



22.02.2012 | JOURNALISMUS & TECHNOLOGIE

Change – Wie Daten den Journalismus verändern

Die Veröffentlichung der [Kriegstagebücher](#) der US-Armee aus Afghanistan und dem Irak waren Sternstunden des Datenjournalismus. Einer neuen Form journalistischen Erzählens. Hunderttausende Dokumente wurden im Jahr 2010 durch die Internetplattform [WikiLeaks](#) veröffentlicht. Allerdings waren die Dokumente in den seltensten Fällen in erzählerischem Ton verfasst, der von Journalisten in größeren Stories hätte zitiert werden können. Vielfach waren es kurze technische Informationen über Vorgänge, Zwischenfälle, Anschläge inklusive präziser Orts- und Zeitangaben. Erst die Methoden des Datenjournalismus, die Durchdringung der Datenmengen mit Statistikwerkzeugen, ihre Veranschaulichung durch Grafiken oder anhand interaktiver Landkarten machten das Rohmaterial plötzlich journalistisch urbar und schufen neue Formen publizistischen Erzählens. Zu den wichtigsten Entwicklungsorten dieses neuartigen Storytellings gehört das [Data Blog](#) des britischen [Guardian](#). [Simon Rogers](#) ist für das Data Blog verantwortlich.

von Simon Rogers

Der Journalismus hat sich seit meinen Anfangstagen derart gewandelt, dass er kaum wiederzuerkennen ist.

Bei meiner ersten Anstellung zu Beginn der 90er Jahre, für eine wöchentlich erscheinende juristische Zeitschrift, war ich der einzige Reporter mit einem Computer. Die Recherche geschah nicht im Internet, sondern in Mikrofiche-Archiven und über Telefon. Unsere Reporter tippten ihre Texte auf Schreibmaschinen und druckten sie auf Papierabschnitte, die die Designer im Layout zusammensetzten. Wir lernten, wie man Bilder mit Lineal und Bleistift abmisst und wie man per Telefon einem Copytaker Texte diktiert. Erst Ende der 90er, als ich selbst für eine Webseite verantwortlich war, hatte ich einen festen Internetzugang und meine erste berufliche E-Mail-Adresse.

Diese Reise in die Vergangenheit soll nicht zeigen, wie alt ich schon bin, sondern wie viel sich in so kurzer Zeit grundlegend geändert hat. Unsere Leser waren weit weg, sie waren stumm. Sie wurden von uns beliefert und nahmen es dankend entgegen. Wir hatten die Verantwortung.

Und dies hat sich geändert. Die Nachrichtendienstleister heutiger Tage haben kaum etwas mit den damaligen Strukturen gemein. In Zeiten ständiger finanzieller Unsicherheit sind wir in einer neuen Rolle. Es sind nicht mehr wir, die entscheiden; stattdessen kämpfen wir verzweifelt um jedes Augenpaar, das auf unserer Webseite und nicht auf der der Konkurrenz liest.

Die Chancen des Ganzen sind aber enorm. Wer kann sich noch vorstellen, einen Artikel zu schreiben, ohne einen Internetzugang zu haben, ohne Twitter oder Facebook zu nutzen? Oder ohne Nachrichtensender? Unsere Vorgänger mussten sich auf Tickermeldungen verlassen – uns hingegen steht die ganze Welt auf Tastendruck offen.

Angesichts ständiger Klagen der Redakteure und Autoren über die heutigen unsicheren Verhältnisse fällt es schwer zu glauben, aber: Wir befinden uns in einer Zeit unglaublicher Möglichkeiten.

Und die größten Chancen sind unerwarteterweise: Daten.

Die WikiLeaks-Enthüllungen zu Afghanistan, Irak und den Depeschen der US-Botschaften, die Spesenaffäre im britischen Parlament, die globale Rezession, sogar die Schweinegrippe-Hysterie – die neue Art und Weise der Berichterstattung zu diesen Ereignissen wurde wohl nur durch Reporter ermöglicht, die keine Angst vor Mathematik hatten, die sich mit Tabellenkalkulationen auskannten, Daten anschaulich aufbereiten konnten und – ganz entscheidend – die richtigen Fragen stellten.

Was ist Datenjournalismus? Es ist ein Merkmal der neuen öffentlichen Transparenz, die sich weltweit bemerkbar macht.

