

Das Potential von Datenjournalismus im World Wide Web*

Von Daniela Woytewicz

1. PROBLEMHINTERGRUND

„Die Zahl derer, die durch zu viele Informationen nicht mehr informiert sind, wächst.“

Rudolf Augstein (* 1923 - † 2002), dt. Journalist, Herausgeber „Der Spiegel“

Die immer engere Vernetzung von Informationen über das Internet führt zu einem stetigen Wachstum der global verfügbaren Datenmasse, die kaum jemand mehr durchdringen kann. Emails, Blogbeiträge, Bilder, Dokumente und Videos können von den Internetnutzern in dem Bruchteil einer Sekunde in das World Wide Web hochgeladen und abgerufen werden. Und diese Entwicklung wird zunehmend beschleunigt: Es sind nun nicht mehr allein die Internetnutzer, welche Informationen im Internet produzieren, abrufen und speichern, - in Form von Blogbeiträgen, wissenschaftlichen Arbeiten, in Medienartikeln, bei Facebook, Twitter etc. - sondern auch Milliarden von Geräten, die dasselbe tun: „[...] Smartphones, GPS-Geräte, Kassen, Minicomputer in unseren Autos und andere Geräte, Browser, die unsere Web-Aktivitäten beobachten, Kameras, Sensoren, Musikgeräte und und und [...]“ (Konitzer 2011). Daten kommen daher aus nahezu unendlich vielen Quellen und sind überall und zu jeder Zeit abrufbar. Der Ausdruck Big Data wird daher gewöhnlich im Zusammenhang mit der Datenflut im Internet verwendet (Vgl. Konitzer 2011). Big Data beschreibt große Massen an Daten, die rasant und unaufhörlich wachsen, „aus vielen verschiedenen Quellen in sehr kurzer Zeit neu entstehen und gleichsam von sehr vielen Anwendern genutzt werden können“ (Landrock/ Zilch 2012), die daher kaum oder schwer beherrschbar scheinen, in deren polystrukturierten Daten aber ein potentieller Mehrwert liegt. Vorausgesetzt man besitzt die dafür notwendige Medienkompetenz, mittels Werkzeugen mit den Daten umzugehen. Das Marktforschungs- und Analyseunternehmen

in der weltweiten Technologie-Industrie Gartner macht daraus eine einfache Gleichung: „Big Data equals Big Opportunity“ (Vgl. IBM 2011: 2). Bereits 2006 wies der Werbefachmann Michael Palmer darauf hin: „Daten sind das neue Öl“ (Palmer 2006). Der Vergleich stimmt: Sie sind einer der wichtigsten und ertragreichsten Rohstoffe der heutigen Informationswirtschaft, ähnlich wie es Kohle und Eisenerz in der industriellen Revolution waren (vgl. IBM 2011). Allerdings sind sie nur dann hilfreich, wenn sie entsprechend verarbeitet werden. Unbearbeitete Daten sind wertlos, erst das Interpretieren des Materials und die Rückschlüsse, die daraus gezogen werden, schaffen den gewünschten Mehrwert (vgl. Matzat 2011b).

2. DER ONLINE-JOURNALIST

So stellt der immer einfacher werdende Zugang zu globalem Wissen und Informationen, die massenhafte Übertragung und Speicherung von digitalen Daten im World Wide Web sowie der Kampf um die Informationshoheit im Netz den Journalismus vor neue Herausforderungen: Es müssen Strategien erprobt werden, um überhaupt erst an relevante Informationen zu gelangen und nicht im Fluss der Big Data zu versinken. „Denn je größer der Datenberg wird, desto mehr gleicht die Suche nach relevanten Informationen der Suche nach der Nadel im Heuhaufen“ (Bensch 2011). Journalisten und vor allem Online-Journalisten müssen lernen, den Überblick innerhalb der Big Data zu behalten, um ihrer Funktion als „Gatekeeper“ im Datenmeer gerecht zu werden. Der Umgang mit Daten führt den Journalisten dazu, sich neue Arbeitstechniken und Werkzeuge anzueignen. Dabei soll die in diesem Sinne neuerworbene Medienkompetenz nicht alleine die Recherche zum Ziel haben. Es soll Wissen produziert werden, damit aus der dispersen Datenmasse Erkenntnisse und Informationen herausgefiltert und dem Rezipienten anschließend verständlich dargestellt werden.



Daniela Woytewicz
Fachhochschule Köln
Online-Redakteurin
(B.A.)
Gotthelfstraße 9
50825 Köln
0178 / 4004544
daniela.woytewicz@
gmail.com

*Vortragsmanuskript
(gehalten auf der
Frühjahrstagung des
vfm am 23. April
2013)

Aus diesen Entwicklungen und den daraus resultierenden Chancen sowie Anforderungen im Umgang mit Daten an den Journalisten von heute hat sich ein neues Genre im Online-Journalismus etabliert, der Datenjournalismus.

3. DEFINITION DATENJOURNALISMUS

Bisher herrscht in der Literatur und unter den Fachexperten keine einheitliche Meinung darüber, was genau unter dem Phänomen Datenjournalismus zu verstehen ist. Deshalb erfolgt im Anschluss durch die Darstellung mehrerer umschreibender Definitionen, die den Begriff abgrenzen, einordnen und möglichst konkretisieren, eine Annäherung an den Terminus Datenjournalismus.

Datenjournalismus ist eine Form des Online-Journalismus, der auf der Analyse und Filterung großer und vielschichtiger Datenmengen basiert. Er dient dazu, komplexe Zusammenhänge, die in Schrift oder Tabellenform unüberschaubar wären, verständlich aufzubereiten und in genuinen Erzählformen darzustellen (vgl. Dean 2012). Dabei handelt es sich um eine Kombination aus Recherche-Ansatz und neuer interaktiver Veröffentlichungsform, bei der das Medium World Wide Web zugleich als Recherche- sowie Publikationsplattform dient (vgl. Matzat 2010c: 6).

