

## Big Data - small problems? Ethische Dimensionen der Forschung mit Online-Kommunikationsspuren

Heise, Nele

Erstveröffentlichung / Primary Publication

Sammelwerksbeitrag / collection article

### Empfohlene Zitierung / Suggested Citation:

Heise, Nele: Big Data - small problems? Ethische Dimensionen der Forschung mit Online-Kommunikationsspuren. In: Maireder, Axel (Ed.) ; Ausserhofer, Julian (Ed.) ; Schumann, Christina (Ed.) ; Taddicken, Monika (Ed.): *Digitale Methoden in der Kommunikationswissenschaft*. Berlin, 2015 (Digital Communication Research 2). - ISBN 978-3-945681-02-2, pp. 39-58. URN: <http://dx.doi.org/10.17174/dcr.v2.3>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:  
<https://creativecommons.org/licenses/by/4.0>

**Empfohlene Zitierung:** Heise, N. (2015). Big Data – small problems? Ethische Dimensionen der Forschung mit Online-Kommunikationsspuren. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Hrsg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 39-58). doi: 10.17174/dcr.v2.3

**Zusammenfassung:** Kaum ein Thema hat in den vergangenen Jahren derart kontroverse Debatten in akademischen Kreisen hervorgerufen wie Big Data. Die automatisierte Verarbeitung riesiger Datenmengen für sozialwissenschaftliche Zwecke wirft nicht nur neue methodologische und theoretische Fragen auf – auch die ethischen Implikationen dieses datengetriebenen Forschungsansatzes geraten zusehends in den Blick. Der Beitrag diskutiert einige Besonderheiten von Big-Data-Analysen, insbesondere in Social Media, und welche Herausforderungen sich für die Umsetzung forschungsethischer Standards ergeben. Im Fokus steht auch die Frage, welchen Status die Nutzerinnen – die Menschen hinter den Daten – bei dieser Art der Forschung haben. Zudem sollen einige generelle Implikationen von Big-Data-Verfahren für Praxis und grundlegende Prinzipien sozialwissenschaftlicher Forschung beleuchtet werden. Ziel des Beitrages ist es, Forscherinnen für forschungsethische Fragen zu sensibilisieren sowie eine Verständigung der scientific community zum Umgang mit Big Data anzuregen.

**Lizenz:** Creative Commons Attribution 4.0 (CC-BY 4.0)

*Nele Heise*

## Big Data – small problems?

### Ethische Dimensionen der Forschung mit Online-Kommunikationsspuren

Im März 2014 publizierte das US-Journal PNAS eine Studie, in der die Timeline von fast 690.000 Nutzerinnen<sup>1</sup> der sozialen Netzwerkplattform Facebook experimentell beeinflusst wurde, um die Theorie der emotionalen Ansteckung zu untersuchen (Kramer et al., 2014). Nicht nur die immens hohe Zahl an verarbeiteten Daten und die Manipulation der Nutzungsoberfläche erregte die Gemüter: Vor allem das Zustandekommen der Studie ohne das explizite Einverständnis der Untersuchungsteilnehmerinnen führte international zu heftig geführten Diskussionen um die Rolle von Unternehmen wie Facebook, wissenschaftliche Standards und ethische Aspekte bei der Nutzung von Big Data (Deterding, 2014).

Die kontroversen Debatten um das „Facebook-Experiment“ verdeutlichen, dass es im Hinblick auf Big-Data-Analysen in der (sozial-)wissenschaftlichen Forschung noch großen Verständigungsbedarf gibt. Zugleich wirft der Fall zahlreiche Fragen nach den ethischen Implikationen dieses neuen, datenintensiven Forschungsansatzes auf. Vor diesem Hintergrund diskutiert der vorliegende Beitrag zunächst einige

1 Zur leichteren Lesbarkeit wird auf die Aufführung des männlichen Begriffsäquivalents verzichtet.

Besonderheiten von Big-Data-Forschung und welche Herausforderungen sich daraus für die Umsetzung forschungsethischer Standards ergeben. Im Fokus steht dabei auch die Frage, welchen Status die beforschten Nutzerinnen, also die Menschen hinter den Daten, bei dieser Art der Forschung haben. Darüber hinaus sollen auch generelle Implikationen von Big-Data-Verfahren für die Praxis und grundlegende Werte sozialwissenschaftlicher Forschung beleuchtet werden.

## 1 **Big Data und ethisch relevante Besonderheiten online-basierter Forschung**

### 1.1 *Was Big-Data-Forschung ist*

Um den Begriff Big Data gibt es viele Diskussionen: Ist er lediglich ein Marketing-Buzzword, ein generalisierter und daher unpräziser Begriff, oder verweist er auf ein gänzlich neues Forschungsparadigma im Sinne einer datengetriebenen Wissenschaft? In der einschlägigen sozialwissenschaftlichen Literatur ist man aufgrund dieser Unbestimmtheit um alternative Begriffe bemüht, wie etwa „traceable information“ oder „massified research“ (Neuhaus & Webmoor, 2012). Im Folgenden meinen wir damit ganz pragmatisch die Erfassung, Speicherung und Verwertung von „massive quantities of information produced by and about people, things, and their interactions“ (boyd & Crawford, 2012, S. 662).

