

## Rezension zu:

Petersson, Gustav Jakob/Breul, Jonathan D. (Hg.): *Cyber Society, Big Data, and Evaluation. Comparative Policy Evaluation. Volume 24.* New Brunswick: Transaction Publishers, 2017. 263 Seiten, 32,95 \$, ISBN 9781412864510

*Elisabeth Nindl*<sup>1</sup>

Im Rahmen der Frühjahrstagung 2018 des AK Methoden zum Thema „Potenziale und Grenzen von Big Data in Evaluationen“ wurden verschiedene Sichtweisen von Big Data und deren Anwendungsmöglichkeiten erörtert. Bei diesem Anlass wurde das Buch „Cyber Society, Big Data, and Evaluation“ vorgestellt, welches die Möglichkeiten von Big Data in der Schaffung von evidenzbasiertem Wissen für Entscheidungsprozesse untersucht. Die Publikation setzt sich zum Ziel, den prominenten aber sehr weitläufigen Begriff ‚Big Data‘ zu konkretisieren und aufzuzeigen, inwiefern Big Data derzeit in Monitoring und Evaluation bereits eingesetzt wird, in welchen Bereichen Potenzial besteht, und welche Hürden dabei auftreten können.

Der von Gustav Jakob Peterson und Jonathan D. Breul herausgegebene Band umfasst 13 Beiträge, von welchen sich die ersten zwei der Definition von Big Data widmen (Gustav Jakob Petersson, Frans Leeuw, Jonathan Breul und H.B.M. Leeuw, sowie Frédéric Lefebvre-Naré, Sebastian Lemire und Gustav Jakob Petersson). Die digitale, vernetzte Welt produziert kontinuierlich immense Datenmengen – die Verknüpfung verschiedener Arten von Daten und Datenquellen zu einem Datensatz wird dann als ‚Big Data‘ bezeichnet. Dies ist bei weitem nicht die einzig mögliche Definition, erscheint jedoch als greifbar für potenzielle Nutzer(innen), da sie im Gegensatz zu den berühmten (mindestens) ‚3 V‘, Volume, Variety und Velocity, nicht selbstreferenziell ist. Besonders instruktiv ist die Unterscheidung nach der Datenstruktur, da sie die zur Verarbeitung notwendigen Methoden verdeutlicht und somit die technischen Anforderungen aufzeigt (S. 29f.). Klassische strukturierte Daten im Querschnitt oder Panel, Zeitreihen

und un- bzw. semistrukturierte Daten wie Text, Bilder, Geodaten, Netzwerkdaten uvm. erfordern verschiedene Arten der Speicherung sowie ein breites Set an Analysemethoden, verbunden mit der notwendigen Software und Anwenderwissen.

Big-Data-Analysen werden bisher weder von Evaluator(inn)en in nennenswertem Ausmaß eingesetzt (Steven Højlund, Karol Olejniczak, Gustav Jakob Peterson und Jakub Rok), noch wird dies von Auftraggebern erwünscht oder ermöglicht. Kim Forss und Jonas Norén stellen in einer Metaanalyse der Terms of Reference von 25 Evaluationen aus dem Bereich der Entwicklungszusammenarbeit fest, dass diese meist sehr explizit sind hinsichtlich Design, Methoden und Datenquellen und damit wenig Raum lassen für z.B. Big-Data-Anwendungen, auch wenn es sinnvolle Einsatzmöglichkeiten gegeben hätte.

Mehrere Beiträge präsentieren Beispiele für den potenziellen Einsatz von Big Data in Evaluierungen im weitesten Sinn, zeigen dabei jedoch auf, dass das Verständnis und die Nutzung noch sehr eingeschränkt sind – dies scheint auch der Breite beider zentralen Begriffe, ‚Big Data‘ und ‚Evaluation‘, zuzuschreiben zu sein.

