Big Data. Zur Datenkonstruktion der Großdatenforschung

Von Ramón Reichert

Nr. 44 – 29.12.2014

Einleitung

Das Schlagwort “Big Data” ist in aller Munde – und beschreibt nicht nur die Entstehung einer neuen Informationswissenschaft, sondern steht auch für die Annahme, dass die Entwicklung der spätmodernen Gesellschaften einerseits von der zunehmenden Verbreitung digitaler Kommunikations- und Vernetzungskulturen und andererseits von der Nutzung von Großdaten abhängig ist. Die digitalen Medien und Technologien verbreiten nicht einfach nur neutrale Inhalte, sondern schaffen im Social Web lebensweltliche Kommunikationsräume und können daher als neue Selbstverständigungsdiskurse digitaler Gesellschaften angesehen werden. Obwohl in den öffentlichen Debatten der Big-Data-Ansatz oft mit Erwartungen digitaler Transparenz und Objektivität gleichgesetzt wird, kann die *Großdatenforschung* kein digitales Fenster zur sozialen Welt in Aussicht stellen. In diesem Sinne setzt sich der folgende Aufsatz 1) mit den methodologischen Einschränkungen der technisch-medialen Dispositive auseinander, berücksichtigt davon ausgehend 2) die Reflexivität der Nutzungspraktiken und kann somit 3) die fiktionalen, inszenatorischen und narrativen Dimensionen der Dateninterpretation konkretisieren, um 4) die maßgeblichen Aspekte der *Datenkonstruktion in der Großdatenforschung zu sondieren, die dazu beitragen sollen,* einen einseitigen Technikdeterminismus zu verabschieden.

Die Erforschung der Sozialen Netzwerke vermittels der Text-, Sediment-, Netzwerk- und Bildanalysen hat bisher aufgezeigt, dass Transaktionsdaten (Cookies, Logdateien, GPS-Daten, Bankdaten, Kundennummern) und Kommunikationsdaten (nutzergenerierte Inhalte) mittlerweile die nachgefragtesten Datenquellen zur Beschaffung und Anwendung von Regierungs- und Kontrollwissen darstellen. Dieses Wissen der Großdatenforschung ist aber ungleich verteilt und nicht öffentlich zugänglich. Die sich dabei verändernden Selbstverständnisse, wie auch die lokalen und globalen Erwartungen an Wissenschaftskulturen und Epistemologien bewirken, dass die der Big-Data-Research zugrundeliegenden fächerübergreifenden Praxisorientierungen eine nuancierte Genealogie, Datenkritik und Medienreflexion der datenintensiven Formen der Wissensproduktion erfordern.

Theorien und Methoden

Gegenwärtig werden in allen Bereichen der digitalen Internetkommunikation große Datenmengen (Big Data) generiert: "More business and government agencies are discovering the strategic uses of large databases. And as all these systems begin to interconnect with each other and as powerful new software tools and techniques are invented to analyze the data for valuable inferences, a radically new kind of 'knowledge infrastructure' is materializing." (Bollier 2010: 3) Seit dem späten 20. Jahrhundert zählen die digitale Großforschung und ihre großen Rechnerzentren und Serverfarmen zu den zentralen Bausteinen der Herstellung, Verarbeitung und Verwaltung von informatischem Wissen. Damit einhergehend rücken mediale Technologien der Datenerfassung und -verarbeitung und Medien, die ein Wissen in Möglichkeitsräumen entwerfen, in die Mitte der Wissensproduktion und der sozialen Kontrolle. In ihrer Einleitung in das "Routledge Handbook of Surveillance Studies" knüpfen die Herausgeber Kirstie Ball, Kevin Haggerty und David Lyon einen Zusammenhang zwischen technologischer und sozialer Kontrolle auf der Grundlage der Verfügbarkeit großer Datenmengen: "Computers with the Power to handle huge datasets, or 'big data', detailed satellite imaging and biometrics are just some of the technologies that now allow us to watch others in greater depth, breadth and immediacy." (2012: 2) In diesem Sinne kann man sowohl von *datenbasierten* als auch von *datengesteuerten* Wissenschaften sprechen, da die Wissensproduktion von der Verfügbarkeit computertechnologischer Infrastrukturen und der Ausbildung von digitalen Anwendungen und Methoden abhängig geworden ist.

Damit einhergehend haben sich auch maßgeblich die Erwartungen an die Wissenschaft des 21. Jahrhunderts verändert und werden zunehmend Forderungen laut, die darauf bestehen, die historisch, kulturell und sozial einflussreichen Aspekte der digitalen Datenpraktiken systematisch aufzuarbeiten – verknüpft mit dem Ziel, diese in den künftigen Wissenschaftskulturen und Epistemologien der Datenerzeugung und -analyse zu verankern. In diesem Sinne kann eingeräumt werden, dass die Repräsentation und die popularisierende Vermittlung der datengenerierten Forschung auch Anschlüsse auf frühere materielle Datenkulturen eröffnen, denen sowohl historische Kontinuitäten als auch mediale Umbrüche inhärent sind und die nur dann verständlich werden, wenn sie im historischen, sozialen und kulturellen Kontext reflektiert werden können (vgl. Gitelman und Pingree, 2004). Eine vergleichende Analyse der Datenverarbeitung unter Berücksichtigung der materiellen Kultur von Datenpraktiken vom 19. bis zum 21. Jahrhundert vermag aufzuzeigen, dass bereits im 19. Jahrhundert die mechanischen Datenpraktiken das taxonomische Erkenntnisinteresse der Forscher maßgeblich beeinflussten – lange bevor es computerbasierte Methoden der Datenerhebung gab (vgl. Driscoll, 2012). Weiterführende Untersuchungen erarbeiten die sozialen und politischen Bedingungen und Auswirkungen des

Übergangs von der mechanischen Datenauszählung der ersten Volkszählungen um 1890 über die elektronischen Datenverarbeitungen der 1950er Jahre bis zum digitalen Social Monitoring der unmittelbaren Gegenwart Im Vorfeld haben sich aber zahlreiche andere Disziplinen und nichtphilologische Bereiche herausgebildet, wie die Literatur-, Bibliotheks- und Archivwissenschaften, die eine längere Wissensgeschichte im Feld der philologischen Case Studies und der praktischen Informationswissenschaft aufweisen und sich seit dem Aufkommen der Lochkartenmethode mit quantitativen und informatikwissenschaftlichen Verfahren für wissensverwaltende Einrichtungen befassten. (vgl. Driscoll, 2012 6f.) Das Sammeln großer Datenmengen ist aber auch in eine Machtgeschichte der datenbasierten Episteme selbst verstrickt. (vgl. Leistert/Röhle 2011) An der Schnittstelle von konzernorientierten Geschäftsmodellen und gouvernementalem Handeln experimentieren Biotechnologie, Gesundheitsprognostik, Arbeits- und Finanzwissenschaften, Risiko- und Trendforschung in ihren Social Media-Analysen und Webanalysen mit Vorhersagemodellen von Trends, Meinungsbildern, Stimmungen oder kollektivem Verhalten.