Vereinfacht ausgedrückt geht es bei Big Data im kommunikationswissenschaftlichen Kontext also um große Datensätze von Online-Kommunikations Spuren und deren Verwendung für datenbasierte Forschung. Ein wesentliches Merkmal „datenintensiver“ Ansätze ist die weitgehende Automatisierung des wissenschaftlichen Prozesses „von der Datenerfassung über die Verarbeitung bis hin zur Modellbildung“, so dass die Wissenschaftlerinnen „erst ziemlich spät einen Einblick in ihre Daten“ (Pietsch, 2013, S. 58-59) erhalten. Erhoben werden die Daten meist automatisiert, beispielsweise mittels der APIs<sup>2</sup> von sozialen Netzwerktopplattformen wie Facebook oder Twitter. Auch die Verarbeitung, Auswertung

2 Die Abkürzung API steht für „application programming interface“, also Programmierschnittstellen eines Softwaresystems, die eine Anbindung an das System ermöglichen, zum Beispiel für Zugriffe auf Datenbanken.

und Darstellung der meist sehr großen Datensätze erfolgt bei Big-Data-Verfahren weitgehend automatisiert über verschiedene Software-Tools, mit deren Hilfe kausale Zusammenhänge, Korrelationen und Abhängigkeiten zwischen einer großen Zahl an Variablen und Indikatoren ermittelt werden können. Während so genannte „predictive analytics“ in Bereichen wie der Informatik, Medizin, Wirtschaft oder Politik bereits vor Jahren Einzug hielten, sind Big-Data-Analysen in den Sozialwissenschaften ein noch eher junges Phänomen. In der Sozialforschung werden sie unter anderem zur Beobachtung der Dynamiken öffentlicher Meinungsbildung in Echtzeit genutzt, zum Beispiel anhand politischer Diskussionen auf Twitter (Kaczmirek et al., 2014; Zeller, 2014).

Der Stellenwert solcher Analysen ist in den Sozialwissenschaften angesichts zahlreicher methodologischer und theoretischer Herausforderungen jedoch nicht unumstritten (siehe Mahrt in diesem Band).<sup>3</sup> Zum einen stören sich Kritikerinnen an der Annahme, dass durch große Datensätze allein Ergebnisse mit höherem Anspruch an Wahrheit, Objektivität oder Genauigkeit erzielt werden (boyd & Crawford, 2012). Zum anderen widerstrebt einigen die mit Big-Data-Analysen einhergehende „Mathematisierung gesellschaftlicher Zusammenhänge“ (Pietsch, 2013, S. 62), weil sie hierin die Gefahr einer positivistisch-reduktionistischen Sozialforschung sehen (Uprichard, 2013).

Seltener indes werden die forschungsethischen Implikationen von Big-Data-Forschung dezidiert in den Blick genommen. Zunächst wird daher im Folgenden diskutiert, inwieweit Besonderheiten onlinebasierter Kommunikation(sdaten) sowie Rahmenbedingungen onlinebasierter Forschung für Big-Data-Analysen forschungsethisch relevant sind, denn: „the information generated by users of social media platforms and services cannot be considered equivalent to conventional types of offline information collected by social researchers“ (Neuhaus & Webmoor, 2012, S. 46). Die Übertragung gängiger Standards gestaltet sich bei Big-Data-Analysen vor allem deshalb schwierig, weil Status und Identität der Teilnehmerinnen und der „Daten“ sowie die Bedingungen, unter denen Forscherinnen mit ihnen interagieren, weniger eindeutig sind als bei offline durchgeführter Forschung (Neuhaus & Webmoor, 2012). Insbesondere die (räumlich-zeitliche)

3 Nähere Ausführungen dazu in Mahrt und Scharkow (2013), Tinati, Halford, Carr und Pope (2014) oder Zeller (2014).

De-Kontextualisierung und Entgrenzung privater und öffentlicher Kommunikation sowie die Anonymität und kommunikative Distanz zwischen Forscherin und Forschungssubjekten können die Prozesse ethischer Entscheidungsfindung bei Big-Data-Analysen erschweren.

## 1.2 *De-Kontextualisierung und (kommunikative) Distanz als Ursachen ethischer Dilemmata*

„Taken out of context, Big Data loses its meaning“ (boyd & Crawford, 2012, S. 670). Mit dieser Aussage bringen die Autorinnen ein zentrales Merkmal von Big-Data-Analysen auf den Punkt: Daten oder vielmehr Spuren sozialer Kommunikation werden bei Big-Data-Analysen aus dem Kontext ihrer Entstehung gegriffen. Dies ist forschungsethisch insofern relevant, als der Entstehungskontext und damit verknüpfte Annahmen über die Privatheit/Öffentlichkeit der Kommunikation die Erwartungen der Nutzerinnen, wer die von ihnen mitgeteilten Informationen wahrnehmen soll, beeinflussen. Insbesondere die Kommunikation in Social Media ist jedoch von einer gewissen Entgrenzung oder Vermischung privater und öffentlicher Kontexte geprägt (Zeller, 2014). Da längst nicht alle Nutzerinnen den Zugang zu ihren Profilen mithilfe von Privacy-Einstellungen regulieren, können Forscherinnen nicht per se davon ausgehen, dass die Vorstellungen der Nutzerinnen mit der tatsächlichen Reichweite der Kommunikation übereinstimmen (Heise & Schmidt, 2014). Die Erfassung und Verwertung von explizit/implizit produzierten Inhalten kann daher mitunter die „kontextuelle Integrität“ (Nissenbaum, 2004) verletzen, das heißt einen Bruch der Handlungs- und Erwartungszusammenhänge hinsichtlich der Frage darstellen, was mit diesen Informationen geschehen soll – und was nicht:

„Data are most often not produced with the intention that it be used as raw research material. Additionally, the original data are, in the process of mining and visualizing, taken out of its context of production and assembled as a new instance“ (Neuhaus & Webmoor, 2012, S. 58).

