

DIE STORY IM DATENHAUFEN

von Jan Lukas Strozyk

Verräterische Klimaanlagen, zweideutige Stellenanzeigen, millionenschwere Bauaufträge: Wer richtig sucht, findet auch in Hunderttausenden von Daten eine Geschichte.

Es war eine einzelne Zeile im schier endlosen Datenstrom. Ein Auftrag, den wir eigentlich gar nicht gesucht hatten. Auf einmal sahen wir: »Procurement Identifier: FA945107C0196; Contracting Office Agency Name: Dept. of the Air Force; Vendor Name: University Stuttgart, Product/Service Code AC53; Description: Weapons (Advanced)«. Dazu eine paar weitere komplizierte Zeichenketten und am Ende ein Betrag: 38.000 US-Dollar.

Übersetzt steht dort, dass die Universität Stuttgart Waffenforschung für die Luftwaffe der USA betrieben hat. Für uns war diese Zeile der Anlass, bei zahlreichen deutschen Hochschulen anzuklopfen, weitere Datenberge zu durchforsten und letztlich eine Debatte um militärische Gelder im deutschen Forschungsbetrieb zu entfachen.

Unser Projekt »Geheimer Krieg« lebte stets von klassischer Recherche: rausgehen, mit den Menschen vor Ort sprechen, Kontakte pflegen, hier und da etwas gesteckt bekommen. Ohne diese Art der journalistischen Arbeit hätte es den Film nicht gegeben, und auch die Beiträge in der *Süddeutschen Zeitung* und auf der Website www.geheimerkrieg.de wären nicht erschienen.

Gleichzeitig haben wir aber versucht, unsere Thesen systematisch mit Daten zu belegen und aus den Daten neue Ansätze zu gewinnen – wie im genannten Beispiel. Dabei half uns eine amerikanische Attitüde: In den USA ist es selbstverständlich, dass die Regierung sich gegenüber dem Bürger

(insbesondere dem Steuerzahler) rechtfertigt. Und das heißt vor allen Dingen: Ausgaben erklärt. Im Jahr 1966 haben die USA den sogenannten »Freedom of Information Act« verabschiedet, schon 20 Jahre zuvor den »Administrative Proce-

Mit der Datenbank FPDS lässt sich ein detailliertes Bild der US-Staatsausgaben in unserem Land zeichnen.

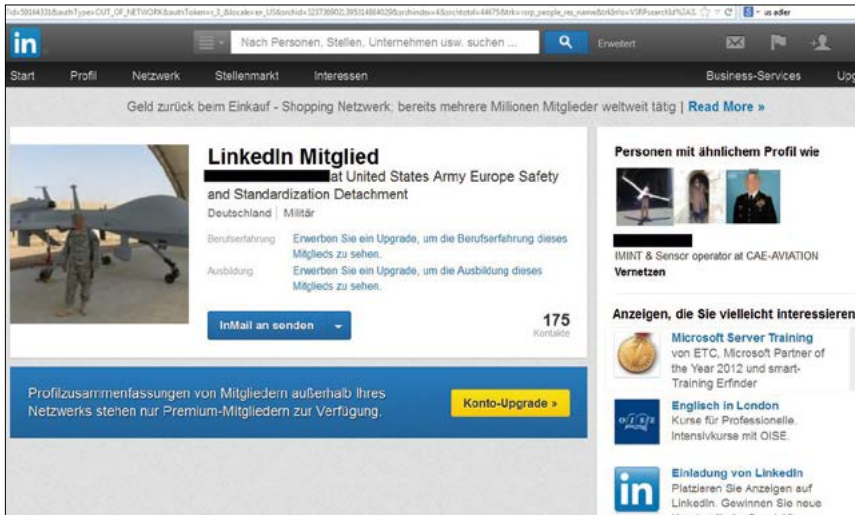
dure Act« – zwei von einer ganzen Reihe von Gesetzen, die US-Bürgern und zum Teil auch Ausländern direkten Zugriff auf Informationen und Daten im Besitz des Staates geben und somit eine transparentere Demokratie ermöglichen sollen.

Ein Weg, sich über die Staatsausgaben zu informieren, ist die Datenbank des »Federal Procurement Data System«, kurz FPDS. Hinter dem kryptischen Namen verbirgt sich eine Sammelstelle für Aufträge aus öffentlicher Hand – fast jeder Dollar, den die US-Regierung bezahlt, findet sich hier wieder. Abgesehen von einigen Ausnahmen, wie Kleinaufträgen unter 3.000 US-Dollar oder als geheim eingestufte Verträge, findet sich hier ein Eintrag für alles, was die US-Regierung sich und ihrem Land leistet. Wir interessierten uns naturgemäß vor allem für Deutschland. Und schnell wurde klar: Mit den Daten aus FPDS lässt sich ein detailliertes Bild der US-Staatsausgaben in unserem Land zeichnen, der größte Teil davon Ausgaben von Militär und Sicherheitseinrichtungen.