Es begann in den USA; weitere Länder zogen nach: Australien, Neuseeland, und das Vereinigte Königreich mit dem Datenportal data.gov.uk, mitentwickelt von Sir Tim Berners-Lee, dem Erfinder des World Wide Web.



Data

Apps

Consultation

Forum

Blogs

Ideas

Linked Data

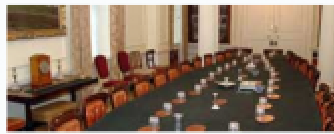
Resources

About

Go

Enter your postcode to find tools and information about your local area using government data

Who does what in Whitehall?



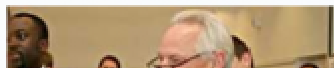
How your money is spent



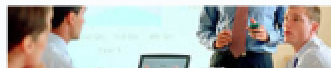
Government contracts in full



Who ministers are meeting



Business plans



Applications



Bislang vertrauten wir für unsere Informationen stets den offiziellen Kanälen. Es war uns egal, ob die Daten schon vorgefiltert und aufbereitet waren. Aber das Vertrauen zu Regierungen ist geschwunden, das Vertrauen zu den traditionellen Medienkanälen noch mehr. Wir streben danach, die Zahlen hinter den Nachrichten selbst zu kennen – um herauszufinden, ob uns wirklich die Wahrheit aufgetischt wird, und um unsere eigenen Nachrichten schreiben zu können.

Dadurch können wir einen Journalismus betreiben, der über Worte oder über Grafiken funktioniert – oder über beides. Manchmal kann eine Story allein durch die Darstellung einer Tabelle erzählt werden.

Vieles davon ist nicht neu – es ist nur einfacher geworden.

Das, was Regierungen bisher anboten, waren de facto Portale für Sammlungen von Daten – Daten, die sowieso schon verfügbar waren.

Aber es gibt auch neue Informationen. Im Vereinigten Königreich beschloss die 2010 gewählte Regierungskoalition im Rahmen ihres Programms für mehr Transparenz einen „Tsunami der Daten“. Riesige Datenbestände wurden der Öffentlichkeit zugänglich gemacht: alle Regierungsausgaben über £ 25.000, die Bezüge höherer Beamter, detaillierte Daten zu Ausgaben aus der Staatskasse, straßengenaue Kriminalstatistiken und die Leistungen einzelner Krankenhäuser bei der Bekämpfung von multiresistenten Erregern wie MRSA.

Die ersten Gehversuche des Guardian im Datenjournalismus kamen zur gleichen Zeit wie die Veröffentlichung unseres Open Platform API. Das [Datablog](#) – für das ich verantwortlich bin – war als

kleines Blog geplant, in dem es um die unseren Artikeln zugrundeliegenden Datensätze geht. Jetzt besteht es aus einer [Startseite](#), einer Suchmaschine für Daten zu Regierungen und zur weltweiten Entwicklung, Datenvisualisierungen u. a. von Grafikern des Guardian sowie Tools zum Forschen nach Daten zu öffentlichen Ausgaben.

Die [Programmierschnittstelle](#) (API) des Guardian ermöglicht es Benutzern, Inhalte strukturiert abzufragen und darauf basierend Anwendungen zu erstellen. Im Datastore finden Google-Tabellen Verwendung, für die es ebenfalls ein API gibt. Diese können Benutzer mit unseren Tabellen kombinieren und so auf Grundlage diverser Datenbestände Mashup-Anwendungen erstellen.

Für mich als Nachrichtenredakteur und Journalist, der mit Grafiken arbeitet, stellte dies eine logische Erweiterung meiner Arbeit dar. Die täglich neu hinzukommenden Daten mussten verarbeitet werden, um in den Stories des Tages zu resultieren.

Nach und nach trug das Datablog immer mehr zu den neuen Stories bei und gab die Hintergründe wieder. Über Crowdsourcing gelang es uns, 458.000 Dokumente zum britischen Spesenskandal zu analysieren und Einzelheiten zu den Aussagen der Abgeordneten herauszufinden. Wir gaben unseren Benutzern die Möglichkeit, die Datenbank mit den Ausgaben aus der Staatskasse einzusehen – wir veröffentlichten die Daten hinter den Meldungen.