In der Ära der Big Data hat sich der Stellenwert von sozialen Netzwerken radikal geändert, denn sie figurieren zunehmend als gigantische Datensammler für die Beobachtungsanordnungen sozialstatistischen Wissens und als Leitbild normalisierender Praktiken. Als Schlagwort steht Big Data für die Überlagerung eines statistisch fundierten Kontrollwissens mit einer medientechnologisch fundierten *Makroorientierung* an der ökonomischen Verwertbarkeit von Daten und Informationen. Die großen Datenmengen werden in verschiedenartigen Wissensfeldern gesammelt: Biotechnologie, Genomforschung, Arbeits- und Finanzwissenschaften, Meinungs- und Trendforschung berufen sich in ihren Arbeiten und Studien auf die Ergebnisse der Informationsverarbeitung der Big Data und formulieren auf dieser Grundlage aussagekräftige Modelle über den gegenwärtigen Status und die künftige Entwicklung von sozialen Gruppen und Gesellschaften. Die Big Data-Research hat sich innerhalb der letzten Jahre erheblich ausdifferenziert: Zahlreichen Studien betreiben mithilfe maschinenbasierter Verfahren wie der Textanalyse (quantitative Linguistik), der Sentimentanalyse (Stimmungserkennung), der sozialen Netzwerkanalyse oder der Bildanalyse vielschichtige Social Media-Analysen. In den meisten Fällen geht es bei der Erforschung sehr großer Datenmengen um die Aggregation von Stimmungen und Trends. Diese Datenanalysen und -visualisierungen versammeln aber in der Regel nur faktische Gegebenheiten und lassen die Frage nach den sozialen Kontexten und Motiven außer Acht. Dessen ungeachtet hat sich der Big-Data-Ansatz in den Human-, Sozial- und Kulturwissenschaften mittlerweile etablieren können.

Durch das Internet und die steigende Beliebtheit von Social Media-Diensten gewinnen Forschungsansätze für den Umgang mit digitalen Kommunikationsdaten an Relevanz. Analoge Methoden, die zur Erforschung interpersonaler oder

Massenkommunikation entwickelt wurden, können aber nicht einfach auf die Kommunikationspraktiken im Social Net übertragen werden. Richard Rogers, ein einflussreicher Forscher im Bereich der Social Media Research, plädiert dafür, nicht mehr allein digitalisierte Methoden (wie zum Beispiel Online-Fragebögen) zur Erforschung der Vernetzungskultur anzuwenden, sondern sich auf digitale Methoden zu konzentrieren, die in der Lage sind, kulturellen Wandel und gesellschaftliche Entwicklungen zu diagnostizieren und zu prognostizieren. Als digitale Methoden lassen sich also Ansätze verstehen, die nicht schon bestehende Methoden für die Internetforschung adaptieren, sondern die genuinen Verfahrensweisen digitaler Medien aufgreifen. (vgl. Rogers 2013: 13f.) Digitale Methoden sind nach Rogers Forschungsansätze, die sich einerseits große Mengen digitaler Kommunikationsdaten zunutze machen, welche von Millionen Nutzern tagtäglich im Social Web produziert werden, und die diese andererseits mit computergestützten Verfahren filtern, analysieren, aufbereiten und darstellen. In der Tradition der Akteur-Netzwerk-Theorie gehen zahlreiche Repräsentanten der Internetforschung von digitalen Akteuren aus wie Hyperlinks, Threads, Tags, PageRanks, Protokolldateien, Cookies, die untereinander und mit Datensatzsubjekten interagieren. Die Akteur-Netzwerke können nur mit digitalen Methoden beobachtet, aufgezeichnet und beurteilt werden – auch wenn sie sich oft als instabile und ephemere Ereignisse herausstellen. Dabei entsteht eine neuartige Methodologie, die Aspekte der Informatik, Statistik, und der Informationsvisualisierung mit sozial- und geisteswissenschaftlichen Forschungsansätzen kombiniert. Die Vision einer solchen nativ-digitalen Forschungsmethodik, ob in Gestalt einer "computational social science" (Lazer 2009: 721-723) oder von "cultural analytics" (Manovich 2009: 199-212) ist jedoch noch unvollständig und verlangt nach einer epistemischen Befragung der digitalen Methoden in der Internetforschung folgender Bereiche:

Digitale Methoden als *geltungstheoretisches Projekt*. Dieses steht für ein bestimmtes Verfahren, das soziale Geltung von Handlungsorientierungen beansprucht. Die Wirtschaftsinformatik, die Computerlinguistik und die empirische Kommunikationssoziologie bilden nicht nur ein Geflecht wissenschaftlicher Felder und Disziplinen, sondern entwickeln in ihren strategischen Verbundprojekten bestimmte Erwartungsansprüche, die soziale Welt erklärend zu erschließen und sind insofern intrinsisch verbunden mit epistemischen und politischen Fragen. Vor diesem Hintergrund setzt sich eine das Selbstverständnis der digitalen Methoden befragende Epistemologie mit der sozialen Wirkmächtigkeit der digitalen Datenwissenschaften auseinander. In einem weiteren Schritt könnte man sich auch mit den diskursiven Zuschreibungen, welche die Sozialen Medien im Nutzungskontext erfahren, auseinandersetzen und sich mit der Frage beschäftigen, inwiefern diese über gesellschaftliche Selbstbilder Auskunft geben. Wenn schließlich von einer habitualisierten Nutzung der Social Media ausgegangen wird, kann im Anschluss daran die Frage nach der normativen, handlungsleitenden

Dimension der Netzwerke aufgeworfen werden, konkret kann gefragt werden, inwiefern sie entweder ein solidarisches oder ein individualistisches Agieren verstärken.