In diesem Zusammenhang weist boyd (2010) darauf hin, dass es sich bei der Aggregation, Zusammenführung und Weiterverarbeitung gesammelter Daten letztlich um eine Art Zweckentfremdung handelt, die potenziell eine Verletzung

der Privatheit bedeuten kann (siehe Abschnitt 2). Auch das „frictionless sharing“ – das Teilen von Inhalten über verschiedene Plattformen hinweg – stellt eine Herausforderung dar, da die Daten teils automatisch von einem Kontext in einen anderen überführt werden, beispielsweise von einem öffentlichen Twitter- auf ein nicht-öffentliches Facebook-Profil.

In solchen Fällen ist fraglich, auf welchen Kontext oder welche Online-Umgebung sich die Erwartungen der Nutzerinnen beziehen. Dies wirkt sich auch auf die Berechtigung der Forscherinnen aus, deren Daten in ihre Analysen einzubeziehen, „because people have different values or expectations in different online settings“ (Eynon, Schroeder & Fry, 2009, S. 188). Hierfür gilt es also Beurteilungskriterien zu entwickeln, die sich etwa an der Zugänglichkeit und Sensibilität der Informationen bemessen können (Heise & Schmidt, 2014); vor allem, da vielen Nutzerinnen die unterschwellige Erhebung ihrer Daten mitunter gar nicht bewusst ist: „Many are not aware of the multiplicity of agents and algorithms currently gathering and storing their data (...) [they] are not necessarily aware of all the multiple uses, profits, and other gains that come from information they have posted“ (boyd & Crawford, 2012, S. 673).

Eine weitere forschungsethische relevante Herausforderung onlinebasierter Forschung ist der im Vergleich zu Offline-Forschung geringere Grad an sozialer Präsenz und wechselseitiger Sichtbarkeit (Heise & Schmidt, 2014). So kann die Anonymität der Beziehung zwischen Forscherin und Beforschter bei Big-Data-Analysen insbesondere im Hinblick auf Verantwortlichkeit problematisch sein: “In terms of accountability, the issue of (researcher) anonymity now becomes a problem rather than being a solution for minimizing risk to the individual” (Neuhaus & Webmoor, 2012, S. 57).

Hinzu kommt die große kommunikative Distanz bei Big-Data-Verfahren. Den gängigen reaktiven sozialwissenschaftlichen Forschungsmethoden ist ein unterschiedlicher Grad an Austauschprozessen und Interaktionen inhärent, über die kommunikative Beziehungen zwischen Forscherin und Forschungsobjekten hergestellt werden (Russel & Kelly, 2002; Ziegaus, 2009). Bei gewissen onlinebasierten Forschungsverfahren erfolgt diese Kommunikation, im Rahmen derer die Teilnehmerinnen in der Regel bei der Kontaktaufnahme oder zu Forschungsbeginn über den Kontext einer Erhebung (Ziele, involvierte Akteurinnen oder Ähnliches) informiert werden, technisch vermittelt und (a-)synchron. Dadurch haben die Teilnehmerinnen ein gewisses Maß an Selbstbestimmung und Kontrolle

darüber, inwieweit sie ihr Einverständnis zur Erhebung, Auswertung und Veröffentlichung bestimmter Daten geben (Heise & Schmidt, 2014).

Bei Big-Data-Analysen ist es nun möglich (oder angesichts hoher Fallzahlen sogar notwendig), auf einen kommunikativen Austausch mit den beforschten Individuen zu verzichten: „a researcher can conveniently collect data from his/her computer without the knowledge of the data producer or the ‘participant’“ (Neuhaus & Webmoor, 2012, S. 57). Das heißt, bei automatisierten Erhebungen findet eine Interaktion zwischen beiden Parteien kaum oder nicht mehr statt. Dies unterminiert zum einen ein basales kommunikationsethisches Prinzip: Reziprozität, also die Wechselseitigkeit der Kommunikation, die als eine Maßgabe der Interaktion zwischen Forscherin und Nutzerin im Online-Kontext verstanden werden kann (Wolff, 2007; Heise, 2013). Eine solche Wechselseitigkeit ist unter den Bedingungen von Big-Data-Forschung nicht mehr gegeben oder aufgrund der großen Anzahl involvierter Nutzerinnen gar unmöglich, was vor allem Folgen für das Einholen einer informierten Einwilligung (siehe Abschnitt 2) hat. Zum anderen hat sich der Austausch verlagert und findet nun vorrangig in Form eines Datenaustauschs zwischen Forscherin und Plattformbetreiberin statt. Durch derartig automatisierte Datenzugänge werden aber nicht nur die Nutzerinnen – als Produzentinnen der Daten – um- oder vielmehr übergangen. Ethisch problematisch ist es auch, dass die Nutzerinnen und ihr Handeln aufgrund der Nicht-Interaktion als Individuen „verschwinden“ und nur als maschinenlesbarer Text wahrnehmbar sind.

