integration und bedarfsorientierten Anpassung der Komponenten.

6 Zukünftige Entwicklung

Für die weitere Entwicklung von Open Data im öffentlichen Sektor ergeben sich einerseits bereits heute viele Möglichkeiten, andererseits stellen sich auch viele Fragen und derzeit nicht gelöste Herausforderungen.

Fasst man das Fünf-Sterne-System für Open Data auch als Anleitung zu einer prozesshaften Vorgehensweise auf, so können Organisationen des öffentlichen Sektors in relativ kurzer Zeit erste Schritte in Richtung offener Verwaltungsdaten unternehmen. Diese Anfänge dienen gleichzeitig der Auseinandersetzung mit der Thematik und können die Grundlage einer folgenden intensiveren und strategischen Auseinandersetzung mit der Datenbereitstellung bilden.

1	Überblick verschaffen, welche Daten der Organisation bereits heute im Internet zugänglich sind
2	Ein einfaches Verzeichnis der bereits öffentlich online verfügbaren Daten erstellen und im Internetportal der Organisation veröffentlichen
3	Bei ausgewählten, bereits im Internet verfügbaren Datensammlungen die existierenden Nutzungsbedingungen klar kennzeichnen und beschreiben
4	Mittels einer Online-Umfrage in Form eines Formulars ermitteln, welche Daten/Themengebiete die Einwohner einer Gebietskörperschaft interessieren

5	Existierende Ansätze in der eigenen Organisation zur Bereitstellung von Verwaltungsdaten und deren Erfahrungen kennenlernen (z.B. Bereitstellung von Geodaten, statistische Daten)
6	Durchführung eines Ideen- und Anwendungswettbewerbes auf Basis bereits öffentlich zugänglicher Verwaltungsdaten

Tabelle 5: Auswahl erster praktischer Schritte in Richtung offener Verwaltungsdaten (die Nummerierung stellt keine Gewichtung dar)

Zur Verstetigung erster Aktivitäten sollte mittel- bis langfristige die Entwicklung eines strategischen Ansatzes für offene Verwaltungsdaten in der Organisation angegangen werden (siehe z.B. Hogge, 2010). Hierbei gilt es dann die regulativ-organisatorischen Rahmenbedingungen ebenso wie die lokalen technologischen Voraussetzungen und eventuell vorhandene Ressourcen zu untersuchen und die angestrebte Bereitstellung offener Daten mit existierenden organisationalen Entwicklungszielen zu verknüpfen.

Die Entwicklung einer Infrastruktur für die Bereitstellung offener Daten des öffentlichen Sektors stellt sicher ein langfristiges Ziel dar, das die Verwaltungseinheiten in Deutschland nachhaltig verändern dürfte. Die Realisierung des Open-Data-Ansatzes verlangt sowohl bottom-up Aktivitäten, die eher kurzfristig angegangen werden können (lokale Initiativen, erste Schritte im 5-Sterne-System) und eher langfristige Anstrengungen top-down, die für Koordination sorgen.

Transparenz organisationaler Vorgänge und Entscheidungen erlangt in Zukunft eine immer stärkere Bedeutung. Industrienationen suchen durch die Fortentwicklung wissensintensiver Produktion Alleinstellungsmerkmale. Für die Entwicklung der digitalen Gesellschaft des 21. Jahrhunderts stellen somit offene Daten und ihre zuverlässige Verfügbarkeit auf elektronischem Wege einen wesentlichen Baustein dar. Gerade für den öffentlichen Sektor ergibt sich hier eine Chance, die eigene Modernisierung voranzutreiben und gleichzeitig den Anforderungen einer offenen Gesellschaft besser gerecht zu werden.