Digitale Methoden als *konstitutionstheoretisches Konstrukt*. Der Gegenstandsbezug der Big Data-Forschung ist heterogen und setzt sich aus unterschiedlichen Methoden zusammen. Mit ihren Technologien der Schnittstellen, den Verfahren des Datentrackings, des Keyword-Trackings, der automatischen Netzwerkanalyse, der Sentiment- und Argumentanalysen oder dem maschinenbasierte Lernen ergeben sich daher vielschichtige Perspektivierungen der Datenkonstrukte. Die Daten selbst firmieren in dieser Sichtweise nicht als Rohdaten, sondern können im rechnergestützten Raum der Möglichkeiten als optionale und modulare Konstellationen reproduziert werden. Vor diesem Hintergrund firmieren die digitalen Methoden auch als Hilfsinstrumente zur Aufrechterhaltung der digitalen Kontrollgesellschaft, deren medienkulturelle Dechiffrierbarkeit einer der großen Anliegen der Software Studies und der Critical Code Studies ist, die versuchen, die Dispositive der Informationsvergabe und die damit einhergehenden politischen Regulative von Layermodellen, Netzwerkprotokollen, Zugangspunkten und Algorithmen aufzuzeigen. In dieser Hinsicht ergibt sich ein vielversprechender Forschungsansatz. Zusätzlich zur Frage, inwiefern durch konkrete habitualisierte Nutzung spezifische sozio-technische Imaginationen entstehen können, fragen Software Studies und Critical Code Studies nach den medialen, technisch-infrastrukturellen Selbstbeschreibungsformeln sozialer Kommunikation. Digitale Methoden können letztlich auch als eine *gründungstheoretische Fiktion* aufgefasst werden. Die einschlägige Forschungsliteratur hat sich eingehend mit der Reliabilität und der Validität der wissenschaftlichen Datenerhebung beschäftigt und ist zum Ergebnis gekommen, dass die Datenschnittstellen des Social Net (Twitter, Facebook, YouTube) mehr oder weniger als dispositive Anordnungen im Sinne eines Gatekeeper fungieren. Als Filterschnittstellen erzeugen die API's ökonomisch motivierte Exklusionseffekte für die Netzforschung, die von ihr aus eigener Kraft nicht kontrolliert werden können. Vor diesem Hintergrund soll die Big Data Research in technologisch-infrastruktureller, forschungspragmatischer und wissenspolitischer Hinsicht datenkritisch hinsichtlich ihrer objektivistischen und positivistischen Postulate entlang von drei praxisnahen Case Studies befragt werden.

Praxisbezüge

Im Forschungsfeld der *Social Media Data* hat sich mit der Gesundheitsprognostik eine evidenzbasierte Praxis der Prävention herausgebildet, die auf die institutionelle Entwicklung der staatlich-administrativen Gesundheitsvorsorge und auf die

Kulturtechniken der Lebensführung Einfluss nehmen. Die Gesundheitsvorsorge beobachtet mit großem Interesse, dass weltweit Millionen von Nutzerinnen und Nutzern täglich mit der Internet-Suchmaschine Google Informationen zum Thema Gesundheit suchen. In Grippezeiten häufen sich die Suchanfragen zur Grippe und die Häufigkeit bestimmter Suchbegriffe kann Anhaltspunkte für die Häufigkeit von Grippeerkrankungen liefern. Studien zum Suchvolumenmuster haben herausgefunden, dass ein signifikanter Zusammenhang zwischen der Anzahl von grippebezogenen Suchanfragen und der Anzahl von Personen mit tatsächlichen Grippesymptomen besteht (Freyer-Dugas et. al. 2012 S. 463-469). Dieses epidemiologische Beziehungsgefüge kann zur Frühwarnung vor Epidemien auf Städte, Regionen, Länder und Kontinente ausgedehnt und differenziert dargestellt werden. Mit der epidemiologischen Auswertung von textuellen Clustern und semantischen Feldern erhält das Social Web den Status einer großen Datenbank, die das soziale Leben in seiner Gesamtheit widerspiegelt und damit eine repräsentative Datenquelle für die präventive Gesundheitspolitik darstellt. Die Kommunikationsprozesse in Online-Netzwerken stehen im Fokus staatlicher Biopolitik, die um die Gesundheit der Bevölkerung besorgt ist und spezifische Wissenstechniken und -modelle zur Erforschung der Big Data entwickelt hat, um die Wahrscheinlichkeit der Verbreitung von Krankheiten in absehbarer Zukunft statistisch zu schätzen.

Die Mehrzahl der Monitoring-Projekte, die große Datenmengen im Social Web untersuchen, wird von Computerlinguist/innen und Informatiker/innen durchgeführt. Generell interpretieren sie die Kommunikation als kollektiv geteilte und kulturspezifische Wissensstrukturen, mit denen Individuen versuchen, ihre Erfahrungen zu interpretieren. Die Erhebung dieser Wissensstrukturen verfolgt den Anspruch, einen sozial differenzierten Einblick in öffentliche Debatten und sozial geteilte Diskursnetze zu erhalten. Die Wissensstrukturen werden hierbei mit Hilfe eines korpuslinguistischen Ansatzes erschlossen. Am Beginn der Forschung steht die Erstellung eines digitalen Korpus, der sich aus begrifflichen Entitäten zusammensetzt, die in der Regel als 'kanonisch' eingestuft werden. So ergeben sich einige Hypothesen erst aus der empirischen Widerständigkeit der Big Data und entwickeln sich erst im Fortgang ihrer Beschreibung. Die Kategorienkataloge suggerieren damit zwar auf den ersten Blick wissenschaftliche Objektivität, andererseits bleibt angesichts der riesigen Datenmengen eine genaue Validierung der Begriffsauswahl, d.h. der interpretativen Selektion der Big Data, oft unklar und vage. Diese Unsicherheit bei der Hypothesenbildung liegt darin begründet, dass das umfangreiche Datenmaterial in keiner Gesamtschau mehr überblickt werden kann und daher auch nicht mehr linguistisch kodiert werden kann. Oft ist die erhobene Datenmenge so umfangreich, dass nach einer ersten Sondierung des Materials weitere Gewichtungen und Einschränkungen zur Komplexitätsreduktion vorgenommen werden müssen. An dieser methodischen Einschränkung des Big-Data-Monitoring wurde kritisiert, dass die erarbeiteten Erkenntnisse nur ein

atomistisches Bild der Daten liefern können und daher auf eine Kontextualisierung des Textmaterials und damit auf eine kontextsensitive Interpretation des Zeichengebrauchs weitgehend verzichten müssen (vgl. Boyd und Crawford 2011). Der Vorteil der Dekontextualisierung bei der nach Worthäufigkeiten fahndenden Big-Data-Analyse besteht nach Manovich (vgl. 2012, S. 465) darin, dass die einzelnen Worteinheiten auf eine enthierarchisierte und dezentrale Repräsentation des Wissens hinauslaufen und damit die Möglichkeit alternativer kollektiver Äußerungsgefüge anbieten.