### 1.3 *Big Data: Bits und Bytes oder Spuren kommunikativen Handelns?*

Der Status und die Identität der Teilnehmerinnen sind bei Big-Data-Analysen weniger eindeutig als bei offline durchgeführter Forschung (Neuhaus & Webmoor, 2012). Dabei stellt die kommunikative Distanz der Forscherin zu den Nutzerinnen insofern eine forschungsethische Herausforderung dar, als sie zu einem gewissen Wahrnehmungsparadox führt: Die Forscherinnen haben es bei Big-Data-Analysen oftmals nur mit Datensätzen oder (Text-)„Objekten“ zu tun, wodurch die Nutzerinnen für sie vor allem als „einrollende Datensätze“ und nicht mehr als Einzelpersonen wahrnehmbar sind. In dieser Distanz zum betroffenen „actual human“ bestehe, so Zimmer (2012), einerseits das Risiko, dass Forscherinnen ihre Arbeit



nicht oder nur bedingt als „‘human subject‘ research“ begreifen, was die Verletzung oder Nicht-Einhaltung ethischer Standards im Umgang mit Forschungsteilnehmerinnen befördern könne.

Andererseits ist fraglich, inwieweit dabei das kommunikationsethische Prinzip der Personalität als eine Norm onlinebasierter Forschung Berücksichtigung findet (Heise, 2013). Bezugsnehmend auf das Grundrecht der Menschenwürde postuliert dieses Prinzip die Anerkennung des Personseins und das Verbot einer Verdinglichung, Objektivierung oder Instrumentalisierung anderer Nutzerinnen (Wolff, 2007). In dieser Hinsicht hält McFarland (2012) die Prozesse der automatisierten Datenerhebung und -verarbeitung im Rahmen von Big-Data-Analysen für ethisch bedenklich, „because it dehumanizes those being judged (...) and thus threatens those profiled objects as collections of facts, rather than as persons“. Auch die Masse von Daten könne dazu führen, dass Risiken für einzelne involvierte Nutzerinnen falsch eingeschätzt werden:

„the individual disappears in the mass, but the issue and potential risk remain on the scale of the individual. (...) [it is] important for both researchers and institutions to accept the fact that this kind of large-scale data mining still involves human subjects“ (Neuhaus & Webmoor, 2012, S. 58-59).

Dies verweist auf ein zentrales Problem im Zusammenhang mit Big-Data-Analysen: Und zwar die Frage, ob es sich bei den gewonnenen Daten um Spuren sozialer Kommunikation, also Handlungsakte, oder Artefakte/Texte handelt – eine Unterscheidung, die bei der Beurteilung ethischer Fragen hochgradig relevant ist. Während Big-Data-Analysen als Erhebungsmethode zwar nicht-reaktiv und Verfahren wie der Inhaltsanalyse durchaus ähnlich sind, untersuchen sie häufig nicht publizistische Inhalte, sondern Nutzer*handeln* („moments of people“), für das bestimmte Standards wie das Einholen der informierten Einwilligung gelten (Heise & Schmidt, 2014).

## **2    Forschungsethische Standards – wie umgehen mit den Menschen hinter den Daten?**

Wir gehen hier von der Prämisse aus, dass Big-Data-Analysen in Abhängigkeit zum Inhalt nicht mit Inhaltsanalysen gleichzusetzen sind, da es sich ins-

besondere bei Daten aus Social-Media-Diensten immer auch um Spuren sozialer Kommunikation handelt. Daher müssen auch ethische Standards berücksichtigt werden, die den Umgang mit Forschungsteilnehmerinnen regulieren und sich insbesondere auf deren Persönlichkeits- und Datenschutzrechte beziehen. Dazu gehören neben der *Freiwilligkeit* die *informierte Einwilligung* („informed consent“), die Wahrung der *Anonymität* und der *Schutz der Privatsphäre* sowie die *Nicht-Schädigung* der Teilnehmerinnen (Diener & Crandall, 1978; Hopf, 2009; Strohm Kitchener & Kitchener, 2009).

Ein zentraler ethischer Standard besagt, dass die Einwilligung zur Teilnahme an wissenschaftlicher Forschung freiwillig von Personen erteilt wird, die dafür kompetent sind und in der Regel im Vorfeld angemessen informiert wurden. Die informierte Einwilligung zählt zu den „primary technical and procedural steps for minimizing risks to human subjects in non-medical research contexts“ (Neuhaus & Webmoor, 2012, S. 45). Sie liegt einmal in der Freiheit des Individuums begründet, selbst darüber zu entscheiden, unter welchen Umständen es anderen Personen Einstellungen, Überzeugungen oder Meinungen mitteilt. Und zum anderen im Recht auf informationelle Selbstbestimmung, also der Kontrolle einer Person über die von ihr selbst mitgeteilten, über sie gesammelte bzw. sie betreffende, von anderen mitgeteilte Daten.<sup>4</sup>

Bei Big-Data-Analysen sind herkömmliche Consent-Verfahren angesichts der Größe der Datensätze und der Anzahl beteiligter Individuen oft nur schwer oder gar nicht umsetzbar. Adäquate Verfahren sind bislang nicht entwickelt. Ungeklärt ist zudem, wie etwa mit einzelnen Postings auf sozialen Netzwerkplattformen umgegangen werden soll, die nach der Datenerhebung von den Nutzerinnen wieder gelöscht wurden. Oder wie man beispielsweise mit Postings verfahren soll, in denen sich Personen nicht selbst äußern, sondern von anderen Nutzerinnen erwähnt werden. Derartige Inhalte können bei einigen Fragestellungen, etwa zu politischen Einstellungen, als durchaus sensibel eingestuft werden (Kaczmarek et al., 2014).