Allerdings arbeiteten Journalisten nicht erst seit Aufkommen dieses neuen Genres mit elektronischen Daten. Seit der flächendeckenden Einführung von Computern und Internetzugang zu Beginn des neuen Jahrtausends bedienen sich klassische journalistische Medien sowie der Online-Journalismus des Internets als Rechercheinstrument und unterfüttern ihre Berichterstattung mit Statistiken, Fotos und Infografiken (vgl. Schneider/ Raue 2003: 251 ff.). Der neue Ansatz des Datenjournalismus ist laut Christina Elmer, Datenjournalistin beim Wochenmagazin Stern, dass der Datensatz Ausgangsbasis und Gegenstand der Berichterstattung zugleich ist: „In der neuen Form geht es anders herum, dass ich mir zuerst die Daten anschau und dann in der Auswertung eine Idee davon bekomme, was ist eigentlich meine Geschichte“ (Zapp 2011: 0:40 bis 0:55 Minute). So dienen laut Elmer Daten nicht mehr allein zur Unterfütterung einer vorangestellten These erst aus ihnen heraus entsteht die eigentliche Fragestellung. Wurden Daten im klassischen Journalismus vorher nur zur Ergänzung von Beiträgen zum besseren Verständnis des Inhalts genutzt, verwendet der Datenjournalismus die unverarbeiteten Rohdaten nicht mehr rein als Recherchequelle, sondern erhebt sie zur Grundlage jeglicher Berichterstattung und zentralem Gegenstand der Geschichte und Präsentation (Vgl. Matzat 2011c). So erklärt der als Medienpädagoge tätige Mitgründer der Agentur OpenDataCity und Journalist Lorenz Matzat die Entwicklung zum heuti-

gen Datenjournalismus: „[...] Das Neue daran ist, Datensätze zum Gegenstand des Berichts und zu einer nützlichen interaktiven Rechercheumgebung zu machen. Bislang sind Datensätze oft als Beiwerk oder Unterstützung gesehen worden, jetzt ist der Datensatz an sich der Gegenstand der Berichterstattung“ (Medium Magazin 2011). Vereinfacht lässt sich dieser Sachverhalt laut der Deutschen Presse-Agentur in folgender Gleichung darstellen (Vgl. Elmer 2012: 3,4):

1. Ursprüngliche Form: These → Daten
Funktion der Daten: Beweismittel, Einordnung, Relevanz verdeutlichen
2. Datenjournalismus: Daten → These
Funktion der Daten: Dienen als Fundgrube für inhaltlichen Mehrwert

Der hier definierte Datenjournalismus existiert nur im und lebt von den Möglichkeiten des World Wide Web. Allein durch die Mittel und Optionen, die das neue Medium bietet, erhält der Datenjournalismus seine eigentliche Form. Für den britischen Online-Journalisten, Blogger und Professor an der City University's School of Journalism in London, Paul Bradshaw kommt er deshalb - rein faktisch - nur online vor. Seine Begründung:

Data journalism may be a new term, but the idea of ploughing through information to find the story is as old as journalism itself - the difference is that the internet and computing power are giving us new tools to help at every stage of the journalistic process. (Bradshaw 2011: 49)

Der Blogger Moritz Shane umschreibt diesen Zustand während der Diskussion über die Veröffentlichung des „Data Journalism Handbook“ im April 2012 folgendermaßen: Datenjournalismus sei eine Kombination von klassischem Journalismus mit den neuen Mitteln, die durch das Vorhandensein und die Zugänglichkeit riesiger Datenmengen im World Wide Web und deren „[...] Möglichkeiten der computer- und Web-2.0-unterstützten Analyse und Darstellung dieser Daten entstanden sind“ (Beckedahl 2012). Stefan Plöchinger, Journalist bei sueddeutsche.de, bringt diesen Fakt auf den Punkt: „Sie ist deshalb so besonders, weil sie nur mit Computertechnik und den Möglichkeiten des Internets denkbar ist. [...] Erst die digitale Welt erschließt dem Journalismus solche ungekannten Präsentationsmöglichkeiten“ (Plöchinger 2012b).

Der Begriff Datenjournalismus kommt ursprünglich aus dem Englischen, man spricht hier vom „data-driven-journalism“ kurz „DDJ“. Wörtlich übersetzt bedeutet dieser Name „datengetriebener Journalismus“ und setzt damit voraus, dass die Berichterstattung nur mit Unterstützung von Datensätzen geschieht; „[...] im Internet per interaktiven Anwendungen, die auf Datenbanken zugreifen“

(Matzat 2011c). Das Wort selbst beinhaltet jedoch auch den Begriff „Journalismus“ und zeigt damit, dass der Fokus trotz des verstärkten Dateneinsatzes letztendlich auf der journalistischen Aufbereitung liegt (vgl. Dean 2012). Neben Datenjournalismus wird im Deutschen auch häufig der Begriff „Datenbankjournalismus“ oder „Datensatzjournalismus“ verwendet. Im Englischen trifft man auf Synonyme wie „computational journalism“, „visual journalism“, sowie „databank journalism“ oder schlicht „datajournalism“ (vgl. Matzat 2011c).

4. CHARAKTERISTIKA

So sehr sich die diversen Definitionsversuche des Datenjournalismus in ihren Ausführungen unterscheiden, so stimmen sie jedoch alle im Kern überein, wenn es um die Aspekte geht, die einen Beitrag zu einem datenjournalistischen Stück werden lassen. Diese Grundcharakteristika können dabei als Gerüst und Voraussetzung für den Datenjournalismus angesehen werden, „um das dann jegliche weitere Schritte gesponnen werden können“ (Krampe 2012a). Im Folgenden werden neun Merkmale für datenjournalistische Werke erläutert, die sich immer wieder in relevanten Texten über Datenjournalismus finden lassen und sich aus der vorherigen Definition ergeben.

- Daten liegen in strukturierter Form digital vor und können digital weiterverarbeitet werden (Matzat 2010a).
- Der Datensatz ist Ausgangsbasis und Gegenstand der Berichterstattung zugleich (Vgl. Bradshaw 2010).
- „Die Daten stehen im Mittelpunkt einer investigativen Recherche“ (Langer 2011: 5).
- Es geht um öffentliche, nicht um private Daten. (Langer 2011: 5)
- Ein vorhandener, maschinenlesbarer Datensatz wird per Software analysiert und miteinander verknüpft. Gegebenenfalls werden auch mehrere maschinenlesbare Datensätze miteinander verbunden (engl. Mashup) (vgl. Krampe 2012).
- Obligatorisch ist die anschließende Darstellung der Erkenntnisse, nicht die reine Recherche. Dabei ist die multimediale und/oder interaktive Darstellung der Daten in einer Anwendung, folglich die Aufbereitung in interaktiven Formaten im World Wide Web frei wählbar. In welcher Form diese geschieht, ob durch Datenbanken, visuelle Grafiken und/oder Animationen, ist zurzeit noch nicht klar umrissen (vgl. Krampe 2012).
- Der Leser kann selbst aktiv mit den Daten umgehen, sie sich individuell erschließen und Vorgänge beobachten (Monitoring) (vgl. Matzat 2010b).
- Die Journalisten veröffentlichen den ganzen Datensatz, also nicht nur die für ihren Beitrag genutzten Daten, so dass die Rohdaten auch von

anderen weiter genutzt werden können und die Recherche gläsern, d.h. transparent wird (vgl. Krampe 2012).