Die Auswertung der Daten kann auf andere Trendentwicklungen erweitert werden. Mittlerweile gibt es zahlreiche Studien, welche die textuellen Daten der Sozialen Medien untersuchen, um *politische Einstellungen* (vgl. Conover et. al. 2011), *Finanztrends* und *Wirtschaftskrisen* (vgl. Gilbert und Karahalios 2010), *Psychopathologien* (vgl. Wald, Khoshgoftaar und Sumner 2012, S. 109-115) und *Aufstände* und *Protestbewegungen* (vgl. Yogatama 2012) frühzeitig vorherzusagen. Von einer systematischen Auswertung der Big Data erwarten sich die Prognostiker/innen eine effizientere Unternehmensführung bei der statistischen Vermessung der Nachfrage- und Absatzmärkte, individualisierte Serviceangebote und eine bessere gesellschaftliche Steuerung. Einen großen politischen Stellenwert hat vor allem die algorithmische Prognostik kollektiver Prozesse. In diesem Konnex ist das Social Web zur wichtigsten Datenquelle zur Herstellung von Regierungs- und Kontrollwissen geworden. Die politische Kontrolle sozialer Bewegungen verschiebt sich hiermit in das Netz, wenn Soziolog/innen und Informatiker/innen gemeinsam etwa an der Erstellung eines *Riot Forecasting* mitwirken und dabei auf die gesammelten Textdaten von Twitter-Streams zugreifen: "Due to the availability of the dataset, we focused on riots in Brazil. Our datasets consist of two news streams, five blog streams, two Twitter streams (one for politicians in Brazil and one for general public in Brazil), and one stream of 34 macroeconomic variables related to Brazil and Latin America." (Ebd., S. 3)

Big Data bietet eine spezifische Methode und Technologie zur statistischen Datenauswertung, die aus der epistemischen Schnittstelle von Wirtschaftsinformatik und kommerzieller Datenbewirtschaftung hervorgeht und die Bereiche der *Business Intelligence*, des *Data Warehouse*[1] und des *Data Mining*[2] in sich vereint. Die Diskussion um den technologisch-infrastrukturellen und machtstrategischen Stellenwert der Big Data zeigt auf, dass die nummerische Repräsentation von Kollektivitäten zu den grundlegenden Operationen digitaler Medien gehört und eine rechnerbasierte Wissenstechnik bezeichnet, mit welcher kollektive Praktiken mathematisch beschreibbar und auf diese Weise quantifizierbar werden. Die Bestimmung der Vielheiten mit Hilfe von nummerisch gegliederten Mengenangaben dient in erster Linie der Orientierung und kann als eine Strategie verstanden werden, die kollektive Datenströme in lesbare Datenkollektive übersetzt. In diesem Sinne firmieren in der medialen Öffentlichkeit soziale

Netzmedien wie Facebook, Twitter und Google+ als Spiegel der allgemeinen Wirtschaftslage (vgl. Bollen, Mao und Zeng 2011, S. 1-8) oder als prognostischer Indikator von nationalen Gefühlsschwankungen (vgl. Bollen 2011, S. 237-251). In diesem Sinn bilden sie selbst Schauplätze einer populären Aufmerksamkeit und popularisierender Diskurse, die ihnen bestimmte Außenwirkungen – etwa als ein Gradmesser der konjunkturellen Entwicklung der Wirtschaft und der sozialen Wohlfahrt – zuschreiben.

Case Study: Facebook Data Research

Welche Musik werden eine Milliarde Menschen in Zukunft hören, wenn sie frisch verliebt sind und welche Musik werden sie hören, wenn sie gerade ihre Beziehung beendet haben? Diese Fragestellungen hat das "Facebook Data Team" im Jahr 2012 zum Anlass genommen, um die Daten von über einer Milliarde Nutzer/innenprofilen (mehr als 10 Prozent der Weltbevölkerung) und 6 Milliarden Songs des Online-Musikdienstes Spotify mittels einer korrelativen Datenanalyse auszuwerten, die den Grad des gleichgerichteten Zusammenhangs zwischen der Variable "Beziehungsstatus" und der Variable "Musikgeschmack" ermittelt.[3] Diese Prognose über das kollektive Konsumverhalten basiert auf Merkmalsvorhersagen, die mittels Data Mining in einer simplen Kausalbeziehung ausgedrückt werden. Unter Leitung des Soziologen Cameron Marlow erforschte die aus Informatiker/innen, Statistiker/innen und Soziolog/innen bestehende Gruppe das statistische Beziehungsverhalten der Facebook-Nutzer/innen und veröffentlichte am 10. Februar des gleichen Jahres zwei Hitlisten von Songs, die Nutzer/innen hörten, als sie ihren Beziehungsstatus änderten, und nannte sie lapidar "Facebook Love Mix" und "Facebook Breakup Mix".[4] Die Forschergruppe im Back-End destillierte aus der statistischen Ermittlungsarbeit der "Big Data" (vgl. Wolf et.al. 2011, S. 217-219) nicht nur eine globale Verhaltensdiagnose, sondern transformierte diese auch in eine suggestive Zukunftsaussage. Sie lautete: Wir Forscher im Back-End bei Facebook wissen, welche Musik eine Milliarde Facebook-Nutzer/innen am liebsten hören werden, wenn sie sich verlieben oder trennen.[5] Unter dem Deckmantel des bloßen Sammelns und Weitergebens von Informationen etabliert die Forschergruppe des "Facebook Data Teams" eine Deutungsmacht gegenüber den Nutzern, indem sie die Nutzer/innen im automatisch generierten Update-Modus "What's going on?" auffordert, regelmäßig Daten und Informationen zu posten.

Die Zukunftsaussagen des "Facebook Data Teams" sind jedoch nur vordergründig mathematisch motiviert und verweisen auf den performativen Ursprung des Zukunftswissens. Trotz fortgeschrittener Mathematisierung, Kalkülisierung und

Operationalisierung des Zukünftigen bezieht das Zukunftswissen seine performative Macht immer auch aus Sprechakten und Aussageordnungen, die sich in literarischen, narrativen und fiktionalen Inszenierungsformen ausdifferenzieren können. (Vgl. Lummerding 2011: 199-216) In diesem Sinne sind die Bedeutungen im Möglichkeitsraum der Zukunft nicht eindeutig determiniert, sondern erweisen sich vielmehr als ein *aggregatähnliches Wissen*, dessen konsenserzwingende Plausibilität sich nicht in *Wahrheitsdiskursen* und *epistemischen Diskursen* erschöpft, sondern auch von *kulturellen* und *ästhetischen* Kommunikationsprozessen und Erwartungshaltungen *(patterns of expectation)* gestützt wird, die Imaginäres, Fiktives und Empirisches in Beziehung setzen.

Das Format der Hitliste und ihrer beliebtesten zehn Songs versucht, durch Vereinfachung komplexe Sachverhalte auf einen Blick darstellbar zu machen. Es handelt sich um ein popularisierendes Zukunftsnarrativ, das eine verhaltensmoderierende, repräsentative und rhetorische Funktion übernehmen und die Zukunftsforschung als unterhaltsame und harmlose Tätigkeit herausstreichen soll. Um in diesem Sinn glaubwürdig zu sein, muss die futurische Epistemologie immer auch auf eine gewisse Weise überzeugend in Szene gesetzt werden, sie muss theatralisch überhöht und werbewirksam inszeniert und erzählt werden, damit sie Aufmerksamkeit generieren kann. Insofern ist den futurischen Aussageweisen immer auch ein Moment der prophetischen Selbst- und Wissensinszenierung inhärent, mit dem die wissenschaftlichen Repräsentant/innen den gesellschaftsdiagnostischen Mehrwert der sozialen Netzwerke unter Beweis stellen wollen (vgl. Doorn 2010, S. 583-602). Soziale Netzmedien wie Facebook agieren heute als Global Player der Meinungsforschung und der Trendanalyse und spielen eine entscheidende Rolle bei der Modellierung von Zukunftsaussagen und futurologischer Wissensinszenierung.