Da es zudem fraglich ist, ob die Kommunikation auf bestimmten Diensten per se als öffentlich einzustufen ist, sofern sie öffentlich zugänglich ist, können wir

4 Dabei handelt es sich laut Schmidt (2012) nicht nur um ein normatives Konzept, sondern auch eine Kompetenz im praktischen Handlungsvollzug.

nicht automatisch von einer impliziten Zustimmung der Nutzerinnen zur Sammlung und Auswertung ihrer Daten ausgehen, denn: „In light of current research (...), it is questionable whether users can effectively distinguish private from public messages and behaviour” (Mahrt & Scharrow, 2013, S. 26).

Ebenfalls nicht geklärt ist, ob die Zustimmung zu den AGBs einzelner Dienste als impliziter informed consent gewertet werden kann bzw. eine explizite Einwilligung „überlagert“ oder gar ersetzt. Die empörten Reaktionen auf das „Facebook-Experiment“ zeigen, dass diese Lesart aus Nutzerinnensicht problematisch ist und die Forscherinnen keinesfalls von der Umsetzung des informed consent entbinden sollte: „Individuals may (...) have agreed to the service conditions of a third party. However, this cancels neither the responsibilities of the researcher nor the institutions that undertake such research” (Neuhaus & Webmoor, 2012, S. 59).

Die bereits in Abschnitt 1.2 angesprochene Frage nach der Öffentlichkeit/Privateit bestimmter Daten oder Kommunikationsformen ist auch in Bezug auf die Standards der Anonymisierung und Nicht-Schädigung relevant (boyd & Crawford, 2012). Der forschungsethische Standard der Nicht-Schädigung postuliert die Vermeidung möglicher Risiken oder Gefährdungen. Das heißt die an der Forschung beteiligten Personen/-gruppen dürfen keinen Nachteilen oder Gefahren ausgesetzt werden, die das Maß des im Alltag Üblichen überschreiten, und müssen über etwaige Risiken aufgeklärt werden (Hopf, 2009). Zur Vermeidung möglicher Schäden, wie der Verletzung der Privatsphäre der Teilnehmerinnen durch die Weitergabe und mangelhafte Sicherung personenbezogener Daten oder unzureichend anonymisierte Publikationen, sind die Daten zu anonymisieren, insbesondere im Hinblick auf die Sensibilität bestimmter Informationen.<sup>5</sup>

Bei Big-Data-Analysen ist die Frage umstritten, was konkret unter persönlichen/sensiblen Informationen zu verstehen ist und welche Schädigungen für Individuen potenziell auftreten können. Einige Autorinnen argumentieren, dass

5 Der Datenschutz zählt zum rechtlich bindenden Kontext empirischer Forschung. So treffen das Bundesdatenschutz- oder das Telemediengesetz Regelungen zur Erhebung, Verarbeitung und Nutzung sowie sicheren Verwahrung von Daten. Demnach dürfen personenbezogene Daten besonderer Art, wie etwa ethnische Herkunft oder religiöse Überzeugungen, nur dann erhoben und verwendet werden, wenn sie zur Durchführung der Forschung unerlässlich sind.

über die Analyse oder Zusammenführung von Datensätzen völlig neue Formen persönlich identifizierender Informationen entstehen:

„By (...) analysing metadata, such as a set of predictive and aggregated findings, or by combining previously discrete data sets, Big Data approaches are not only able to manufacture novel PII [personally identifiable information; Anm. d. Verf.], but often do so outside the purview of current privacy protections.” (Crawford & Schultz, 2013, S. 94)

Individuen sollten daher das Recht haben, zu erfahren, wie persönliche Informationen konkret genutzt werden. Dies scheint umso drängender angesichts der Tatsache, dass Nutzerinnen dazu tendieren, über mehrere Dienste hinweg ähnliche „Identifizier“ zu verwenden. Über die (Re-)Kombination oder das „cross screening“ verschiedener Datensätze lassen sich so sekundäre Informationen generieren, über die Einzelpersonen identifizierbar sind: „This is especially the case if traces are available over a longer period of time or in greater quantity so that a more comprehensive picture about an individual may be pieced together“ (Neuhaus & Webmoor, 2012, S. 58). Diesbezüglich argumentieren Crawford und Schultz, dass das Sammeln von Informationen über das Verhalten Einzelner eine neuartige Form der Verletzung von Privatsphäre („predictive privacy harms“) darstellen und durchaus persönliche Schädigungen zur Folge haben könne: „Personal harms emerge from the inappropriate inclusion and predictive analysis of an individual’s personal data without their knowledge or express consent“ (Crawford & Schultz, 2013, S. 94).<sup>6</sup> Ein weiteres Problem bei der Umsetzung des Anonymisierungsstandards ist die potenzielle De-Anonymisierung von Datensätzen nach der Veröffentlichung (siehe Abschnitt 4.1).<sup>7</sup>

- 6 Ein Beispiel für unberechtigte Privatsphäreneingriffe ist der Fall Target: Das Unternehmen hatte mittels der Analyse verschiedener (Meta-)Daten schwangere Nutzerinnen identifiziert und deren Kontaktdaten ohne Zustimmung an Marketingfirmen weitergegeben. Einige Betroffene hatten ihre Schwangerschaft noch nicht öffentlich gemacht, wodurch es möglicherweise auch zu persönlichen Schädigungen kam (Crawford & Schultz, 2013).
- 7 Zimmer (2010) zeigte am Beispiel des US-amerikanischen Forschungsprojekts T3, dass sich selbst größere Datensätze de-anonymisieren lassen: Nachdem die auf Facebook gesammelten Daten eines ganzen Studierendenjahrgangs online gestellt wurden, konnte die betreffende Universität anhand des öffentlich einsehbaren Codebuchs der Studie identifiziert werden, womit die Anonymisierung des Datensatzes hinfällig war.