Mit Hilfe der Programmierer von der Berliner Firma OpenDataCity konnten wir die Datenbank des FPDS für unsere Zwecke »scrapen«, also eine lokale Kopie erzeugen. Dazu haben wir



Datenbanken belegen, in welchem Umfang das US-Militär die Logistik für Drohneinsätze aufbaut.



Auf ihren Profifotos in Online-Berufsnetzwerken posieren US-Soldaten ganz offen vor Drohnen.

zunächst die komplette Datenbank in einer Volltextsuche nach relevanten Einträgen durchsucht. Relevant bedeutet in dem Fall, dass eines der von uns festgelegten Suchwörter darin vorkommt, zum Beispiel »Darmstadt« oder »Stuttgart«. Durch unsere Recherchen wussten wir zudem von einigen wichtigen Militäreinheiten in Deutschland und ihren Abkürzungen. So landete zum Beispiel der Code »66MI« für die 66. Military Intelligence Brigade auf der Liste unserer Suchbegriffe. Die so erzeugte Kopie ließ sich deutlich zügiger und komfortabler bearbeiten, als einzelne Abfragen an den FPDS-Server es je zugelassen hätten.

RUND 42 AUFTRÄGE PRO TAG

Das Ergebnis hat uns alle überrascht: Über 150.000 Einträge hat die Liste der Aufträge und Auftragsänderungen der vergangenen zehn Jahre – allein für Deutschland. Das heißt, dass die US-Regierung jeden Tag im Schnitt rund 42 Aufträge abschließt, die einen Bezug zur Bundesrepublik haben. Allein 2012 hat die US-Regierung in Deutschland drei Milliarden US-Dollar ausgegeben – nur nach Afghanistan floss mehr Geld. Freilich weckt nicht jeder Eintrag sofort das journalistische Interesse. Wer hat schon Lust, einer Toilettenpapier-Bestellung der Kasernen in Heidelberg hinterherzurecherchieren? Schnell stellt sich also in Anbetracht einer derartigen Masse an Daten die Frage, wie man daraus die relevanten Informationen ableitet.

Um die Daten-Spreu vom Daten-Weizen zu trennen, benutzen wir unter anderem das Programm »OpenRefine«. Die Software ist open-source-lizenziert und frei verfügbar und eignet sich auch für Anfänger gut, um Daten-Durcheinander zu sortieren. An dieser Stelle sei erwähnt, dass die hier vorgestellte Recherche kein »Best-Practice«-Beispiel sein soll. Vielmehr haben wir uns Schritt für Schritt herangetastet. Bestimmt

neue, präzisere Listen erstellen und durchstöbern: Nur Aufträge mit Bezug zum US Africa Command, nur Bauaufträge, nur Aufträge mit einem Wert von mindestens einer Million US-Dollar und so weiter. Die so herausgearbeiteten Daten nutzten wir als Grundlage für weitere Recherchen und um unsere eigenen Thesen zu überprüfen. Und letztlich auch, um unsere Leser zur Recherche zu animieren. Dazu haben wir die Daten bereinigt und auf der Website www.geheimerkrieg.de zur Verfügung gestellt. So kann jeder Nutzer selbst auf die Suche gehen: Arbeitet mein Arbeitgeber mit den Amerikanern zusammen? Gibt es in meiner Heimatregion einen wichtigen Stützpunkt?

Natürlich macht eine Datenbank noch keine Geschichte. Aber ein paar Zeilen unter Hunderttausenden können manchmal das entscheidende Puzzleteil sein, und sie zu finden, hat den Reiz ausgemacht. Ähnlich wie wir mit den FPDS-Daten verfahren sind, haben wir weitere Quellen systematisch durchsucht und ausgewertet: Job-Börsen und Business-Netzwerke wie Monster, LinkedIn oder Xing und eine weitere Auftragsdatenbank namens »Federal Business Opportunities«. Dazu immer wieder Telefonate, Treffen, E-Mails. Am Ende ist »Geheimer Krieg« für uns beides gewesen: ein klassisches Rechercheprojekt und ein erfolgreiches datenjournalistisches Experiment. ■



Jan Lukas Strozzyk ist freier Mitarbeiter im Team Recherche beim NDR.

Foto: Screenshot/LinkedIn; Martin Jäschke

»WIE EIN NEUER KONTINENT«

Ein Interview mit dem Datenjournalisten Lorenz Matzat über seine Mitarbeit am Projekt »Geheimer Krieg«

Herr Matzat, Sie haben mit den Firmen OpenDataCity und Lokaler die technische Infrastruktur für das Projekt »Geheimer Krieg« zur Verfügung gestellt. Was waren dabei die Herausforderungen?

Matzat: Die Kunst bei einer Anwendung im Netz ist es, dass alles reibungslos läuft, sich schlüssig und elegant bedienen lässt. Dem Betrachter darf die Anwendung nicht im Wege stehen. Zum Beispiel war es eine Herausforderung, die Kartenflüge von einem Ort zum anderen zu ermöglichen, dafür müssen viele Kartendaten zum Gerät des einzelnen Betrachters geladen werden. Der Aufwand, sich mit zwei Medienhäusern zu koordinieren, darf auch nicht unterschätzt werden.