Auch die Glücksforschung nutzt heute vermehrt Freundschaftsnetzwerke wie Facebook zur Auswertung ihrer Massendaten. Innerhalb der Big-Data-Prognostik stellt die sogenannte "Happiges Research" eine zentrale Forschungsrichtung dar. Doch die sozioökonomische Beschäftigung mit dem Glück wird überwiegend unter Ausschluss der akademischen Öffentlichkeit durchgeführt. In diesem Zusammenhang warnen einflussreiche Theoretiker wie Lev Manovich (vgl. 2012) und Danah Boyd (vgl. 2011) daher vor einem "Digital Divide", der das Zukunftswissen einseitig verteilt und zu Machtasymmetrien zwischen Forschern *innerhalb* und *außerhalb* der Netzwerke führen könnte. Manovich kritisiert den limitierten Zugang zu sozialstatistischem Daten, der von vornherein eine monopolartige Regierung und Verwaltung von Zukunft schafft: "[…] only social media companies have access to really large social data – especially transactional data. An anthropologist working for Facebook or a sociologist working for Google will have access to data that the rest of the scholarly community will not." (Manovich 2012: S. 467)

Dieses ungleiche Verhältnis festigt die Stellung der sozialen Netzwerke als computerbasierte Kontrollmedien, die sich Zukunftswissen entlang einer vertikalen und eindimensionalen Netzkommunikation aneignen: (1) Sie ermöglichen einen kontinuierlichen Fluss von Daten (digitale Fußabdrücke), (2) sie sammeln und ordnen diese Daten und (3) sie etablieren geschlossene Wissens- und Kommunikationsräume für Expert/innen und ihre Expertisen, welche die kollektiven Daten zu Informationen verdichten und interpretieren. Das Zukunftswissen durchläuft folglich unterschiedliche mediale, technologische und infrastrukturelle Schichten, die hierarchisch und pyramidal angeordnet sind: "The current ecosystem around Big Data creates a new kind of digital divide: the Big Data rich and the Big Data poor. Some company researchers have even gone so far as to suggest that academics shouldn't bother studying social media data sets – Jimmy Lin, a professor on industrial sabbatical at Twitter argued that academics should not engage in research that industry 'can do better'." (Boyd/Crawford 2011) Diese Aussagen verdeutlichen – neben der faktisch gegebenen technologisch-infrastrukturellen Abschottung des Zukunftswissens –, dass das strategische Entscheidungshandeln im Backend-Bereich und nicht in der Peer-to-Peer-Kommunikation angelegt ist. Die Peers können zwar in ihrer eingeschränkten Agency die Ergebnisse verfälschen, Fake-Profile anlegen und Nonsens kommunizieren, besitzen aber keine Möglichkeiten der aktiven Zukunftsgestaltung, die über taktische Aktivitäten hinausgehen.

Warum ist eigentlich die Erforschung des Glücks für die Gestaltung des Zukunftswissens so relevant geworden? Die Dominanz der Glücksforschung hat zwei historische Gründe (vgl. Frey/Stutzer 2002, S. 402). Seit der griechischen Antike wird dem Glück eine zentrale Stelle im menschlichen Leben eingeräumt und nach Aristoteles besteht das Ziel alles menschlichen Tuns darin, den Zustand der Glückseligkeit zu erlangen.[6] Ein weiterer maßgeblicher Diskursstrang ist der seit Jeremy Bentham einflussreich gewordene Utilitarismus der Glücksdiskurse. Mit dem *Greatest Happiness Principle* entwickelte Bentham die Vorstellung, dass das größte zu erreichende Gut das Streben nach dem größtmöglichen Glück für die größtmögliche Anzahl von Menschen bedinge (vgl. Bentham 1977, 393f.). An diese sozioökonomische Konzeption des Glücks knüpft die "Happiges Research" an, die Glück nach rationalem Kalkül als individuellen Nutzen interpretiert und in der Hochrechnung von aggregierten Glücksbekundungen das soziale Wohlbefinden berechnet.

Eine maßgebliche Spielart der futurologischen Prophetie stellt der seit 2007 eingeführte "Facebook Happiges Index" dar, der anhand einer Wortindexanalyse in den Statusmeldungen die Stimmung der Nutzer/innen sozialempirisch auswertet. Auf der Datengrundlage der Status-Updates errechnen die Netzwerkforscher/innen in ihrem "Gross National Happiges Index" (GNH) das sogenannte "Bruttonationalglück" von Gesellschaften. Der Soziologe Adam Kramer

arbeitete von 2008 bis 2009 bei Facebook und errechnete gemeinsam mit den Mitarbeiterinnen und Mitarbeitern des Facebook Data Team, der Sozialpsychologin Moira Burke, dem Informatiker Danny Ferrante und dem Leiter der Data Science Research Cameron Marlow den Happiness Index. Kramer konnte dabei das intern verfügbare Datenvolumen des Netzwerks nutzen. Er evaluierte die Häufigkeit von positiven und negativen Wörtern im selbstdokumentarischen Format der Statusmeldungen und kontextualisierte diese Selbstaufzeichnungen mit der individuellen Lebenszufriedenheit der Nutzer (*convergent validity*) und mit signifikanten Datenkurven an Tagen, an denen unterschiedliche Ereignisse die Medienöffentlichkeit bewegten (*face validity*): "'Gross national happiness' is operationalized as a standardized difference between the use of positive and negative words, aggregated across days, and present a graph of this metric." (Kramer 2010, S. 287) Die von den Soziologinnen und Soziologen analysierten individuellen Praktiken der Selbstsorge werden mit Hilfe von semantischen Wortnetzen letztlich auf die Oppositionspaare "Glück"/"Unglück" und "Zufriedenheit"/"Unzufriedenheit" reduziert. Eine binär strukturierte Stimmungslage wird schließlich als Indikator einer kollektiven Mentalität veranschlagt, die auf bestimmte kollektiv geteilte Erfahrungen rekurriert und spezifische Stimmungen ausprägt. Die soziologische Massenerhebung der Selbstdokumentationen (*self reports*) in sozialen Netzwerken hat bisher die Stimmungslage von 22 Nationalstaaten ermittelt. Mit der wissenschaftlichen Korrelation von subjektiven Befindlichkeiten und bevölkerungsstatistischem Wissen kann der "Happiges Index" nicht nur als Indikator eines "guten" oder "schlechten" Regierens gewertet werden, sondern als Kriterium einer möglichen Anpassungsleistung des Politischen an die Wahrnehmungsverarbeitung der Sozialen Netzwerke. In diesem Sinne stellt der „Happy Index" ein erweitertes Instrumentarium wirtschaftlicher Expansion und staatlicher-administrativer Entscheidungsvorbereitung dar.