Literatur

- Almirall, E. (2010). Open Innovation Mechanisms in Smart Cities (Description of Work No. 270896). Competitiveness and Innovation Framework Programme.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American* 284, no. 5: 34.
- Berners-Lee, T. (2009). Linked Data. *Design Issues*. Zugriff 10. November 2011 unter <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2009). Putting Government Data online. *Design Issues*. World Wide Web Consortium. Zugriff 10. November 2011 unter <http://www.w3.org/DesignIssues/GovData>
- Bundesministerium des Innern (Hrsg.). (2010). Regierungsprogramm - Vernetzte und transparente Verwaltung. Zugriff 10. November 2011 unter http://www.verwaltung-innovativ.de/cln_117/SharedDocs/Publikationen/Pressemitteilungen/regierungsprogramm__vernetzte__und__transparente__verwaltung.templateId=raw.property=publicationFile.pdf/regierungsprogramm_vernetzte_und_transparente_verwaltung.pdf
- Dekkers, M., Polman, F., te Velde, R. & de Vries, M. (2006). MEPSIR - Measuring European Public Sector Information Resources. Final Report of Study on Exploitation of public sector information - benchmarking of EU framework conditions. (Hrsg.: European Commission).
- Florini, A. (2007). *The right to know: transparency for an open world*. Columbia University Press.
- Fornfeld, M., Boele-Keimer, G., Recher, S. & Fanning, M. (2009). Assessment of the Re-use of Public Sector Information (PSI) in the Geographical Information, Meteorological Information and Legal Information Sectors. (Hrsg.: MICUS Management Consulting GmbH).
- Glock, C. & Broens, M. (2008). Wissensmanagement in der öffentlichen Verwaltung mithilfe von Social-Software-Applikationen. *Verwaltung & Management - Zeitschrift für allgemeine Verwaltung*, 4(5), 273-276.
- Heath, T. (2010). Linked Data – Connect Distributed Data across the Web. Zugriff 10. November 2011 unter <http://www.linkeddata.org>
- Hogge, B. (2010). Open Data Study. (Hrsg.: Transparency and Accountability Initiative). Zugriff 10. November 2011 unter http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519/open-data-study-100519.pdf

- Holtkamp, L. (2006). Partizipative Verwaltung - hohe Erwartungen, ernüchternde Ergebnisse. In J. Bogumil, W. Jann & F. Nullmeier (Hrsg.), *Politik und Verwaltung* (S. 185-207). Wiesbaden: Verlag für Sozialwissenschaften.
- Kubicek, H. (2007). Electronic Democracy and Deliberative Consultation on Urban Projects - "Putting E-Democracy into Context" (Report for the Congress of Local and Regional Authorities). Zugriff 10. November 2011 unter http://www.ifib.de/publikationsdateien/Creative_final.pdf
- Kubicek, H. (2008). Next generation FoI between information management and web 2.0. *dg.o ,08 Proceedings of the 2008 international conference on Digital government research* (S. 9-16). Digital Government Society of North America. Zugriff 10. November 2011 unter <http://portal.acm.org/citation.cfm?id=1367832.1367838>
- Leitl, M. (2008). Was ist ein Prosument?. *Harvard Business Manager. Heft 3/2008*, Schwerpunkt Personal. Zugriff 10. November 2011 unter www.harvardbusinessmanager.de/heft/artikel/a-618967.html
- Linnert, K., Kubicek, H. & Sommer, I. (2011). Bremer Empfehlung zu Open Government Data - Ein elektronischer Weg zu besserem Informationszugang und mehr Transparenz in der öffentlichen Verwaltung. Zugriff 10. November 2011 unter http://www.finanzen.bremen.de/sixcms/media.php/13/2011-01-17_BremerErklaerung.pdf
- von Lucke, J. & Geiger, C. (2010). Open Government Data - Frei verfügbare Daten des öffentlichen Sektors (Gutachten für die Deutsche Telekom AG zur T-City Friedrichshafen). Deutsche Telekom Institute for Connected Cities, Zeppelin University Friedrichshafen.
- Newbery, D., Bently, L. & Pollock, R. (2008). Models of Public Sector Information Provision via Trading Funds. (Office for Public Sector Information (OPSI), United Kingdom, Hrsg.). Zugriff 10. November 2011 unter <http://www.opsi.gov.uk/advice/poi/models-psi-via-trading-funds.pdf>
- NRW SPD (Hrsg.). (2010). Gemeinsam neue Wege gehen - Koalitionsvertrag zwischen der NRW SPD und Bündnis 90 / Die Grünen NRW. Zugriff 10. November 2011 unter http://www.gruene-nrw.de/fileadmin/user_upload/landesverband/gruene-nrw/aktuelles/2010/koalitionsvertrag/Koalitionsvertrag_Rot-Gruen_NRW_2010-2015.pdf
- Office of Fair Trading (Hrsg.). (2006). The commercial use of public information (CUPI). Zugriff 10. November 2011 unter http://www.oft.gov.uk/shared_oftr/reports/consumer_protection/oft861.pdf
- Open Knowledge Foundation. (2010). Open Definition - Defining the Open in Open Data, Open Content and Open Services. Open Definition. Zugriff 10. November 2011 unter <http://www.opendefinition.org/okd/>