Das Kernelement von www.geheimerkrieg.de ist eine interaktive Karte. Welche Vorteile hat diese Art der Navigation aus Ihrer Sicht, welche Nachteile?

Es ist unmöglich, in einer Anwendung allen Aspekten eines Sachverhalts gerecht zu werden. Es lässt sich auch darüber streiten, ob für die Geschichten beim Projekt »Geheimer Krieg« wirklich eine Karte vonnöten war. Letztlich war es eine dramaturgische Entscheidung. In der Konzeption des Webprojekts haben wir uns in mehreren Runden zusammen mit den Beteiligten von NDR und *Süddeutscher Zeitung* darauf festgelegt: Es ging uns darum, zeigen zu können, an welchen Orten der »Geheimer Krieg« stattfindet. Der Vorteil an Karten ist, dass sie einiges an Kontext in sich tragen, weil die meisten Leute Karten lesen können und mit vielen Orten direkt Informationen verbinden. Zudem erlauben interaktive Karten im Netz eine non-lineare Navigation durch eine Geschichte und per Zoom den fixen Wechsel von Makro- zu Mikroperspektive.

Für die Website www.geheimerkrieg.de und die Recherche zum Projekt wurden zwei Datenbanken der US-Regierung ausgelesen und aufbereitet. Wie funktioniert so etwas?

Wir versuchen den Aufbau einer Datenbank zu verstehen, bevor wir sie mit einem maßangefertigten Programmschnipsel »scrapen«, also ausschaben; das kann bei komplexen Datenbanken sehr aufwändig sein, aber auch erstaunlich simpel, wenn sie bestimmten Standards folgt. Ansatzpunkt ist oft, wie Datenbanken ihre Ergebnisse an den Webbrowser ausspielen – irgendwie müssen die Antworten ja auf eine Anfrage hin von der Datenbank auf den Bildschirm gelangen. An dieser Stelle lassen sich die Daten mitlesen, oder man kann zumindest das Verfahren identifizieren. Danach geht es darum, sie nach einem vorher festgelegten Modell strukturiert in eine eigene Datenbank auszulesen. Über die hat man dann die volle Kontrolle.

Warum sind eigentlich ausgerechnet Daten im Moment journalistisch so interessant? Datenjournalismus ist ja ein relativ modernes Phänomen.

Die fortschreitende Digitalisierung steigert den Datenbestand rasant. Es ist wie ein neuer Kontinent, der aus dem Meer steigt. Den müssen Journalisten besuchen und mit den ihm gerechten Mitteln darüber berichten.

Was muss ein Journalist heute mitbringen, wenn er mit Daten arbeiten will?

Ohne einen gewissen Nerdfaktor wird es schwer. Man sollte zumindest die Grundlagen von Statistik verstehen und bereit sein, sich in Teamarbeit mit digitalen Werkzeugen auseinanderzusetzen. Es schadet auch nicht, sich zumindest etwas mit einer gängigen

Programmiersprache, aktuellen Webtechnologien und Datenbanken auszukennen. Der Einstieg ist aber nicht schwer. Viele Tools sind kostenfrei, und es gibt eine riesige Auswahl von Tutorials und anderen Hilfen.

Die Fragen stellte Jan Lukas Stroyzk.

Zu Hunderttausenden listen die USA ihre Millionenaufträge auf – auch für die Militärbasis Ramstein.

2003-10-01	\$ 11,772,040.00	DEPT OF DEFENSE	W S A GMBH WORLD SERVICE AL		KAISERSLAUTERN	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 6,995,375.00	DEPT OF DEFENSE	DRS TRAINING&CONTROL SYSTEMS	32548	FORT WALTON BE	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 4,862,973.00	DEPT OF DEFENSE	PAE GOVERNMENT SERVICES, INC	90017	LOS ANGELES	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 3,536,345.00	DEPT OF DEFENSE	PAE GOVERNMENT SERVICES, INC	90017	LOS ANGELES	469 ABS/LGC Rhein Main AB, Gf
2003-10-01	\$ 3,256,872.00	DEPT OF DEFENSE	LOCKHEED MARTIN CORPORATION	89115	LAS VEGAS	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 3,096,579.00	DEPT OF DEFENSE	ITT FEDERAL SERVICES INTERNATI	80935	COLORADO SPRIN	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 1,616,192.00	DEPT OF DEFENSE	LOCKHEED MARTIN CORPORATION	89115	LAS VEGAS	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 1,325,584.00	DEPT OF DEFENSE	A KLEES GESELLSCHAFT MIT BESC		NEUNKIRCHEN	USAFE CONS/LGC ROB, Germany,
2003-10-01	\$ 1,298,056.00	DEPT OF DEFENSE	PIEPENBROCK	67663	KAISERSLAUTERN	USAFE CONS/LGC ROB, Germany,