So gibt der vierte Beitrag von Maria Barados und Jonathan I. Mitchell einen Überblick über Teile des kanadischen Gesundheitssystems und einiger darin verfügbarer Daten. Die Autor(inn)en schlussfolgern, dass die Wirkungsanalyse der Akkreditierung von Krankenhäusern anhand von Big-Data-Analysen von Data Scientists durchgeführt werden könnte. Der siebte Beitrag von Jonathan D. Breul beschreibt die Vielzahl an Informations- und Datenquellen, die bei den Sicherheitsvorkehrungen zum Super Bowl 2014 eingesetzt wurden. Die Evaluation des Sicherheitskonzepts nach der Sportveran-

1 KMU Forschung Austria, Wien

staltung selbst wird dann jedoch anhand von Interviews mit verschiedenen Repräsentant(inn)en durchgeführt und kann somit auch keinen Bezug zu Big Data herstellen. Im achten Beitrag von Peter Wilkins wird die Verwendung von Big Data zur Verkehrsplanung beschrieben. Auch wenn die Kombination von GPS-Daten mit Postings in sozialen Netzwerken (z.B. zu Unfällen oder dem Straßenzustand) zum Management von Verkehrsflüssen innovativ sein mag, schafft es auch dieser Beitrag nicht, einen Konnex zu Evaluation herzustellen.

Die Verwendung von online generierten Daten zur Untersuchung einer Policy-Maßnahme untersucht H.B.M. Leeuw. Er beschreibt eine Wirkungsanalyse einer Maßnahme gegen illegale Downloads in den USA, das Copyright Alert System (sechster Beitrag). Da es sich hierbei um eine e-Intervention handelt, erscheint es naheliegend, das Online-Nutzerverhalten als Datenbasis zu verwenden. Inwiefern allerdings die Betrachtung der Veränderungen der durchschnittlichen Anzahl an Suchanfragen vor, während und nach der Intervention von Daten aus 115 Abfragen bei Google Trends als Big-Data-Anwendung bezeichnet werden kann, sei dahingestellt. Auch der Autor selbst merkt an, dass dieses Screening von Nutzerverhalten nur eine Ausgangsbasis für bzw. eine Komponente von einer umfassenderen (empirischen) Evaluation sein kann.

Francesco Mazzeo Rinaldi, Giovanni Giuffrida und Tom Negrete präsentieren eine Ex-ante-Evaluation einer Politikmaßnahme (elfter Beitrag). Anhand von Online-Kommentaren zu Zeitungsartikeln über ein potenzielles Verbot von Einweg-Plastiktaschen in Kalifornien soll die Einstellung der Menschen bezüglich dieser Maßnahme eruiert werden. Dazu werden 28 Artikel einer Zeitung zu diesem Thema ausgewählt, die durchschnittliche Anzahl der Leser(innen) ausgewertet und mit jener von anderen Berichten verglichen. In weiterer Folge wurden die Anzahl der Kommentare pro Artikel berechnet, die als Wahrscheinlichkeiten eines User-Kommentars interpretiert wird. Die insgesamt siebzehn Kommentare werden schließlich anhand eines speziellen Opinion-Mining-Algorithmus auf ihre Einstellung zum Thema (Sentimentanalyse) ausgewertet. Die überschaubare Anzahl erlaubt es den Autoren, alle Kommentare zu lesen und mit den Ergebnissen der automatisierten Analyse zu vergleichen. Hier zeigt sich, wie schwierig es für Algorithmen ist, geschriebenen Text richtig zu interpretieren. Algorithmenbasierte Sentimentanalysen auf Basis von Text sind somit kritisch zu sehen.