Der wirkliche Anschlag für den Datenjournalismus kam aber 2010 mit einer großen Tabelle: 92.201 Datenzeilen, jede davon enthielt Detailangaben zu einem militärischen Vorgang in Afghanistan. Die WikiLeaks-Kriegstagebücher. Allerdings erst Kapitel Eins.

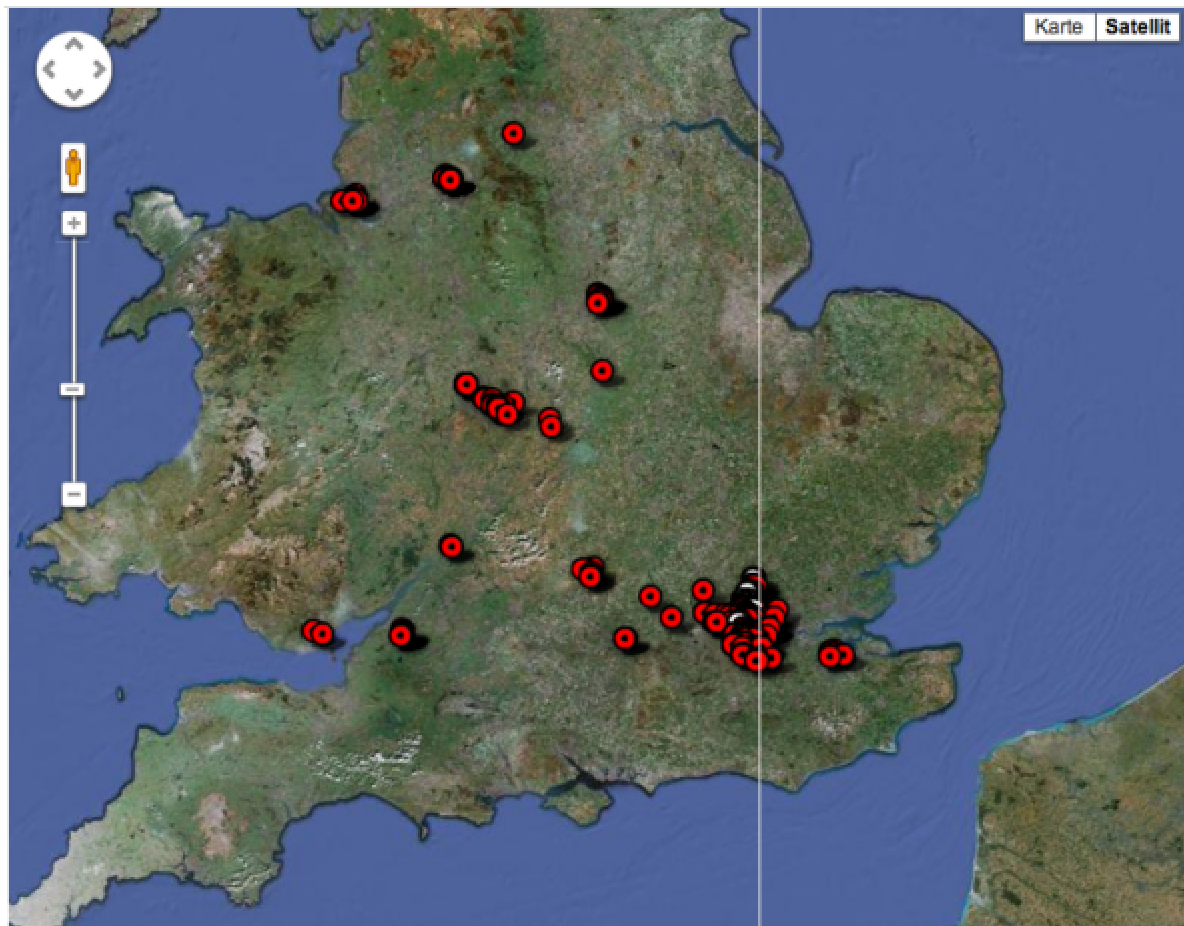
Zwei weitere Kapitel sollten noch folgen: Irak und die Depeschen. Der offizielle Name der ersten beiden Teile lautete SIGACTS: die „Significant Actions Database“ der US-Streitkräfte. Aus diesen Informationen hätten ohne den Datenjournalismus niemals Nachrichten werden können.