- Ein weiterer entscheidender Faktor datenjournalistischer Projekte kann die Einbindung des Crowdsourcing-Prinzips sein. Der Prozess beschreibt die Einbindung einer Gruppe von Internetnutzern (die sich außerhalb der Strukturen des publizierenden Mediums befindet) in die Recherche und/oder Produktion von Inhalten (vgl. Pelzer 2011). So werden zum Beispiel für das Projekt Zugmonitor der Süddeutschen Zeitung die Leser öffentlich dazu aufgerufen, selbst aktiv am Projekt teilzunehmen, zu recherchieren und zu analysieren (vgl. Plöchinger 2012b). Letztlich meint Crowdsourcing im Datenjournalismus jedoch nicht, dass Laien die Rollen der Profis übernehmen, sondern diese bei ihrer Arbeit unterstützen. „Die Bewertung, Einordnung und Veröffentlichung bleibt üblicherweise im Verantwortungsbereich einer Redaktion“ (Forum Journalismus und Medien Wien 2012).

5. ERZÄHLFORMEN DES DATENJOURNALISMUS

Die verschiedenen Erzählformen des DDJ lassen sich laut Mirko Lorenz, einem der führenden deutschen Datenjournalisten, in drei verschiedene Formen unterteilen, nämlich in Data Stories, in Data Specials und in Data Apps (vgl. Davies 2012: 24). Angelehnt an Definitionen und Charakteristika des Datenjournalismus soll darauf verwiesen werden, dass es sich bei den kommenden Beschreibungen nur um eine Annäherung an die verschiedenen Erzählformen handelt und nicht um verbindliche Muster, da oftmals Mischformen der nachfolgend beschriebenen Kategorien existieren und die Grenzen zwischen ihnen fließend verlaufen.

Data-Stories bilden die einfachste Form von Datenjournalismus-Projekten und kommen zum Einsatz, wenn ein kleiner Datensatz vorliegt. Meistens werden die Ergebnisse der Datenanalyse in einem Balken-, Linien- oder Tortendiagramm dargestellt. Data Stories können im redaktionellen Alltag als Tages-Projekt angesiedelt und durch einen bis zwei Journalisten verwirklicht werden.

Ein Beispiel für eine Data Story lässt sich auf dem Blog „DataRemixed“ von Datenjournalist Ben Jones finden. Er fragte sich nach Felix Baumgartners Sprung aus der Stratosphäre und den ganzen Zahlen und Fakten, die einem in den Medien wortwörtlich um die Ohren geschlagen wurden: „Compared to what?“ und setzte den Sprung in Relation zu anderen bemerkenswerten Messgrößen wie zum Beispiel zur Höhe des Mount Everest oder der Schallgeschwin-

digkeit. Entstanden sind dadurch insgesamt drei datenjournalistische Visualisierungen, die unter dem Titel „A Tribute to fearless Felix“ laufen und folgende Fragestellung beinhalten (vgl. Jones 2012):

- „How high was he relative to other things I am familiar with, like the cruising altitude of a commercial jetliner?“
- How fast did he travel relative to an NHRA dragster or a bullet being fired from a gun?“
- How long was he falling compared to the fastest mile time?“

Die zweite Form, die sogenannten Data Specials, erweisen sich im Gegensatz zu Data Stories bereits als wesentlich komplexer und können in ihrer Umsetzung mehrere Wochen in Anspruch nehmen. Sie werden bei mittleren bis großen Datensätzen verwendet und beinhalten häufig Interaktionsmöglichkeiten, die dem Rezipienten erlauben, mit dem Datensatz zu „spielen“. Das Team für solch ein Data Special umfasst in der Regel einen (oder mehrere) Journalisten und einen Konzeptionier, der die Design-Komponente verwirklicht.

Ein treffendes Beispiel dafür zeigt die Onlineausgabe der Wochenzeitung „Die Zeit“ im Frühjahr 2011 am Beispiel der Berichterstattung „Verräterisches Handy“ (vgl. Zeit.de 2011). Das Projekt soll die Auswirkungen des im Jahr 2007 verabschiedeten Gesetzes zur Vorratsdatenspeicherung, nach dem „Telefon- und Internetverbindungsdaten [...] ein halbes Jahr gespeichert werden“ (Focus-Online 2012) dürfen, auf die Privatsphäre der deutschen Bevölkerung verdeutlichen. Im Zuge dieses Projekts stellte der Grünen-Netzpolitiker Malte Spitz nach erfolgreichem Einklagen von sechs Monaten seiner Ortsdaten bei der Telekom die gespeicherten Vorratsdaten zeit.de zur Verfügung. zeit.de verarbeitete zusammen mit dem Unternehmen OpenDataCity das Exceldokument mit 35.831 Zeilen (vgl. Biermann 2011) in einer interaktiven Datenvisualisierung „auf der die Nutzer alle Aufenthaltsorte, Telefonate und SMS des Politikers nachvollziehen [...]“ (Grimme Online Award 2012) und ein halbes Jahr von Spitz Leben durch selbstständiges Navigieren innerhalb der interaktiven Karte nach eigenem Belieben nacherleben konnten. Angereichert wurde die Visualisierung mit frei verfügbaren Inhalten aus dem Internet, wie Tweets und Blogbeiträgen von Spitz, Links zu Websites auf denen der Politiker erwähnt wurde, Videos zum Thema, sowie dem ergänzenden Artikel „Was Vorratsdaten über uns verraten“ von Zeit.de-Redakteur und Projektmitglied Kai Biermann (vgl. Langer 2011: 5), der dort die Entwicklungsgeschichte der Datenvisualisierung und den thematischen sowie rechtlichen Hintergrund der Daten erläutert.

Durch diese Zusammensetzung aus datengestützter sowie interaktiver Visualisierung in Kombination mit klassischer Berichterstattung und multimedia-

len Inhalten veranschaulicht dieses Data Special die Problematik des Themas wesentlich eindringlicher, als es in reiner Textform durch Artikel hätte vermittelt werden können (vgl. Langer 2011: 5). Die Rohdaten, die als Basis für die Datenvisualisierung verwendet wurden, stehen als Google-Spreadsheets im Sinne der Charakteristika des DDJ der Allgemeinheit frei zum Download zur weiteren Verwendung und Verarbeitung zur Verfügung (vgl. Biermann 2011). Für diesen Beitrag gewann das Projekt 2011 den Grimme Online Award in der Kategorie „Spezial“.

In die dritte Kategorie; Data App fallen Werke, die auf sehr großen, komplexen Datensätzen beruhen (gegebenenfalls werden auch mehrere maschinenlesbare Datensätze miteinander verbunden, sogenannte Mashups aus verschiedenen Datensätzen) oder jene, die sich an Echtzeitdaten bedienen. Solch ein Projekt zeichnet sich häufig durch eine Crowdsourcing-Einbindung aus. Die Umsetzung braucht von allen Erzählformen am meisten Zeit und kann auch in manchen Fällen als Jahresprojekt angesiedelt werden.

