



CHRISTIAN
RUDDER
INSIDE
BIG
DATA

UNSERE DATEN ZEIGEN,
WER WIR WIRKLICH SIND.

HANSER

weiß Google auch noch, ob Sie schwul oder wütend oder einsam oder rassistisch sind oder sich Sorgen machen, ob Ihre Mutter womöglich Krebs hat. Twitter, Reddit, Tumblr, Instagram – all diese Firmen sind natürlich in erster Linie profitorientierte Unternehmen, aber gleich danach erheben sie auch demografische Daten bisher unbekannter Reichweite, Vollständigkeit und Bedeutsamkeit. Praktisch nebenbei zeigen uns digitale Daten jetzt realistisch, wie wir Menschen kämpfen, wie wir lieben, wie wir altern, wer wir sind und wie wir uns verändern. Man muss nur hinschauen: Aus ganz kurzer Distanz zeigen uns die Daten, wie sich Leute verhalten, wenn sie glauben, unbeobachtet zu sein. Im Folgenden zeige ich

Ihnen, was ich gesehen habe. Und Deosprays sind mir dabei wirklich egal.



Wenn Sie Erfahrung mit allgemein verständlichen Sachbüchern haben, dann werden Sie an Dataclysm vielleicht dies und das ungewöhnlich finden. Erstens die Verwendung der Farbe Rot. Zweitens, dass es darin um Menschengruppen und große Zahlen geht, woraus sich die seltsame Erscheinung ergibt, dass in einem Buch über Menschen kaum welche vorkommen: Einzelpersonen werden so gut wie nicht erwähnt. Es gibt viele Grafiken, Statistiken und Tabellen, aber kaum Namen. Es ist ein beliebtes Stilmittel der Populärwissenschaft, große Ereignisse anhand von etwas Kleinem und Kuriossem zu schildern – die Geschichte der Welt aus der

Sicht der Speiserübe, ein Fisch als Kriegsgrund, ein ganzer Regenbogen an der Zimmerwand, wenn man eine Taschenlampe im genau richtigen Winkel auf ein Prisma richtet. Ich mache es genau umgekehrt. Ich nehme etwas Großes – eine ungeheure Datenmenge über Handlungen, Gedanken und Äußerungen von Menschen, viele Terabytes an Daten – und filtere viele kleine Fakten heraus: Was das Netzwerk Ihrer Freunde über die Stabilität Ihrer Ehe denkt, wie Asiaten (oder Weiße, Schwarze und Latinos) sich am seltensten selbst beschreiben, wo und warum Schwule Ihre Neigung nicht öffentlich leben, wie sich die Schreibgewohnheiten in den vergangenen zehn Jahren geändert haben, und wie sich Wutausbrüche nicht geändert haben. Das Prinzip ist, unser Selbstverständnis auf

Zahlen statt auf Beschreibungen zu gründen, oder vielleicht eher die Zahlen als Beschreibungen zu verstehen.

Dieser Ansatz entstand in den Schlackehalden der Statistik. *Dataclysm* ist eine Erweiterung der Arbeit meiner Kollegen und meiner selbst aus den letzten Jahren. Eine Internetpartnervermittlung möchte Menschen zusammenbringen, und um das seriös zuwege zu bringen, muss sie die Wünsche, Gewohnheiten und Abneigungen ihrer Nutzer kennen. Also macht man sich fleißig daran, einen Haufen detaillierter Daten zu sammeln und sie in allgemeine Theorien menschlichen Verhaltens zu verwandeln. Wer mit solchen Informationen arbeitet anstatt, zum Beispiel, für die Hochzeitsredaktion der Sonntagszeitung,

bekommt mit der Zeit ein besonderes Verwandtschaftsgefühl zum bunt gemischten Durcheinander der Menschheit als Ganzes und nicht zu zwei Einzelpersonen. Man fängt an, die Menschen so zu begreifen, wie ein Chemiker vielleicht die umherwirbelnden Moleküle seiner Tinktur begreift und zu lieben beginnt.

Dennoch müssen natürlich alle Webseiten – und alle Informatiker – ihre Daten formalisieren. Mit allem, was keine Zahl ist, kann ein Algorithmus nichts anfangen; wenn ein Computer also etwas begreifen soll, muss man es ihm so weit wie möglich in Zahlen übersetzen. Webseiten und Apps stehen damit vor der Herausforderung, das Kontinuum menschlichen Verhaltens in kleine Schubladen mit den Etiketten 1, 2, 3