Aber schon zuvor hatten wir große Datenmengen zu bewältigen. Die veröffentlichte Datenbank mit den Ausgaben aus der Staatskasse (COINS – Combined On-line Information System) war noch größer und überwältigender. Die letzte Regierung wollte sie nicht veröffentlichen – die neue Koalition gab sie innerhalb weniger Wochen frei. Jede einzelne öffentliche Ausgabe aus allen Ressorts war aufgelistet, seit dem Jahr 2005. Millionen von Datensätzen, enorm komplex und schwer zu bearbeiten.

Wir begannen, die Daten zu filtern, um eins der wesentlichen Details dieses Krieges ans Licht zu bringen: die zunehmenden Angriffe mit USBVs (Unkonventionellen Spreng- oder Brandvorrichtungen) – das sind selbstgebaute Bomben, die unberechenbar und schwer abzuwehren sind. Auch diese Datenmenge war riesig – aber einfacher zu bewältigen. Zwischen 2004 und 2009 gab es ca. 7500 USBV-Explosionen bzw. -Hinterhaltangriffe (dabei wird die USBV mit anderen Waffen kombiniert, z. B. Handfeuerwaffen oder Panzergranaten). Weitere 8000 USBVs wurden entdeckt und geräumt. Wir wollten herausfinden, wie die Entwicklung im Laufe der Jahre war – und Vergleiche anstellen. Die Daten ergaben, dass der Süden, in dem britische und kanadische Truppen stationiert waren, am schlimmsten betroffen war – was die Berichte unserer Reporter vom Krieg bestätigte.

Der Datenjournalismus war live. Das Ziel war, durch diese Informationen unserem Reporterteam vor Ort eine Berichterstattung über die Menschen zu ermöglichen – und durch die Auswertung der Daten ein Gesamtbild zu zeichnen, in dem der wahre Verlauf des Krieges sichtbar wird.

Dann folgten die Unruhen in England. Ausgelöst durch die Erschießung eines jungen Mannes in London breiteten sich in fünf Tagen im August 2011 gewalttätige Unruhen von Tottenham über die gesamte Hauptstadt bis hin in den Norden Englands, nach Manchester und Liverpool, aus.



Für den Guardian bedeutete dies direkten Datenjournalismus. Zuerst ging es darum, die zu versorgen, die wissen wollten, was geschah. Wir erstellten eine [Liste](#) mit allen Vorfällen, für die es einen bestätigten Bericht gab, und visualisierten diese mit Google Fusion Tables auf einer [Landkarte](#). Die zugrundeliegenden Daten stellten wir zum freien Download zur Verfügung – vermutlich die einfachste, aber unsere erfolgreichste Idee.

Als die Unruhen sich abgeschwächt hatten und immer mehr Personen festgenommen und angeklagt wurden, kam es zu Hunderten von Gerichtsverhandlungen wegen Straftaten im Zusammenhang mit den Unruhen. Jeden Tag wurden in den Gerichten detaillierte Protokolle angefertigt und nach Abschluss der Verhandlungen in Auszügen den Gerichtsreportern zur Verfügung gestellt.

Diese waren, wenn man nicht manuell vorgehen wollte, besonders schwierig zu verarbeiten – wir wollten die Ortsangaben und die Ergebnisse der Verhandlungen herausfiltern. Es ging um die Daten von buchstäblich Tausenden von Angeklagten, also machten wir weiter. Bald hatten wir über 1000 Fälle zusammen – eine beispiellose Kriminalstatistik; ein einmaliger Datenbestand.

Eine schnelle Analyse der 1000 Fälle belegte, ebenso wie eine der ersten Detailauswertungen der Daten, was viele Anwälte befürchtet hatten: Die Gefängnisstrafen, zu denen Teilnehmer an den Unruhen verurteilt wurden, fielen im Schnitt 25 % länger aus als üblich. Mit weiteren analysierten Fällen zeigte sich, dass selbst dieser drastische Wert noch übertroffen wurde.