Schließlich beinhaltet der Band auch einen Beitrag, der eine sehr aufschlussreiche Big-Data-Anwendung vorstellt und anhand des konkreten Beispiels erforderliche Rahmenbedingungen aufzeigt. Steffen Bohni Nielsen, Nicolaj Ejler und Maryanne Schretzman beschreiben ein äußerst umfangreiches Vorhaben des New Yorker Center for Innovation through Data Intelligence (CIDI). In den meisten Ländern ist es schwierig oder verboten, personalisierte Daten aus verschiedenen Quellen zu verbinden und für kommerzielle oder politische bzw. administrative Zwecke zu nutzen. Das CIDI allerdings wurde ermächtigt, personenbezogene Daten von verschiedenen Behörden bzw. Agenturen (u.a. Gesundheit, Sozialwesen) zusammenzufügen, um individuelle Verhaltensmuster zu identifizieren und anhand eines Frühwarnsystems Interventionen zu gestalten. Menschen werden im Zeitverlauf beobachtet, um zu untersuchen, in welchen Kontexten und Verläufen z.B. Obdachlosigkeit oder Teenager-Schwangerschaften auftreten. Durch das Zusammenführen der Daten von fünf verschiedenen Agenturen (und die geografische Zuordnung via Google Maps) geht die Evaluation über die Wirkungsanalyse der Leistungen einer einzelnen Agentur hinaus und erlaubt die Analyse der Wirkungen aus dem Zusammenspiel der verschiedenen Angebote und Maßnahmen. Die sehr aufschlussreiche tabellarische Gegenüberstellung der Vorgehensweise in einer ‚klassischen‘ Evaluation mit dem CIDI-Big-Data-Ansatz (S. 163f.) verdeutlicht, inwiefern sich die Methoden ähneln und ergänzen, und dass die strukturellen Anforderungen für die Implementierung eines Big-Data-Systems für Monitoring und Evaluation beachtlich sind. Dies zeigt, dass Big-Data-Methoden, nicht wie in anderen Kapiteln wiederholt konstatiert, per se eine kostengünstigere Alternative zu klassischem Monitoring und Evaluationen darstellen. Die Autor(inn)en schlussfolgern, dass sich beide Ansätze gegenseitig anreichern können – dazu müssen sich Evaluator(inn)en jedoch sukzessive die notwendigen analytischen Kompetenzen aneignen.

Die dringend notwendige, kritische Reflektion des Big-Data-Hypes eröffnen Kim Forss und Jonas Norén. Dazu gliedern sie Big Data nach den Datenquellen in vier Bereiche (S. 177f.): 1) Active Driven Data: unstrukturierete Daten, die absichtlich von Nutzer(inne)n in spezifischen Systemen gespeichert werden; 2) Passive Driven Data: unstrukturierte Daten, die von Nutzer(inne)n unabsichtlich in einem System gespeichert werden, wie z.B. Suchabfragen oder Geolokationsdaten; 3) von Algorithmen generierte unstrukturierte Daten, die z.B. in

der Maschine-Maschine-Kommunikation entstehen und aufgezeichnet werden, und 4) öffentliche Statistiken, also strukturierte Daten, die aktiv in Datenbanken eingetragen werden. Die Daten aus den ersten drei Bereichen sind zumeist im Besitz von privaten Unternehmen und daher für Evaluator(inn)en nur schwer zugänglich, zudem ist der Datenzugang meist gesetzlich eingeschränkt und reguliert. Google oder Twitter beispielsweise machen ihre Daten nur in Teilen der Öffentlichkeit zugänglich, außerdem bestehen Unklarheiten hinsichtlich der Datenabdeckung und der Algorithmen, die die Daten generieren. Somit ist es zweifelhaft, in welchem Ausmaß solche Daten verlässliche Quellen darstellen können, die tatsächliche soziale Vorgänge reflektieren. Wenn die Entstehung und Abdeckung der Daten nicht transparent ist, dann ist es für Evaluator(inn)en nur schwer, deren Verwendung zu argumentieren. Solche Bedenken mögen bei Sentimentanalysen für politische Initiativen auf Basis von sozialen Netzwerken oder Kommentaren von Zeitungsanalysen zweitrangig sein, bei der Ex-ante-Evaluation eines Sicherheitskonzepts oder in Evaluationen von Programmen oder Agenturen ist dies aber keinesfalls vertretbar.

Forss und Norén ziehen daraus einen wesentlichen Schluss: Derzeit gibt es weder aufgrund von Datenverfügbarkeit und -qualität, noch von Auftraggeberseite Anreize, Big Data in Evaluationen einzusetzen, stattdessen vielmehr dafür, ebendies nicht zu tun. Gleichzeitig bedingen etablierte Vorgehensweisen und eingespielte Routinen in der Umsetzung von Evaluationen das Festhalten an bestehenden Abläufen. Damit Big-Data-Anwendungen als reguläre Methoden in Evaluationen einfließen können, muss dies von den Auftraggebern gewünscht werden, Evaluators(inn)en die notwendigen analytischen Fähigkeiten erwerben, die Datennutzung auf institutioneller Ebene geregelt werden und die Transparenz bei den Eigentümern der Daten sichergestellt sein, so dass die Daten auf ihre Qualität und Verlässlichkeit überprüft werden können.