Case Study: Twitter Research

Für die Big Data-Forschung stellen Anwendungsprogrammierungen ein zentrales Instrument der datengetriebenen Wissensproduktion dar. Der Begriff Anwendungsprogrammierung referiert auf die englischsprachige Bezeichnung *application programming interface*, kurz: API.

Es ist eine Ironie der Geschichte, dass die Usability der kommerziell motivierten Twitter-API eine Konjunktur der Big-Social-Data-Forschung im Bereich der Medien- und Kommunikationswissenschaften eingeleitet hat. Die Online-Forschung ist hierbei von den Versionen abhängig, die das Unternehmen Twitter ihnen zur freien Verfügung überlässt. Aktuell ist es die Version 1.1. der REST API *(Representational*

State Transfer), die von Twitter für die Nutzung vorgesehen ist, die ältere Version wurde eingestellt und ist nicht mehr verwendbar. Die API-Schnittstelle kann als eine nutzerspezifische Gebrauchsanordnung für die Genese und Konstitution epistemischer Praktiken aufgefasst werden. Die API-Schnittstelle gibt u.a. die Menge und den Modus von Auswahlfreiheiten und Zugangsbeschränkungen vor. Eine mediale Dispositivanalyse könnte an der von Danah Boyd, Jane Crawford und Lev Manovich artikulierten Datenkritik der Forschungsinfrastrukturen ansetzen und die Analyse von medialen Anordnung und ihren Diskursen mit der Analyse der an dieser Schnittstelle entstehenden Machteffekte weiterentwickeln.

Richard Rogers hat den Begriff *online groundedness* entwickelt, "um Forschung zu beschreiben, die dem Medium folgt, die dessen Dynamiken erfasst und fundierte Aussagen über den kulturellen und gesellschaftlichen Wandel trifft." (2013: 64) Kann die Onlineforschung selbständig entscheiden, einem Medium zu folgen, oder lagert sie sich nicht vielmehr als ein taktischer Effekt an vorgegebene Tendenzen an? Die Twitter-API hat bisher eine konstituierende Macht über den Aufschwung der angewandten Twitter-Forschung ausgeübt. Die von Twitter zur Verfügung gestellte Anwendungsprogrammierung kann in zweifacher Hinsicht als mediendispositive Infrastruktur problematisiert werden. Einerseits schreibt sie als Programmierparadigma für Webanwendungen die Logik von Back- und Frontend fort und firmiert folglich nicht als Fenster zur sozialen Datenwelt, sondern erstellt im Sinne einer automatisierten Vorselektion softwarebasierte Filter selektiver Wissensgenerierung, die für die gewöhnliche Forschung nicht hintergehbar ist. Die Filterfunktion der Anwendungsschnittstelle etabliert daher systematisch Intransparenz, regelungstechnische Lücken und epistemische Unklarheiten. Die im Rahmen der Anwendungsprogrammierungen operierenden Erhebungs- und Verarbeitungsmethoden können in dieser Hinsicht auch als *gründungstheoretische Fiktion* aufgefasst werden. Die einschlägige Forschungsliteratur (vgl. Burgess/Puschmann 2013) hat sich eingehend mit der Reliabilität und der Validität der Twitterdaten beschäftigt und ist zum Ergebnis gekommen, dass die Twitter-Datenschnittstellen mehr oder weniger als dispositive Anordnungen im Sinne eines Gatekeeper aufgefasst werden können. Dies lässt sich an den unterschiedlichen Zugangspunkten des Mikroblogging-Dienstes Twitter, der Twitter-API aufzeigen: "statuses/filter" gibt alle öffentlichen Tweets zurück, welche einer Menge übergebener Filterparameter entsprechen. Die Standardzugriffsrechte erlauben bis zu 400 Schlüsselwörter, 5000 Nutzerids und 25 Orte. "Statuses/samples" gibt eine kleine, zufällige Untermenge aller öffentlichen Tweets zurück. "Statuses/firehouse" ist ein Zugriffspunkt, der spezielle Rechte benötigt, um sich mit ihm verbinden zu können. Als Filterschnittstellen erzeugen die API's also immer auch ökonomisch motivierte Exklusionseffekte für die Netzforschung, die von ihr aus eigener Kraft nicht kontrolliert werden können. Digitale Methoden und Analysen situieren sich folglich im Wechselspiel zwischen Medieninnovationen und den Limitationen technischer Infrastrukturen. Um also präzise die Machtasymmetrien sozio-

technischer Interaktionen herauszuarbeiten, müsste man schließlich nicht an der schieren Existenz einzelner Kommunikationstechnologien und ihrer digitalen Methoden ansetzen, sondern sich vielmehr auf die mit ihnen verknüpften konkreten Datenpraktiken und technisch-infrastrukturellen Beschränkungen einlassen.

Data Mining und Social Monitoring

Die operative Erforschung der Big Data weist zwei unterschiedliche Ausrichtungen auf. Einerseits versucht sie, kollektive Figurationen und Tendenzen kollektiver Dynamiken zu modellieren, andererseits geht es ihr darum, aus großen Datenmengen *personenzentrierte* und *zielgruppenspezifische* Merkmalsausprägungen herauszulesen. Im Folgenden sollen die historischen Diskursfelder dieser datenbasierten Techniken zur Herstellung subjektzentrierten Wissens herausarbeitet werden, um das strategische Bezugsverhältnis zwischen Wissen und Macht – oder genauer: zwischen Regierungswissen und der Herausbildung von Subjektivierungsmodellen – akzentuieren zu können.

In seinen Anfängen wurde das Profiling als Bewertungsmethode im Personalausleseverfahren der Testpsychologie in den USA entwickelt (vgl. Giordano 2005). Die standardisierten Verfahren der Testpsychologie zur Ermittlung von Leistungsfähigkeit bilden direkte Vorläufer des Profiling (aber auch der Rasterfahndung). Begriffe wie das "Persönlichkeitsprofil" oder das "Profiling" entstammen dem psychologisch-therapeutischen Diskurs und markieren heute Leitdiskurse in den Praxisformen der Selbstthematisierung. Unter den Vorzeichen des Postfordismus hat sich das Profiling als ein Ökonomisierungs- und Standardisierungsinstrument gesellschaftlich verallgemeinert und ist als eine vielschichtige Such- und Analysemethode der Informations- bzw. Wissensgesellschaften in Verwendung. Das hohe Ansehen der Selbstevaluation verweist auf zwei soziale Prozesse. Einerseits hat sich die Anzahl der Testparameter und -verfahren und der daran beteiligten Testobjekte mit dem Auftritt der Web-2.0-Interfacetechnologien vervielfältigt, andererseits hat sich – in Abgrenzung zur beruflichen Eignungsdiagnostik – die Evaluationspraxis auch in qualitativer Hinsicht verändert und umfasst heute die gesamte Persönlichkeit und kreativen Potenziale des Subjekts.