Ein Beispiel hierfür lieferte The Guardian im Frühjahr 2009, als die britische Tageszeitung 458.000 Dokumente zum britischen Spesensskandal der Unterhausabgeordneten zum Herunterladen ins World Wide Web stellte. Die Dokumente wurden zuvor vom Parlament veröffentlicht. Sie lagen jedoch teilweise in handschriftlicher Form vor und waren somit maschinell nicht auswertbar (Vgl. Matzat 2011a). Zudem fehlte es The Guardian an personellen Ressourcen, um den unstrukturierten Datensatz in Eigenarbeit zu analysieren. Er bat deshalb seine Nutzer im Rahmen des Projektes „Investigate your MP's Expenses“: „Können Sie uns mit der Überprüfung dieser Daten helfen?“ (vgl. Lisakowski 2012).

Um den Nutzer besser in die Medienanalyse einzubeziehen und ihm die Arbeit zu erleichtern, erstellte die Tageszeitung eine eigens dafür konzipierte Webapplikation zur Handhabung und Analyse der Datensätze. Die Reaktionen auf dieses Projekt waren laut dem mitwirkenden Journalisten Simon Rogers überwältigend (vgl. Bouchart 2012). Über 28.000 Internetnutzer halfen online mit und werteten fast 50 Prozent der Spesensskandal-Dokumente aus (vgl. guardian.co.uk 2009), um die Redaktion auf mögliche Verfehlungen der Politiker hinzuweisen (vgl. Matzat 2011a). Die daraus gewonnenen Erkenntnisse publizierte The Guardian sowohl im Print als auch auf seiner Website und deckte dadurch einen der größten Politikskandale Großbritanniens auf, der dafür sorgte, dass in Großbritannien noch im selben Jahr Rücktritte von Politikern und Gesetzesänderungen folgten (vgl. Oswald 2011).

6. DATENZWIEBEL - SYSTEMTHEORETISCHES MODELL DER DATENTYPEN

Grundlage jeglicher Berichterstattung und zentraler Gegenstand für alle datenjournalistischen Werke ist das Vorhandensein eines Datensatzes. So ist laut Simon Rogers, leitender Redakteur des Datablogs des britischen *The Guardian*, „der erste Schritt des Datenjournalisten [...], Daten überhaupt als interessante Quelle zu erkennen, aufzuspüren und „journalistisch urbar“ zu machen“ (Leßmöllmann 2012). Doch welche Daten sind in diesem Meer aus Zahlen für den DDJ relevant und lassen diese sich für bestimmte Zwecke sinnvoll kategorisieren? Zur Beantwortung dieser Fragen wird die sogenannte „Datenzwiebel“ von Andreas Grieb zur Hilfe genommen und für die vorliegende Thematik modifiziert. Im Hinblick auf den Datenjournalismus gibt es demnach drei zentrale Typen von Daten im Netz: Open Data, Transparent Data und Leaked Data, die im Weiteren näher erläutert werden.

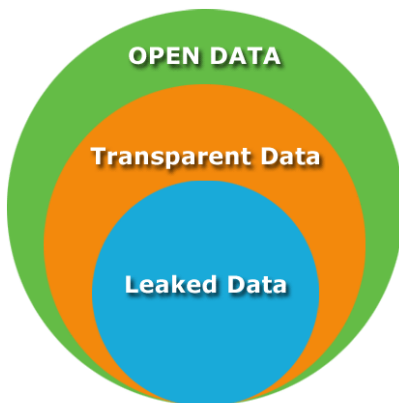


Abbildung 1: Die Datenzwiebel. Eigene Grafik nach der Datenquelle: Grieb 2011.

6.1. Das Prinzip Open Data

Unter „offenen Daten“ (englisch Open Data) werden laut Open Knowledge Foundation (vgl. Open Knowledge Foundation 2012) alle Datentypen subsumiert, die ohne jegliche Einschränkung zur Weiterverwendung und -bearbeitung für jeden frei zugänglich sind. Das Konzept umfasst dabei nicht nur die Daten der Regierung und der öffentlichen Verwaltung. Laut der Co:laboratory Internet & Gesellschaft geht es bei Open Data auch um Daten aus Bibliotheken und Archiven, von privatwirtschaftlich agierenden Unternehmen, der Wissenschaft sowie Non-Profit-Einrichtungen (Vgl. Internet & Gesellschaft Co:laboratory 2010: 27). Wichtig ist in diesem Kontext, dass es sich bei den Daten ausschließlich um nicht personenbezogene oder weitestgehend anonymisierte Daten handelt, also um Daten, die keinen Datenschutz- oder Sicherheitsbeschränkungen unterliegen. Das Open-Data-

Prinzip und der DDJ verfolgen somit beide denselben Grundgedanken im Gebrauch von Daten; Sie wollen mehr Transparenz durch den offenen Umgang mit Rohdaten und deren Bereitstellung als eine für alle frei zugängliche Informationsquelle. Open Data ist somit neben anderen Faktoren eine wichtige Voraussetzung für guten Datenjournalismus (Vgl. Kaute 2011).

6.2. Webscraping: Saugen aus dem Netz

„Da nicht alle Webseitenbetreiber offene Schnittstellen für ihre Inhalte anbieten, müssen Informationen extrahiert und in strukturierte Form gebracht werden“ (Fomin 2010: 17). Dabei kommt das sogenannte Screen-Scraping zum Einsatz. Screen-Scraping, zu Deutsch „Bildschirm auskratzen“, ist eine Methode zum automatischen Auslesen von Texten von Computerbildschirmen (vgl. Hoely 2011). In den meisten Fällen ist jedoch der Begriff Web Scraping geläufiger, da das Screen-Scraping fast ausschließlich in Bezug auf Webseiten gesetzt und im Folgenden als Synonym benutzt wird. In diesem Rahmen heißt scrapen (schaben, kratzen) das Auffinden und „Einsammeln“ von Informationen von Websites, um diese in einem maschinenlesbaren Datensatz zusammenzuführen (vgl. Bergemann 2010); folglich das Kopieren von Daten aus dem Netz. Dabei werden die auf Websites in Textform gelagerten Informationen durch entsprechende Software-Werkzeuge, Screen-Scraper oder Wrapper genannt, extrahiert und in einer Tabellen-Datei gespeichert (vgl. Fomin 2010: 17). So können sie anschließend weiteranalysiert und verarbeitet werden.