- Piotrowski, S. J. & Van Ryzin, G. G. (2007). Citizen Attitudes Toward Transparency in Local Government. *The American Review of Public Administration*, 37(3), 306-323. doi:10.1177/0275074006296777
- Pollock, R. (2009). The Economics of Public Sector Information (Cambridge Working Papers in Economics). Faculty of Economics, University of Cambridge. Zugriff 10. November 2011 unter <http://econpapers.repec.org/RePEc:cam:camdae:0920>
- Raimond, Y., Scott, T., Sinclair, P., Miller, L., Betts, S., McNamara, F. (2010). Case Study: use of Semantic Web Technologien on the BBC Web Sites. W3C Semantic Web Use Cases and Case Studies.
- Relly, J. E. & Sabharwal, M. (2009). Perceptions of transparency of government policymaking: A cross-national study. *Government Information Quarterly*, 26(1), 148-157. doi:10.1016/j.giq.2008.04.002
- Schenk, F., Kruse, F. & Klenke, M. (2010). From Simple Data Sources to a Complex Information System: Integrating Heterogeneous Data Models into an Information Infrastructure for the Public Administration. gehalten auf der 24th International Conference on Informatics for Environmental Protection in cooperation with Intergeo 2010, Integration of Environmental Information in Europe, Bonn. Zugriff 10. November 2011 unter http://www.kst.portalu.de/pub/praes/2010/praes_enviroinfo_bonn_sk_10-2010.pdf
- Schmidt, M.G. (2008). Beteiligungszentrierte Demokratietheorien. *Demokratietheorien* (4th Aufl., S. 236-253). Wiesbaden: Verlag für Sozialwissenschaften.
- Sebastopol Group. (2007, Dezember 7). 8 Principles of Open Government Data. Zugriff 10. November 2011 unter http://resource.org/8_principles.html
- Shadbolt, N., O'Hara, K., Salvadores, M. & Alani, H. (2011). eGovernment. *Handbook of Semantic Web Technologies* (S. 840-900). Springer-Verlag. Zugriff 10. November 2011 unter <http://eprints.ecs.soton.ac.uk/21711/>
- Sunlight Foundation. (2010, August 11). Ten Principles for Opening Up Government Information. Sunlight Foundation. Zugriff 10. November 2011 unter <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>
- W3C. (2004). Web Services Architecture. Zugriff 10. November 2011 unter <http://www.w3.org/TR/ws-arch/>

- Rodriguez, M. A. (2009). A Reflection on the Structure and Process of the Web of Data. *Bulletin of the American Society for Information Science and Technology*. 35(6).
- W3C (2008). SPARQL Query Language for RDF. W3C Recommendation. Zugriff 10. November 2011 unter <http://www.w3.org/TR/rdf-sparql-query/>