Das Web 2.0 mit seinen Social Networks und Communities verspricht ein großes prognostisches Potenzial, weil Marketingaktivitäten auf bestimmte Zielgruppen mittels modularer Technologien für User Tracking, Webmining, Profiling, Testing, Optimierung, Ad-Serving und Targeted Advertising abgestimmt werden können. Das Profiling im Web 2.0 verläuft nach dem Prinzip des Closed Circuit. Die Anordnung des Closed Circuit beschreibt ein Aufzeichnungsverfahren, bei der das

Eingabemedium direkt mit dem Abbildungsmedium verbunden ist. Bei der Beobachtungsanordnung im Closed Circuit machen die User/innen die Erfahrung der Synchronität ihrer Handlungen. Die sofortige Verfügbarkeit der Datenstrukturen und ihre gleichzeitige Manipulationsmöglichkeit durch das Targeted Advertising ist eine besondere Eigenschaft des Echtzeit-Profilings, das vergangene Nutzungsgewohnheiten von Online-Rezipient/innen analysiert (Click Advertising, Graphenanalyse), um zielgerichtete Werbung (Quality Market) für ein künftiges Konsumverhalten zu modellieren. Vor diesem Hintergrund entwickelte Microsoft ein Profiling-System, das soziometrische Daten wie etwa Alter, Geschlecht, Einkommen und Bildung mit möglichst großer Wahrscheinlichkeit ableiten sollte. Die Prognosefähigkeit der Sozialen Netzwerke bleibt aber davon abhängig, ob es gelingt, die biografisch und demografisch relevanten Daten und Informationen in distinkte und segregierte Bausteine der weiteren Datenverarbeitungen aufzugliedern. Als ein gemischtes Medium muss sich das Profiling zwangsläufig aus heterogenen Repräsentationen zusammensetzen. Es übernimmt das Modell der Prüfung von Persönlichkeitsmerkmalen der älteren Eignungsdiagnostik und macht es zur Sache kollektiver Approbationsleistungen, um seine Wirkungsweisen zu vervielfältigen und zu verstärken.

Die Profilbildung enthält Wissenstechniken, die auf binären Unterscheidungen beruhen (z.B. die Geschlechtszugehörigkeit), mit quantitativen Skalierungen operieren (z.B. hierarchische Ranking-Techniken) oder die auf die Erstellung qualitativer Profile abzielen (z.B. das Aufzeigen kreativer Fähigkeiten und Begabungen in „freien" Datenfeldern). Profile reproduzieren einerseits soziale Normen und bringen andererseits auch neue Formen von Individualität hervor. Sie verkörpern den Imperativ zur permanenten Selbstentzifferung auf der Grundlage bestimmter Auswahlmenüs, vorgegebener Datenfelder und eines Vokabulars, das es den Individuen erlauben soll, sich selbst in einer boomenden Bekenntniskultur zu verorten. Das „bedienerfreundliche" Profiling besteht in der Regel aus sogenannten Tools, das sind Checklisten, Fragebögen für Selbst-Evaluierung, analytische Rahmen, Übungsabschnitte, Bilanzen, Statistiken mit Kommentar, Datenbanken, Listen von Adressen und pädagogische Module zur Ermittlung individueller Fähigkeiten, Neigungen und Lieblingsbeschäftigungen.

Kommerzielle Suchmaschinen analysieren mittels Behavioural Targeting die Profile ihrer Nutzer. Diese Suchtechnologie erlaubt es, auf verhaltensorientierte Kriterien wie Produkteinstellung, Markenwahl, Preisverhalten, Lebenszyklus zu reagieren und relevante Werbung zu schalten. Das Behavioural Targeting evaluiert kontinuierliche Nutzungsgewohnheiten, private Interessen und demografische Merkmale und erstellt damit ein statistisches Relief pluraler und flexibler Subjektivität (vgl. Castelluccia 2012, S. 21-33). Das wesentliche Merkmal des digitalen Targeting ist der Sachverhalt, dass das Individuum nur noch als dechiffrierbare und transformierbare Figur seiner Brauchbarkeiten ins Blickfeld rückt. Es erzeugt ein

multiples und „dividuelles Selbst“ (Deleuze 1993, S. 260), das zwischen Orten, Situationen, Teilsystemen und Gruppen oszilliert – ein Rekurs auf eine personale Identität oder ein Kernselbst ist unter dividuellen Modulationsbedingungen nicht mehr vorgesehen. Im Unterschied zur klassisch analogen Rasterfahndung geht es beim digitalen Data Mining nicht mehr um die möglichst vollständige Ausbreitung der Daten, sondern um eine Operationalisierung der Datenmassen, die für prognostische Abfragen und Auswertungen effektiv in Beziehung zueinander gesetzt werden können. Es verändert nicht nur die Wissensgenerierung persönlicher Daten und Informationen, sondern auch die Prozesse sozialer Reglementierung. Insofern erzeugt das computergestützte Behavioural Targeting mehr als eine technische Virtualisierung von Wissensformen, denn es transformiert nachhaltig das Konzept des Raums, was zur Folge hat, dass sich das Raster vom topografischen Raum verflüchtigt und an seine Stelle der topologische Datenraum tritt. Dieser topologische Datenraum steht in Opposition zur Anwendungsschicht, die dem Kommunikationsraum der Nutzer/innen entspricht. Das futurische Wissen (bestehend aus der statistischen Erhebungsmethode des Data Mining, der Visualisierungstechnik des Data Mapping und des systematischen Protokollierungsverfahren des Data Monitoring) ist konstitutiv aus der Anwendungsschicht ausgeschlossen und den Nutzer/innen nicht zugänglich. Damit basiert das Zukunftswissen der sozialen Netzwerke auf einer asymmetrisch verlaufenden Machtbeziehung, welche sich in die technische Infrastruktur und in den Aufbau des medialen Dispositivs verlagert hat.

Literatur

Adelmann, Ralf (2011): Von der Freundschaft in Facebook: Mediale Politiken sozialer Beziehungen in Social Network Sites. In: *Generation Facebook: Über das Leben im Social Net*, hrsg. Oliver Leistert und Theo Röhle, S. 127-144. Bielefeld: transcript Verlag.

Ball, Kirstie/Haggerty, Kevin D./Lyon, David (2012): *Routledge Handbook of Surveillance Studies*, London: Routledge.

Bentham, Jeremy (1977): A Comment on the Commentaries and A Fragment on Government, hg. v. James H. Burns and Herbert L. A. Hart, in: The Collected Works of Jeremy Bentham, London, 1977.

Bollen, Johan (2011): Happiness Is Assortative in Online Social Networks. In: *Artifical Life* 17 (3): 237-251.

Bollen, Johan, Huina Mao und Xiaojun Zeng (2011): Twitter mood predicts the stock market. In: *Journal of Computational Science* 2 (1): 1-8.

Bollier, David (2010): The promise and peril of big data, Washington, DC: The Aspen Institute, Online: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf, zuletzt gesichtet am 27. Dezember 2013.