Ähnlich gestaltet sich auch der Beitrag von Sebastian Lemire und Gustav Jakob Petersson, in welchem die sogenannten vier fundamentalen Veränderungen, die das revolutionäre Potenzial von Big Data ausmachen, kritisch hinterfragt werden: 1) datenbasierte statt theoriebasierte Wissensproduktion, 2) Korrelation statt Kausalität, 3) Populationen statt Stichproben und 4) „messy data“ statt saubere Daten. Im Einklang mit den Autoren muss auch ich den ersten drei Punkten vehement widersprechen. Zuerst benötigt jede Analyse einen logisch kon-

sistenten theoretischen Rahmen, dies gilt umso mehr, je größer die verfügbaren Datenmengen sind. Es erscheint als Ausflucht der Denkfaulen, die logische Argumentation gegen Algorithmen zur Datenanalyse einzutauschen; auch Lemire und Petersson halten fest „data never speak for themselves. Data are inanimate, silent.“ (S. 218). Die Absage von Punkt 1 führt direkt zur rigorosen Ablehnung der Aussage, dass Korrelationen Kausalität ersetzen können. Es bleibt zu hoffen, dass Evaluators(inn)en nicht diesem alten und durch unzählige Beispiele widerlegten Denkfehler unterliegen. Der dritte Aspekt, dass Big Data den Fokus auf gesamte Populationen statt auf Stichproben erlaubt, klingt verlockend, ist aber ebenso kritisch zu hinterfragen. Liegt der Fokus einer Evaluation beispielsweise auf der zukünftigen Nutzung eines Programms, so ist auch eine bestehende, ‚gesamte Population‘ nur eine Stichprobe einer zukünftigen Population. Zudem sind z.B. auch Daten von Nutzer(inne)n sozialer Netzwerke nicht mehr als Stichproben, die zudem starken Selektionseffekten unterliegen oder Fake-Profile enthalten können. Der letzte Punkt betrifft im Wesentlichen zwei Aspekte, die auch Forss und Norén ansprechen, nämlich dass Daten „anfallen“, d.h. ohne Zweck gesammelt werden und deren Entstehung oft unklar ist. Des Weiteren ist es derzeit auch noch unklar, wie mit der Dynamik in den Daten umgegangen werden kann und was die Implikationen daraus sind.

Aufgrund wesentlicher ungeklärter Aspekte bezüglich der Entstehung und Abdeckung der Daten, Selektionseffekte, datenschutzrechtlicher Bedenken und der fehlenden Transparenz bei den Eigentümern der Daten (zumeist multinationale Konzerne), halte ich es für unrealistisch, dass die prophezeiten Quantensprünge tatsächlich so eintreten werden. Dass die Verfügbarkeit von Daten in Echtzeit und die großen Stichproben (‚Populationen‘) statistisch validere und genauere Aussagen ermöglichen, muss aufgrund von Bedenken zu Selektionseffekten und der Güte statistischer Tests in sehr großen Stichproben relativiert werden. Die Möglichkeit, menschliches Verhalten vielfältig zu erfassen und damit die Relevanz von Evaluationen zu erhöhen, ist nicht nur durch gesetzliche Restriktionen hinsichtlich personenbezogener Daten schwierig, Daten aus sozialen Netzwerken sind auch wegen Fake-Profilen, Bots und den Selektionsprozessen insgesamt zu hinterfragen. Kostenersparnisse sind, wie am Beispiel von CIDI klar wird, nicht unbedingt zu erfahren, da auch das Zusammenführen verschiedener Datenquellen kostenintensiv ist.

Nach der Lektüre des Buches stelle ich fest, dass gar nicht so viel neu ist an Big Data, son-

dern dass ganz ähnlich gelagerte Fragen auftreten wie bei klassischen, strukturierten Daten. Schlussendlich sollte gelten, dass evidenzbasiertes Wissen stets auf einem möglichst breiten, soliden Fundament begründet sein muss. Big-Da-

ta-Anwendungen können und sollen sicherlich einige Aspekte in Evaluationen komplementieren und vielleicht auch substituieren. Dies muss allerdings mit der nötigen Nüchternheit geschehen, die kritisches Hinterfragen erlaubt.