Insgesamt zeigt sich, dass forschungsethische Standards auch bei Big-Data-Analysen relevant sind, insbesondere im Kontext der Kommunikation auf sozialen Netzwerkplattformen. Und auch wenn deren konkrete Umsetzung noch ausgehandelt werden muss, scheint klar: „Concerns over consent, privacy and anonymity do not disappear simply because subjects participate in online social networks; rather, they become even more important“ (Zimmer, 2010, S. 324).

### **3 Was Big Data für Praxis und Grundwerte sozialwissenschaftlicher Forschung bedeutet**

Neben forschungsethischen Standards, die den Umgang mit Akteurinnen *außerhalb* der Wissenschaft, insbesondere den Untersuchungsteilnehmerinnen regeln, gehören auch normative Leitlinien zur Wertbasis wissenschaftlicher Tätigkeit, die das Handeln *innerhalb* der Wissenschaft steuern. Darunter fallen forschungstechnische Mindeststandards und Grundsätze wissenschaftlichen Arbeitens sowie höhere Prinzipien, die Forscherinnen bei der Beurteilung ihrer Arbeit und der ethischen Entscheidungsfindung im Hinblick auf grundlegende Werte, Ziele und Verantwortung von Forschung helfen sollen (Diener & Crandall, 1978; Strohm Kitchener & Kitchener, 2009; Fenner, 2010). Im Folgenden soll diskutiert werden, inwieweit Big-Data-Verfahren eine Herausforderung für die Praxis und grundlegende Werte sozialwissenschaftlicher Forschung darstellen.

#### *3.1 Big Data als Herausforderung für sozialwissenschaftliche Praxis*

In der konkreten Anwendungspraxis berührt Big-Data-Forschung einige Grundsätze guten wissenschaftlichen Arbeitens, wie sie etwa die Deutsche Forschungsgemeinschaft (DFG, 2013) formuliert, beispielsweise den Grundsatz der Nachvollziehbarkeit aller Forschungsschritte. Im Vergleich zu anderen Verfahren sind bei Big-Data-Analysen die weitgehend automatisierten Arbeitsschritte und algorithmische Modellierung, bei denen die Ergebnisse unter Berücksichtigung einer großen Anzahl von Parametern und Daten erlangt werden, zum Teil nur schwer und auch nicht für alle nachvollziehbar (Pietsch, 2013): „the advent of Big Data makes it easier for those with programming skills and knowledge of

social media APIs to ask social questions” (Tinati, Halford, Carr & Pope, 2014, S. 5). Derzeit verfügen nur wenige Forscherinnen über das notwendige Wissen oder die Ressourcen, um etwa die Qualität von Big-Data-Analysen beurteilen zu können. Dies kann ein Problem für innerwissenschaftliche Kontrollprozesse, wie peer feedback und die Beurteilung der Qualität des Vorgehens darstellen: „instead of (...) [an] information silo, we potentially have instead the information scientist silo“ (Neuhaus & Webmoor, 2012, S. 59).

Auch bei der Veröffentlichung von Ergebnissen sind der Offenlegung algorithmischer Datenerhebung und -auswertung aufgrund ihrer Komplexität und Datenmenge Grenzen gesetzt. Zwar werden, ähnlich wie bei Codebüchern zu Inhaltsanalysen, häufig Hashtag- oder Schlagwortlisten publiziert – die Abbildung des zugrunde liegenden Codes oder Algorithmus ist aber weitaus problematischer bis unmöglich. Ein aus ethischer Perspektive schwerwiegendes Problem bei der Veröffentlichung ganzer Datensätze ist zudem die potenzielle De-Anonymisierung, weshalb laut Mahrt und Scharrow (2013) nur noch wenige Anbieter zur Kooperation und Weitergabe anonymer Datensätze bereit seien. Neben datenschutzrechtlichen Problemen besteht beim (internationalen) Austausch und der Bereitstellung von Forschungsdaten ein nachträgliches Risiko darin, dass sich Content-Providerinnen oder einzelne Nutzerinnen gegen die Veröffentlichung ihrer Daten aussprechen: „The post-hoc withdrawal (...) makes replications of the findings impossible and therefore violates a core principle of empirical research” (Mahrt & Scharrow, 2013, S. 26; siehe auch Zeller, 2014).

Aus diesen und anderen Gründen sollten Big-Data-Forscherinnen als Teil guter wissenschaftlicher Praxis verstärkt Maßnahmen zur Gewährleistung von Transparenz, Sicherheit und „Accountability“ (Verantwortlichkeit/Zurechenbarkeit) ergreifen (Wen, 2012). Einerseits trägt dies zur Qualitätssicherung und Güte wissenschaftlichen Arbeitens bei. Andererseits können Maßnahmen, die Zurechenbarkeit und einen verantwortungsvollen Umgang mit den Daten (Stichwort: Datenschutz) anzeigen, das Vertrauen in die Forschung steigern. Bei der Veröffentlichung von Datensätzen und Grafiken könnte dies über Erklärungen zur Autorinnenschaft bei der Datenerhebung und -verarbeitung erfolgen: „This would both break the spell of anonymity (...) surrounding these data sets and assign a claimed responsibility“ (Neuhaus & Webmoor, 2012, S. 59).