Diese von Websites extrahierten Daten fallen in die zweite Schicht der dargestellten Grafik der Datenzwiebel und werden als Transparent Data bezeichnet. Der Journalist Andreas Grieb erläutert, dass diese Transparent Data tendenziell auch „open“ sind, sich jedoch grundlegend in der Aufbereitung unterscheiden: „So liegt Transparent Data zwar frei zugänglich vor, jedoch nicht in Tabellen-Form und nicht in die gewünschte Relation gesetzt. Sie ist verstreut, nicht zusammengefasst, ihre Lizenzierung ist mitunter unklar“ (Grieb 2011). Als Beispiel dafür nennt er die Beziehungsstatistik auf Facebook. Sie sind laut Grieb prinzipiell zugänglich und anschaulich, aber sie liegen nicht als komplett vorhandener Datensatz vor, der über eine offene Schnittstelle heruntergeladen werden kann. So folgert er: „Du kannst sie erstellen, aber du müsstest sie erstellen“ (Grieb 2011).

6.3. Leaked Information

Das Wort Leaks wird aus dem Englischen abgeleitet und bedeutet so viel wie „Lecks“, „Löcher“ oder „undichte Stellen“. Der Vorgang, der zur Leaked Data führt, wird als Data Leakage bezeichnet. Data Leakage bedeutet die nicht autorisierte Übertragung von organisationsinternen Daten an ein externes Ziel

oder einen externen Empfänger (vgl. Gordon 2007: 5). Als Synonym für Data Leakage ist auch das Begriff Information Leakage geläufig.

Im Zusammenhang mit dem hier dargestellten DDJ spielen dabei Enthüllungsplattformen, die sich im World Wide Web gebildet haben, wie zum Beispiel Wikileaks und Open Leaks eine zentrale Rolle für die datengestützte Berichterstattung, da sie riesige Dokumentensätze, die nicht zur Veröffentlichung vorgesehen waren, von teilweise anonymen Whistleblowern zentral auf einer Plattform für journalistische Zwecke zur Verfügung stellen. Nach dem Journalisten Andreas Grieb ist Leaked Data eine „Mischform aus Closed Data und Open Data. Genauer: Sie ist Closed Data, die Open wurde“ (Grieb 2011). Auch mit dieser Prämisse verfolgt man mit Leaked Data grundsätzlich das gleiche Ziel wie bei Open Data, nämlich Daten zu öffnen, sie jedem frei zugänglich zu machen und die Datensätzen dabei in strukturierter, maschinenlesbarer Form oder über eine offene Schnittstelle (API) zur Weiterverarbeitung und Weiterverbreitung anzubieten (Vgl. Paderta 2012). Leaked Data dienen also in ihrem Grundsatz der Transparenz und Freiheit von digitalen Informationen. Sie sind jedoch auch Closed Data in dem Sinne, dass die Daten zwar für einen bestimmten Zweck erhoben wurden, sie jedoch unter Verschluss gehalten werden und ihre Publikation nicht für die Öffentlichkeit bestimmt ist. So entstehen bei der Verwendung von Leaked Data, genauso wie bei Transparent Data, einige Einschränkungen für den DDJ, besonders in Bezug auf ethische und rechtliche Fragen. Zudem sind ihre Vollständigkeit und ihr Kontext oft nicht sicher abzuklären. Laut Grieb ist eine Einordnung deshalb schwierig „und stellt sie immer unter Vorbehalt. Auch die Motivation des Whistleblowers und möglicher Zwischenkanäle färbt das Material“ (Grieb 2011).

Anm. d. A.: Im Hinblick auf diese drei Datentypen sollte auch immer auf Datenschutz und valide Quellen geachtet werden. Es sollte daher immer eine Datenkontrolle durch den Journalisten erfolgen. Darauf im Einzelnen einzugehen, würde jedoch den Rahmen dieses Aufsatzes sprengen.

7. DAS POTENTIAL DES DATENJOURNALISMUS

Die digitale Revolution als zentrales Element unserer heutigen Gesellschaft prägt gleichzeitig den Kampf um die Informationshierarchie im Netz: „Wer kann Informationen zurückhalten, wer hat die Deutungshoheit über das, was geschieht. Und wer kann seine Deutung mit größtmöglicher und weitreichendster Autorität verbreiten“ (Matzat 2010a). Lorenz meint, dass der Journalismus sich nicht vor der Auseinandersetzung mit Datenanalyse und -Auswertung scheuen darf und auf die Entwicklungen der Datenzeit rea-

gieren müsse. Die Antwort auf die neuen Herausforderungen, die an den Journalismus von morgen gestellt werden, ist der Datenjournalismus und das in ihm liegende Potential. „Datenjournalismus antwortet auf eine verdatete Welt, die ihre gesellschaftlichen Prozesse häufig in Datensätzen dokumentiert und ablegt - ein Fundus für relevante journalistische Geschichten“ (Leßmöllmann 2012). Eine datenjournalistische Berichterstattung kann somit neue Zusammenhänge und Fakten aufdecken, die sonst verborgen geblieben wären. Im Weiteren wird näher auf die einzelnen Punkte eingegangen, die das Potential, welches der Datenjournalismus im World Wide Web birgt, verdeutlichen.

7.1. Exklusivität und Unabhängigkeit

Nach Lorenz Matzats Ansicht (Lokalrundfunktag 2012: 2:10 bis 2:50 Minute) schöpft der Online-Journalismus bislang noch nicht sein volles Potential aus. „Die Website einer Zeitung imitiert irgendwie immer noch das Zeitungsblatt. Ich denke Daten bieten eine neue Chance in diesem Bereich, weil sie dem Trägermedium Internet entsprechen“ erläutert Matzat. „Das Internet kann etwas, was all die anderen Nebenformen nicht können: journalistische Geschichten non-linear und interaktiv erzählen“. Im DDJ liegt also die Kraft, das volle interaktive und multimediale Potential des Mediums Web zu nutzen. Journalistische Webpräsenzen könnten so gegenüber ihren gedruckten Mutterblättern ein Alleinstellungsmerkmal etablieren und ihre Stärken ausspielen (vgl. Buettner 2012), da sie eine völlig neue Form der Berichterstattung ermöglichen. Zudem wollen Rezipienten von journalistischen Angeboten einzigartige Inhalte und keine Nachrichten, die sich überall im Netz wieder finden lassen, meint Aron Pilhofer (Langer 2011: 7), Leiter der 14-köpfigen Abteilung Interactice News in der Redaktion der New York Times. Datenjournalistische Werke bieten dabei den Vorteil, einzigartig und in ihrer Form nicht leicht imitierbar zu sein: „Interaktive Datenanwendungen kann man nicht einfach kopieren. Ich kann nicht, wie ein Artikel oder Foto das einfach in den Zwischenspeicher meines Rechners legen und irgendwo anders einfügen, das ist nicht möglich“ (Lokalrundfunktag 2012: 18:00 bis 18:40 Minute). Ein journalistisches Medium, welches ein datenjournalistisches Werk publiziert, kann sich somit der Exklusivität seiner Geschichte sicher sein. Zudem sind datengestützte Berichterstattungen nachhaltig und aktualisierbar, da die Anwendungen auf Datenbanken basieren und auch nach deren Publikation mit weiteren Daten angereichert werden können. „Wenn im nächsten Jahr der Datensatz vom Vorjahr veröffentlicht wird, kann ich die Anwendung einfach damit ergänzen, die Anwendung ist somit nicht tot, sie kann je nach Thema immer weiter wachsen und zu-

gänglich bleiben“ (Lokalrundfunktag 2012: 18:40 bis 19:00 Minute). Dadurch schafft der Datenjournalismus in seiner Form und Ausprägung Exklusivität im Bereich des Journalismus und Mehrwert für den Online-Journalismus.