Reading the Riots

Investigating England's summer of disorder

In partnership with the London School of Economics
Supported by the Joseph Rowntree Foundation and the Open Society Foundations

A study of the causes of the English riots



Rioters say anger with police fuelled summer unrest

Guardian-LSE study of riots – involving hundreds of interviews with participants – reveals deep antipathy towards officers

1044 comments

- Riots were 'a sort of revenge' against the police
- The four-day truce among gangs during the riots
- Cameron and the Queen - rioters' sense of injustice
- Don't blame our parents, say rioters



A fire lit in Tottenham that burned Manchester: the rioters' story

Rioters from across England describe how it felt to be caught up in the chaos that engulfed the country last summer

- A consumerist feast amid the summer riots
- The women who rioted
- Were the riots about race?
- The morality of rioters
- A new kind of riot? From 1981 to 2011



How riot rumours spread on Twitter

Analysis of 2.6 million tweets shows Twitter is adept at correcting misinformation - particularly if the claim is that a tiger is on the loose in Primrose Hill

- Twitter and the riots: how the news spread
- How Twitter was used during the riots
- BBM was rioters' communication method of choice
- Riots in their own words - interactive
- The England 'riot commute' mapped: an animation
- Why did the riots end?

Auf der Grundlage dieses Datenbestandes bauten wir das Projekt „Reading the Riots“ auf – eine einzigartige Sicht auf die Unruhen, aus der Perspektive der Menschen, die vor Ort waren. Das Projekt stand unter der Leitung von Paul Lewis vom Guardian und wurde mit der London School of Economics und einem speziell geschulten Team von Interviewern durchgeführt.

Nicht zuletzt konnten wir durch den Zugriff auf 2,57 Millionen Twitter-Nachrichten mit Bezug auf die Unruhen erstmalig herausfinden, welche Rolle die neuen Medien wirklich spielten.

Es zeigte sich, wie stark Twitter für eine große Bürgerbewegung, die Koordinierung der Aufräumarbeiten, eingesetzt wurde. Mehr als 206.000 Twitter-Nachrichten – 8 % des Gesamtvolumens – hatten zum Thema, wie die Verwüstungen von vier Nächten voller Gewalt und Plünderungen beseitigt werden können.

In Anbetracht der Tatsache, dass der Datenjournalismus traditionell eine langwierige Angelegenheit ist, dessen Ergebnisse nur am Rande Aufmerksamkeit finden, zeigten die Unruhen einen neuen Umgang mit Daten auf. Der direkte Datenjournalismus unterstützt schon da, wo Nachrichten noch am aktuellsten sind.

Dies sind aber immer noch erst zwei Beispiele, zwei Schritte auf dem Weg des Datenjournalismus zur breiten Akzeptanz. Viele Reporter halten das alles noch für Unsinn und wollen nichts von Mathematik oder Tabellenkalkulationen wissen. Andere hingegen sehen diese neue Welle als geeignet an, den Journalismus zu retten.

Eine neue Rolle für Journalisten als Brücke zwischen denen, die die Daten haben (und es nicht hinkriegen, sie verständlich zu machen), und der Öffentlichkeit, die nach Antworten fragt, die Daten einsehen und verstehen möchte, dabei aber Unterstützung benötigt. Diese Brücke, das können wir sein.

Vielleicht ist es in Zukunft sogar unsere Aufgabe, Daten als vertrauenswürdige Quelle bereitzustellen, als „sicherer Hafen“ wertvoller Informationen.

Oft hört man, das Internet zerstöre den Journalismus. Der Datenjournalismus verbindet beides: das traditionelle journalistische Handwerk und die Stärken der Technologie – und dadurch kann eine spannende Story entstehen.

Momentan findet man Datenjournalismus vor allem auf Newsseiten und in Blogs. Aber funktioniert der Datenjournalismus nur rein visuell? Oder wäre er nicht auch im Radio denkbar? Doch, davon bin ich überzeugt. Beim Datenjournalismus geht es nicht explizit um hübsche Grafiken oder Visualisierungen. Es geht darum, Storys auf Grundlage von Daten und Zahlen zu entwickeln. Das ist doch sozusagen reiner Journalismus. Ob visuelle Story oder investigative Berichterstattung – die Art der Präsentation ist fast schon zweitrangig, das Format ist hochflexibel.

In ein paar Jahren ist der Datenjournalismus vielleicht schon keine Neuerung mehr; heute ist er es noch. Die Welt ändert sich, und sie ändert sich durch Daten.