### 3.2 Big Data und ethische Grundprinzipien

Neben der Umsetzung der Grundsätze wissenschaftlichen Arbeitens stellt sich auch die Frage, inwieweit Big-Data-Analysen ethische Grundprinzipien wissenschaftlicher Forschung berühren. Dazu gehören laut Strohm Kitchener und Kitchener (2009) die vier Dimensionen *Non-Maleficence* (Nicht-Schädigung/Schadensvermeidung), *Beneficence* (Fürsorge), *Justice* (Gleichheit/Gerechtigkeit) und *Fidelity* (Vertrauen), aber auch Grundwerte wie Forschungsfreiheit. Für uns sind hier zunächst das Gerechtigkeitsprinzip (das Ideal einer Gleichbehandlung aller) und das Prinzip der *Beneficence* interessant, welches besagt, dass wissenschaftliche Forschung nicht nur um das Wohlergehen aller bemüht sein soll, sondern auch die Autonomie und Selbstbestimmung sowie die Werte und Entscheidungen der Teilnehmerinnen anerkennen und respektieren soll.

Im Hinblick auf diese Prinzipien stellt sich bei Big-Data-Analysen zum einen die Frage, wer in den Datensätzen repräsentiert ist und welche Menschen davon ausgeschlossen sind, etwa weil sie bestimmte soziale Netzwerkplattformen nicht nutzen. Dabei geht es nicht nur um demographische Diskrepanzen, sondern auch um Unterschiede in der Nutzung einzelner Angebote, die die Generalisierbarkeit der Ergebnisse erheblich einschränken. So erlaubt ein Dienst wie Twitter unterschiedlichste Beziehungsgeflechte und eine breite Vielfalt an Kommunikationsmodi, von denen die Nutzerinnen in verschiedener Weise Gebrauch machen. Die Aussagekraft von Analysen einzelner Hashtags oder @-replies ist daher oft fraglich:

„users take advantage of these communicative facilities in different ways. Some may enjoy the increased ability to speak, whilst others may appreciate merely the wider array of content available to consume. (...) analysing only tweet data may not offer a fair overview of Twitter usership per se, let alone a national population more generally“ (Cows, 2014).

Angesichts dieser dateninhärenten „biases“ ist fraglich, inwieweit Big-Data-Forschung überhaupt den Grundprinzipien einer Gleichbehandlung und der Förderung des Wohlergehens aller genügen kann. Hier sollte auch in den Sozialwissenschaften eine Debatte zu den gesellschaftlichen Folgen von Big-Data-Analysen stattfinden, etwa zum Einfluss so genannter „predictive analytics“ (Wen, 2012)

auf politische Entscheidungen.<sup>8</sup> Zudem sollten Sozialwissenschaftlerinnen ein Bewusstsein dafür entwickeln, dass die Interpretation von Big-Data-Analysen, wie bei jeder anderen Forschung auch, gewisse Nachteile für gesellschaftliche Gruppen nach sich ziehen kann, zum Beispiel durch die Erzeugung oder Verstärkung von Stereotypen (Hopf, 2009). Letztlich stellt das algorithmenbasierte Erfassen, Auswerten und Darstellen von Daten immer auch eine Klassifikation sozialen Handelns dar, die durchaus (schwer abschätzbare) gesellschaftliche Folgen haben kann: „How we shape the social through our counting and classifying are highly political and ethical issues“ (Uprichard, 2013).

Neben dieser gesellschaftlichen Ebene spielen ethische Grundprinzipien bei Big-Data-Analysen auch auf individueller Ebene der Nutzerinnen eine Rolle, insbesondere im Hinblick auf die Wahrung und Anerkennung ihrer Autonomie. Dies betrifft nicht nur ihr Recht auf informationelle Selbstbestimmung (siehe Abschnitt 2), sondern auch die Machtverhältnisse zwischen Forscherinnen und Nutzerinnen, die bei Big-Data-Forschung hinsichtlich der Kontrolle und des Zugangs zu den Daten häufig verschoben sind: “researchers have the tools and access, while social media users as a whole do not“ (boyd & Crawford 2012, S. 673).

Die Frage nach Handlungsmacht betrifft aber nicht nur die Beziehung zwischen Forscherin und Beforschten, da bei Big-Data-Analysen eine dritte Partei hinzukommt: die „Besitzerinnen“ der Daten. Bei sozialwissenschaftlichen Studien handelt es sich hier meist um Dienste wie Twitter, Facebook oder Google, also kommerzielle Unternehmen, die als Intermediäre die zur Forschung genutzten Informationen in ihren Datenbanken bereithalten und (partiell) bereitstellen. Diese Mittler-Position der Intermediäre trägt nicht nur zur kommunikativen Distanz zwischen Forscherinnen und Nutzerinnen bei (siehe Abschnitt 1), sondern macht sie auch zu einer neuartigen Anspruchsgruppe ethischer Entscheidungsfindung, vor allem bezüglich der informierten Einwilligung (siehe Abschnitt 2).