DDJ schafft weiterhin Unabhängigkeit in der Berichterstattung, da sich Journalisten nicht mehr ausschließlich auf die Datenauswertungen von externen Unternehmen, Behörden oder aus der Wissenschaft verlassen müssen. Datenjournalisten können laut Christina Elmer (Vgl. ZAPP 2011: 1:41 bis 2:00) durch die Fähigkeit, Rohdatensätze zu analysieren und aufzubereiten, selbst entscheiden, was den Kern eines Datensatzes ausmacht und dadurch exklusive journalistische Geschichten schaffen. Dadurch wird zugleich auch unabhängiges Recherchieren der Journalisten möglich. „Wenn man Rohdaten selber hat, ist man unabhängig von denen, die die Daten auswerten“ (Focke 2011). So kann der DDJ Zusammenhänge aufzeigen, die vorher nicht ersichtlich waren und in dieser Form nicht hätten erzählt werden können. Der Journalist ist nicht mehr auf Statistiken und Diagramme aus zweiter oder dritter Hand angewiesen und kann alle Daten, die er benötigt einfach selbst auswerten „was dem Journalismus in seinem Selbstverständnis als unabhängiger Vermittler durchaus entgegen kommen würde“ (Macromedia Hochschule für Medien und Kommunikation 2012).

7.2. Transparenz und Glaubwürdigkeit

Zu der konsequenten Umsetzung von datenjournalistischen Projekten gehört auch immer das Bereitstellen der dazugehörigen Rohdaten für den Nutzer (vgl. Langer 2011: 4), folglich eine vollkommene Offenlegung der Daten. Der Datenjournalismus übt somit eine bisher in dieser Form noch nicht existente Quellen-Offenheit im Journalismus aus, die erlaubt, dass Rezipienten die Fakten einer Geschichte selbst prüfen und nachvollziehen können. Lorenz Matzat erläutert dazu, dass Datenjournalismus ein Format sei, welches der heutigen, selbstbestimmten Medienutzung der Internetgesellschaft entgegen komme. DDJ reagiere auf Nutzer, die selbst Themen erkunden wollen und sich in einer möglichst kurzen Zeit, schnell und aus vielen Quellen Informationen beschaffen würden. „Für die ist ein normaler Text-Artikel nicht mehr das journalistische Format, welches sie konsumieren. Sie wollen selber aktiv werden und sich selbst mit der Thematik auseinandersetzen. Für solche Menschen ist eine interaktive Datenanwendung, zumindest aus journalistischer Sicht, sicherlich der richtige Weg“ (Lokalrundfunktag 2012: 19:00 bis 19:25). Dabei bewirkt die non-lineare Erzählform der DDJ-Werke, dass der Rezipient den Beitrag selbst erlebt und selbstständig erkunden kann: Dem Nutzer wird die Wahl gelassen, welche Ausschnitte eines datengestützten Werkes er rezi-

piert, wie tief er in die Materie einsteigt und welche Zusatzinformationen er abrufen kann. Der DDJ schafft dadurch Transparenz und baut gleichzeitig Vertrauen beim Rezipienten in die journalistische Berichterstattung und den Online-Journalismus auf.

Neben der Vertrauensbasis, die der DDJ zwischen Nutzern und Journalisten schafft, stärkt der DDJ auch die Glaubwürdigkeit in die Medien und den Journalismus. Durch die selbstständige Recherche der Daten und deren intensive Aufbereitung besinnt sich der DDJ auf den Ursprung des Qualitätsjournalismus zurück: Er konzentriert sich auf die Hintergrundberichterstattung, recherchiert sorgfältig, prüft Fakten und bereitet die Inhalte redaktionell-reflektiert auf (vgl. Thurner 2010). Web-Angebote mit datenjournalistischen Stücken kräftigen dadurch ihre eigene Rolle als Primärinformationsquelle sowie Kontrollorgan der Gesellschaft und können ihrem Angebot zu mehr Seriosität helfen. In Zeiten der „Gratiskultur“, in der Nutzer kostenlosen journalistischen Inhalt gewohnt sind und sich journalistische Online-Angebote von News-Portalen, Blogs, PR-Seiten oder Bürgerjournalismus abzugrenzen versuchen, könnte die durch die hohe, qualitative Berichterstattung des Datenjournalismus gewonnene Glaubwürdigkeit helfen, journalistische Web-Angebote finanziell ertragreich zu machen (vgl. Gehmlich 2012).

8. FAZIT UND AUSBLICK

„Datenjournalismus ist gekommen, um zu bleiben“ (OpenDataCity 2012).

Zusammenfassend bleibt festzuhalten, dass der DDJ ein immenses Potential für den Journalismus und vor allem für den Online-Journalismus in sich trägt: Er ermöglicht Exklusivität, bewirkt Unabhängigkeit, schafft Transparenz und fördert Glaubwürdigkeit. Datenjournalismus verbessert durch den professionellen Umgang mit Daten den Online-Journalismus und bietet die Möglichkeit, dass sich Journalisten an die aktuellen Entwicklungen des digitalen Zeitalters anpassen. Dabei kann Datenjournalismus nicht nur als vorübergehender Trend angesehen werden, sondern nimmt die Position einer neuen, journalistischen Avantgarde im World Wide Web ein, deren feste Etablierung in journalistischen Online-Angeboten in den nächsten Jahren zunehmend zu betrachten sein wird: Denn die Datenberge im World Wide Web werden nicht kleiner. Sie sind der Rohstoff, der den Journalismus beflügeln und ihm helfen wird, politische, soziale und wirtschaftliche Vorgänge auf eine neue, qualitativ-hochwertige und spektakuläre Weise abzubilden. Journalisten, die mit Daten umgehen können, sind bereits heute gefragte Experten (vgl. Langer 2011: 1).