Boyd, Danah und Kate Crawford (2011): Six Provocations for Big Data. In: *Conference Paper, A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011, Oxford: University of Oxford und Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 zuletzt gesichtet am 27. Dezember 2013.

Burgess, Jean, und Cornelius Puschmann : "The Politics of Twitter Data", 2013. www.papers.ssrn.com/sol3/papers.cfm?abstract_id=2206225 (20. Oktober 2014).

Castelluccia, Claude (2012): Behavioural Tracking on the Internet: A Technical Perspective. In: *European Data Protection. Good Health?*, hrsg. Serge Gutwirth et. al., S. 21-33. New York/London: Springer Verlag.

Conover, Michael D., Bruno Goncalves, Jacob Ratkiewicz, Alessandro Flammini und Filippo Menczer (2011): Predicting the Political Alignment of Twitter Users. Proceedings of the 3rd IEEE Conference on Social Computing, forthcoming, Online: http://cnets.indiana.edu/wpcontent/uploads/conover_prediction_socialcom_pdfexpress_ok_version.pdf, zuletzt gesichtet am 27. Dezember 2013.

Deleuze, Gilles (1993): *Unterhandlungen 1972-1990*. Frankfurt am Main: Suhrkamp Verlag.

Doorn, Niels Van (2010): The ties that bind: the networked performance of gender, sexuality and friendship on MySpace. In: *New Media & Society* 12 (4): 583-602.

Driscoll, Kevin: »From punched cards to ›Big Data‹: A social history of database populism«, in: *Communication +1* (1), 2012. http://kevindriscoll.info/ (20. Mai 2014).

Frey, Bruno S. und Alois Stutzer (2002): What can economists learn from happiness research? In: *Journal of Economic Literature* 40 (2): 402-435.

Freyer-Dugas, Andrea et. al. (2012): Google Flu Trends: Correlation With Emergency Department Influenza Rates and Crowding Metrics. In: *Clinical Infectious Diseases,* 54(15): 463-469.

Gilbert, Eric und Karrie Karahalios (2010): Widespread Worry and the Stock Market. In: *4th International AAAI Conference on Weblogs and Social Media (ICWSM).* Washington, DC: George Washington University.

Giordano, Gerard (2005): *How testing came to dominate American schools: The history of educational assessment*. New York/Wien: Lang.

Gitelman, Lisa, und Geoffrey B. Pingree: *New Media: 1740-1915*, Cambridge, MA: MIT Press 2004.

Klaasen, Abbey (2007): The Right Ads at the Right Time – via Yahoo; Web Giant Looks to Offer Behavioral-Targeting Tools outside its Own Properties. In: *Advertising Age* 3/Februar: 34-39.

Kramer, Adam D. I. (2010): An Unobtrusive Behavioral Model of 'Gross National Happiness'. In: *Conference on Human Factors in Computing Systems* 28 (3), hrsg. Association for Computing Machinery, New York, S. 287-290.

Lazer, David et al. (2009): „Computational Social Science". Science, 323 (5915), S. 721-723.

Leistert, Oliver und Theo Röhle, Theo (Hrsg.) (2011): *Generation Facebook: Über das Leben im Social Net*. Bielefeld: transcript Verlag.

Lemke, Thomas (2004): Test. In: *Leviathan. Zeitschrift für Sozialwissenschaft* 32 (1): 119-124.

Lummerding, Susanne (2011): Facebooking. What You Book is What You Get – What Else? In: *Generation Facebook: Über das Leben im Social Net*, hrsg. Oliver Leistert und Theo Röhle, S. 199-216. Bielefeld: transcript Verlag.

Manovich, Lev: »How to Follow Global Digital Cultures: Cultural Analytics for Beginners«, in: Konrad Becker und Felix Stalder (Hg.): *Deep Search: The Politics of Search beyond Google*, Edison, NJ, 2009, S. 198–212.

Manovich, Lev (2012): Trending: The promises and the challenges of Big Social Data. In: *Debates in the digital humanities,* hrsg. Matthew K. Gold, S. 460-475. Minneapolis: University of Minnesota Press.

Rogers, Richard (2013): *Digital Methods*. Cambridge, Mass.: MIT Press.

Taewoo Nam und Jennifer Stromer-Galley (2012): The Democratic Divide in the 2008 US Presidential Election. In: *Journal of Information Technology & Politics* 9 (2):133-149.

Wald, Randall, Taghi M.Khoshgoftaar und Chris Sumner. 2012. Machine Prediction of Personality from Facebook Profiles. In: *13th IEEE International Conference on Information Reuse and Integration*, Washington, S. 109-115.

Webb, Stephen (2006): *Social Work in a Risk Society. Social and Political Perspectives*. Houndmills: Palgrave Macmillan, S. 165.

Wolf, Fredric et. al.. (2011): Education and data-intensive science in the beginning of the 21st century. In: *OMICS: A Journal of Integrative Biology* 15 (4): 217-219.

Yogatama, Dani (2012): *Predicting the Future: Text as Societal Measurement*, Online: http://www.cs.cmu.edu/~dyogatam/Home_files/statement.pdf, zuletzt gesichtet am 27. Dezember 2013.

Fussnoten

1. Das Data Warehousing ist eine infrastrukturelle Technologie, die zur Auswertung großer Datenbestände dient.
2. Im kommerziellen Bereich etablierte sich der Begriff Data Mining für den gesamten Prozess des Knowledge Discovery in Databases. Data Mining meint die Anwendung von explorativen Methoden auf einen Datenbestand mit dem Ziel der Mustererkennung. Ziel der explorativen Datenanalyse ist über die Darstellung der Daten hinaus die Suche nach Strukturen und Besonderheiten. Sie wird daher typischerweise eingesetzt, wenn die Fragestellung nicht genau definiert ist oder auch die Wahl eines geeigneten statistischen Modells unklar ist. Ihre Suche umfasst, ausgehend von der Datenselektion, alle Aktivitäten, die zur Kommunikation von in Datenbeständen entdeckten Mustern notwendig sind: Aufgabendefinition, Selektion und Extraktion, Vorbereitung und Transformation, Mustererkennung, Evaluation und Präsentation.
3. Facebook Data Science, https://www.facebook.com/data (letzter Zugriff: 28.12.2013).
4. Unter dem Titel "Facebook Reveals Most Popular Songs for New Loves and Breakups" äußerte sich "Wired" begeistert über die neuen Möglichkeiten des Data Minings: www.wired.com/underwire/2012/02/facebook-love-songs/ (letzter Zugriff: 28.12.2013).
5. Die kollektive Figur "Wir" meint in diesem Fall die Forscher im Backend-Bereich und hat futurologische Verschwörungstheorien angeheizt, die das Weltwissen in den Händen weniger Forscher vermuten.
6. Dieses unveräußerliche Recht des Menschen auf Glück (the pursuit of happiness) nahmen die Vereinigten Staaten von Amerika in die Eröffnungspassage ihrer Unabhängigkeitserklärung auf.