9. LITERATURVERZEICHNIS

- Beckedahl, Markus 2012: Data Journalism Handbook. URL: <http://netzpolitik.org/2012/data-journalism-handbook/> [Stand: 30.04.2012; Abruf: 20.05.2012]
- Bensch, Danielle 2011: Die Menschheit erstickt an ihren Daten. URL: <http://www.welt.de/wirtschaft/webwelt/article12507803/Die-Menschheit-erstickt-an-ihren-Daten.html> [Stand: 18.06.2011; Abruf: 02.05.2012]
- Bergemann, Benjamin 2010: Datenjournalismus: Die zwei Seiten der Transparenz. URL: <http://politik-digital.de/datenjournalismus-die-zwei-seiten-der-transparenz/> [Stand: 30.11.2010; Abruf: 01.06.2012]
- Biermann, Kai 2011: URL: Datenschutz: Was Vorratsdaten über uns verraten. URL: <http://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz> [Stand: 24.02.2011; Abruf: 10.06.2012]
- Bouchart, Marianne 2012: Crowdsourcing Data at the Guardian Datablog. URL: http://datajournalismhandbook.org/1.0/en/getting_data_5.html [Stand: 30.04.2012; Abruf: 18.06.2012]
- Bradshaw, Paul 2010: How to be a data journalist. URL: <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide> [Stand: 01.10.2010; Abruf: 09.05.2012]
- Bradshaw, Paul/ Rohmaa, Liisa 2011: The Online Journalism Handbook. Skills To Survive And Thrive In The Digital Age. Harlow in England: Longman.
- Buettner, Mathias 2012: Geschichten aus 1001 Database. Daten sind Storys! URL: <http://blog.der13stock.de/2012/05/geschichten-aus-1001-database/>
- Davies, Dan: 2012: Mirko Lorenz - Data Journalism Event. URL: <http://de.slideshare.net/dandavies23/mirko-lorenz-data-journalism-event> [Stand: 19.09.2012; Abruf: 23.05.2013]
- Dean, Talisa 2012: Datenjournal: Online-Journalismus zwischen Daten und Multimedia URL: <http://www.datenjournal.de/dossier/einstieg> [Stand: N.N.; Abruf: 07.05.2012]
- Elmer, Christina 2012: Data Driven Journalism: Neue Erzählformen mit Datenjournalismus. URL: <http://www.bpb.de/system/files/pdf/9NGOGE.pdf> [PDF-Datei] [Stand: N.N.; Abruf: 09.05.2012]
- Focke, Jonathan 2011: Die Nachricht aus der Datenbank. URL: <http://www.derblindfleck.de/index.php/2011/07/06/die-nachricht-aus-der-datenbank/> [Stand: 06.07.2011; Abruf: 21.05.2012]
- Focus-Online 2012: Vorratsdatenspeicherung: Vorratsdatenspeicherung. URL: http://www.focus.de/politik/deutschland/antiterrorkampf/vorratsdatenspeicherung_aid_67215.html [Stand: N.N.; Abruf: 10.06.2012]
- Fomin, Alexander 2010: Semi-automatische Generierung von REST-Services aus Webseiten. URL: www.se.uni-hannover.de/pub/File/pdfpapers/Fomin2010.pdf [PDF-Datei] [Stand: 14.04.2010; Abruf: 01.06.2012]
- Forum Journalismus und Medien Wien 2012: Crowdsourcing - Die Weisheit der Massen nutzen. URL: <http://www.fjum-wien.at/wp/kurse/crowdsourcing-die-weisheit-der-massen-nutzen/> [Stand: N.N.; Abruf: 21.05.2012]
- Gehmlich, Pierre 2012: Schnell, Exklusiv, Transparent. URL: <http://www.onlinejournalismus.de/2012/02/27/schnell-exklusiv-und-transparent/> [Stand: N.N.; Abruf: 22.06.2012]
- Gordon, Peter 2007: Data Leakage - Threats and Mitigation. URL: http://www.sans.org/reading-room/whitepapers/awareness/data-leakage-threats-mitigation_1931 [PDF-Datei] [Stand: 15.10.2007; Abruf: 06.06.2012]
- Griß, Andreas 2011: Die Datenzwiebel. URL: <http://www.youdaz.com/2011/06/20/die-daten-zwiebel/> [Stand: 20.06.2011; Abruf: 28.05.2012]
- Grimme Online-Award 2012: Preisträger 2011. URL: <http://www.grimme-institut.de/html/index.php?id=1122#c8530> [Stand: N.N.; Abruf: 10.06.2012]
- Guardian.co.uk 2009: Investigate your MP's expenses. URL: <http://mps-expenses.guardian.co.uk/> [Stand: 2009; Abruf: 18.06.2012]
- Hoely, Daniel 2011: Scraping: Ist das Kopieren von Daten aus dem Netz erlaubt? URL: <http://blog.kooptech.de/2011/09/scraping-ist-das-kopieren-von-daten-aus-dem-netz-erlaubt/> [Stand: 08.09.2011; Abruf: 01.05.2012]
- Holovaty, Adrian 2006: A fundamental way newspaper sites need to change. URL: <http://www.holovaty.com/writing/fundamental-change/> [Stand: 06.09.2006; Abruf: 20.05.2012]
- IBM, Business Analytics & Optimization Insider 2011: Big Data: Datenregen oder Datensegen? URL: <http://www-01.ibm.com/software/de/insider/ausgaben/2011/2011-3-1a.html?key=Big%20Data> [PDF-Datei] [Stand: 01.03.2011; Abruf: 04.05.2012]
- Internet & Gesellschaft Co:laboratory 2010: „Offene Staatskunst“: Bessere Politik durch „Open Government“? URL: <http://bit.ly/Lofdlk> [PDF-Datei] [Stand: 04.10.2010; Abruf: 29.05.2012]
- Jones, Ben 2012: A Tribute To Fearless Felix. URL: <http://dataremixed.com/2012/10/a-tribute-to-fearless-felix/> [Stand: 15.10.2012; Abruf: 21.05.2013]
- Kaute, Sonja 2011: Datenjournalismus & Open Data: Überblick, Beispiele und Quellen für Journalisten. URL: <http://stift-und-blog.de/2011/11/datenjournalismus-opendata/> [Stand: 06.11.2011; Abruf: 21.05.2012]
- Konitzer, Michael 2011: Big Data. URL: <http://konitzer.wordpress.com/tag/information-overflow/> [Stand: 15.05.2011; Abruf: 02.05.2012]
- Krampe, Anna-Lena 2012a: Bestandsaufnahme: Datenjournalismus in Deutschland (Teil 1). URL: <http://datenschaetze.wordpress.com/2012/04/12/bestandsaufnahme-datenjournalismus-in-deutschland-teil-1/> [Stand: 12.04.2012; Abruf: 20.05.2012]
- Landrock, Holm / Zilch, Andreas 2012: Big Data: Eine neue Dimension für Informationen - Teil 3: Tempo und Nutzerzahlen. URL: <http://www.experton-group.de/research/ict-news-dach/news/article/big-data-eine-neue-dimension-fuer-informationen-teil-3-tempo-und-nutzerzahlen.html> [Stand: N.N.; Abruf: 06.05.2012]
- Langer, Ulrike 2011: Datenjournalismus. Journalisten Werkstatt. Salzburg-Eugendorf: Medienfachverlag Oberbauer.
- Leßmöllmann, Annette 2012: Datenjournalismus: Chance für den Journalismus von morgen. URL: <http://journalistik-journal.lookingintomedia.com/?p=843> [Stand: 29.03.2012; Abruf: 21.05.2012]
- Lisakowski, Vera 2012: Wozu Datenjournalismus? URL: <http://blog.grimme-online-award.de/2012/05/wozu-datenjournalismus/> [Stand: 24.05.2012; Abruf: 18.06.2012]
- Lokalrundfunktage: 2012: Lorenz Matzat, OpenData-City Berlin, bei der Local Web Conference 2012. URL: <http://www.youtube.com/watch?v=ZrV7VDNwjgQ> [Stand: 08.03.2012; Abruf: 17.05.2012]
- Macromedia Hochschule für Medien und Kommunikation 2012: Datenjournalismus ein wenig mehr journalistische Unabhängigkeit. URL: <http://journalismus-mhmk.de/medialab/2012/01/datenjournalismus-ein-wenig-mehr-journalistische-unabhaengigkeit/> [Stand: 27.01.2012; Abruf: 14.06.2012]
- Matzat, Lorenz 2010a: Data Driven Journalism: Versuch einer Definition. URL: <http://opendata-network.org/2010/04/data-driven-journalism-versuch-einer-definition/> [Stand: 13.04.2010; Abruf: 07.05.2012]
- Matzat, Lorenz 2010b: Datenjournalismus vor dem Internet: Wetterbericht, Finanzdaten und Co. URL: <http://www.datenjournalist.de/datenjournalismus-vor-dem-internet-wetterbericht-finanzdaten-und-co/> [Stand: 01.09.2010; Abruf: 21.05.2012]
- Matzat, Lorenz 2010c: Einführung in Datenjournalismus / Data Driven Journalism. URL: http://www.slideshare.net/wir_sie/einfuehrung-in-datenjournalismus-data-driven-journalism [Foliensatz] [Stand: 16.04.2010; Abruf: 09.05.2012]
- Matzat, Lorenz 2010d: Hyperlocal: Lokaljournalismus und OpenData in den USA und Deutschland. URL: <http://www.datenjournalist.de/hyperlocal-microlocal-usw-lokaljournalismus-und-opendata-in-den-usa-und-deutschland/> [Stand: 19.06.2010; Abruf: 17.06.2012]
- Matzat, Lorenz 2011a: Datenjournalismus: Neue Erzählformen. URL: <http://www.bpb.de/gesellschaft/medien/opendata/64069/datenjournalismus?p=2> [Stand: 26.10.2011; Abruf: 10.06.2012]
- Matzat, Lorenz 2011b: Daten sind der Treibstoff des Journalismus. URL: <http://blog.zeit.de/opendata/2011/05/04/daten-journalismus-textastribune/> [Stand: 04.05.2011; Abruf: 04.05.2012]
- Matzat Lorenz 2011c: Dossier Open Data: Datenjournalismus. URL: <http://www.bpb.de/gesellschaft/medien/opendata/64069/datenjournalismus> [Stand: 26.10.2011; Abruf: 07.05.2012]
- Medium Magazin 2011: „Man macht sich überprüfbar“. URL: <http://www.mediummagazin.de/archiv/2011-2/ausgabe-01022011/man-macht-sich-ueberpruefbar> [Stand: 08.01.2011; Abruf: 09.05.2012]
- OpenDataCity 2012: Wir sind zweimal für den Grimme Online Award 2012 nominiert. URL: <http://www.opendatacity.de/wir-sind-zweimal-fur-den-grimme-online-award-2012-nominiert/> [Stand: 08.05.2012; Abruf: 20.05.2012]
- Open Knowledge Foundation 2012: Was ist Open Data? URL: <http://opendatahandbook.org/de/what-is-open-data/index.html> [Stand: N.N.; Abruf: 29.05.2012]
- Oswald, Bernd 2011: Die Intelligenz des Schwarms. URL: <http://www.mediummagazin.de/archiv/2011-2/ausgabe-32011/die-intelligenz-des-schwarms/> [Stand: 18.03.2011; Abruf: 18.06.2012]
- Paderta, Damian 2012: OpenData - raumbezogene Daten. URL: <http://paderta.com/stuff/opendata-raumbezogene-daten-full/> [Stand: 01.04.2012; Abruf: 29.05.2012]
- Palmer, Michael 2006: Data is the New Oil. URL: http://ana.blogs.com/maestros/2006/11/data_is_the_new.html [Stand: 03.11.2006; Abruf: 04.05.2012]
- Pelzer, Claudia 2011: Crowdsourcing Terminologie. URL: <http://www.crowdsourcingblog.de/blog/2011/02/12/terminologie/> [Stand: 12.02.2011; Abruf: 21.05.2012]
- Plöschinger, Stefan 2012a: Geschichte des Zugmonitors: Decoder für den Bahn-Stau. URL: <http://www.sueddeutsche.de/reise/geschichte-des-zugmonitors-decoder-fuer-den-bahn-stau-1.1304468> [Stand: 09.03.2012; Abruf: 15.06.2012]
- Plöschinger Stefan 2012b: Wir speichern, Sie stöbern. URL: <http://www.sueddeutsche.de/kolumne/datenjournalismus-projekt-zugmonitor-wir-speichern-sie-stoebren-1.1304376> [Stand: 10.03.2012; Abruf: 17.05.2012]
- Schneider, Wolf/ Raue, Paul-Josef 52003 (1996): Das neue Handbuch des Journalismus. Reinbeck bei Hamburg: Rowohlt Verlag GmbH.
- Turner, Thomas 2010: Daten Journalismus - der Weg aus der Journalismuskrise <http://www.zukunftsweg.at/daten-journalismus-der-weg-aus-der-journalismuskrise> [Stand: 26.09.2010; Abruf: 22.06.2012]
- ZAPP 2011: Recherche: Datenjournalismus. URL: <http://www.youtube.com/watch?v=poUa8Wu9eTk> [Video] [Stand: 16.11.2011; Abruf: 07.05.2012]
- Zeit.de 2011: Vorratsdatenspeicherung: Verätherisches Handy. URL: <http://www.zeit.de/datenschutz/malte-spitz-vorratsdaten> [Stand: 10.02.2011; Abruf: 10.06.2012]