Fraglich ist, inwieweit dadurch neue Abhängigkeiten entstehen und ein zentrales wissenschaftliches Grundprinzip, nämlich die innere und äußere Freiheit der Forscherinnen in der Definition ihres Untersuchungsgegenstandes, der Wahl

8 Dass Big-Data-Analysen durchaus realweltliche Folgen haben, zeigte sich 2013 in den USA: Dort erfolgte die Zuteilung von Notversorgungsgütern nach dem Hurrikan Sandy unter anderem auf Basis einer Analyse der Twitter-Kommunikation aus betroffenen Gebieten (Crawford, 2013).



ihres Forschungsweges sowie das Recht auf Veröffentlichung der Ergebnisse beeinträchtigt wird. Dies betrifft zum einen die Datenerhebung: Viele Plattformbetreiberinnen schließen die automatisierte Erfassung von Nutzerinneninformationen/ -interaktionen in ihren AGBs aus oder knüpfen sie im Sinne ihres „Hausrechtes“ an bestimmte Bedingungen. Facebook etwa fordert von Forscherinnen explizit, die betroffenen Nutzerinnen über Art und Zweck der gesammelten Informationen in Kenntnis zu setzen und ihre Zustimmung einzuholen (Heise & Schmidt, 2014). Mitunter sind Wissenschaftlerinnen daher auf Kooperationen mit Plattformbetreiberinnen angewiesen, die aber eher selten sind und, wie das „Facebook-Experiment“ zeigt, aus forschungsethischer Perspektive nicht immer glücklich verlaufen. Da die Unternehmen in der Regel auf datenschutzrechtliche Absicherung und den Schutz ihrer Geschäftsmodelle – zu denen auch Datenbanken gehören – bedacht sind, können solche Kooperationen zudem mit Einschränkungen für die Forscherinnen verbunden sein, beispielsweise bezüglich der Publikation von Daten (siehe Abschnitt 4.1).

Auch über die Art und Form der verfügbaren Daten haben die Forscherinnen häufig keine unmittelbare Entscheidungsgewalt, was die Bandbreite an möglichen Fragestellungen bzw. ihre Forschung generell einschränken kann: „the tools we use can limit the range of questions that might be imagined, simply because they do not fit the affordances of the tool. (...) This then introduces serious limitations in terms of the scope of research that can be done“ (Vis, 2013). Entscheidungen über die Gestaltung von Nutzungsoberflächen oder bestimmte Nutzungsoptionen werden von den Intermediären gemäß eigener Zielvorstellungen (wie Profit- oder Reichweitenmaximierung) implementiert und sind daher nicht wertneutral. Mit Blick auf das Prinzip der Forschungsfreiheit gilt es bei Big-Data-Analysen daher die Rolle der Plattformbetreiberinnen zu berücksichtigen und zugleich kritisch zu reflektieren, inwieweit sie den Möglichkeitshorizont der Forscherinnen (mit-)bestimmen.

#### 4 Ausblick

Im öffentlichen Diskurs um Big Data wurden zuletzt immer wieder eine gesellschaftliche Diskussion sowie politische Entscheidungen darüber eingefordert, für welche Bereiche derartige Analysen eingesetzt werden dürfen – und für

welche nicht. An diesen Diskussionen sollten auch wir Sozialwissenschaftlerinnen uns beteiligen: „We need to think about what it means to measure the social world and how our models of causality are constructed (...), how these counts and measurements are used and for whom? These answers are not trivial and social scientists need to be part of those conversations” (Uprichard, 2013). Die kontroversen Debatten um das „Facebook-Experiment“ machen deutlich, dass auch eine Reflexion der ethischen Implikationen und ein Verständigungsprozess über unseren Umgang mit Big-Data-Forschung stattfinden muss – gerade jetzt, da die Forschung mit „traceable information“ zögerlich, aber unumkehrbar Einzug in unsere Forschungsfelder erhält.

Dabei gilt es nicht, den Wert sozialwissenschaftlicher Big-Data-Forschung grundsätzlich anzuzweifeln, sondern „to examine exactly what is known through such totalizing inquiries (...), to query the discursive horizon they construct, as well as what vanishes beyond that horizon” (Baym, 2013). Zugleich gilt es auch den Nutzen von Big-Data-Analysen und ihren Beitrag zum Gemeinschaftswohl klar herausstellen, um den Eindruck zu vermeiden, Daten würden schlichtweg aufgrund ihrer Verfügbarkeit oder rein um des Datensammeln-Willens erhoben. Nicht zuletzt führen uns die Enthüllungen von Abhör- und Big-Data-Zugriffstaktiken verschiedener Regierungsinstitutionen und die dadurch erneut angestoßene Datenschutzdebatte einmal mehr vor Augen: Es ist wichtig, datenintensive Forschung klar vom auch nur ansatzweisen Verdacht der verdeckten Beobachtung oder „Überwachung“ abzugrenzen, um das gesellschaftliche Vertrauen in Wissenschaft nicht aufs Spiel zu setzen. Denn selbst wenn die kommerzielle Verwendung von Nutzungsdaten durch die Plattformbetreiberinnen fragwürdig erscheint, so beruht sie doch auf einem „exchange or transaction of mutual benefit between service providers and customers. However, for academics (...) who deploy such data, the ethical implications must be carefully evaluated“ (Neuhaus & Webmoor, 2012, S. 57).

Die Absicht des vorliegenden Beitrags war es, einige ethische Dimensionen und Herausforderungen von Big-Data-Forschung aufzuzeigen, um die Sensibilität für forschungsethische Fragen zu schärfen und die Annahme zu entkräften, diese Art der Forschung sei „disconnected from ethics“ (Neuhaus & Webmoor, 2012, S. 60). Da es sich insbesondere bei Social-Media-Analysen um die Untersuchung menschlich-sozialer Ausdrucksformen handelt, müssen auch in Big-Data-Studien sozialwissenschaftliche ethische Grundprinzipien greifen (Zeller, 2014).

Angesichts vieler offener Fragen und Handlungsunsicherheiten braucht es eine Verständigung innerhalb der Scientific Community zum Umgang mit ethischen Dilemmata: „Big Data research has to solve the problem of guaranteeing privacy and ethical standards while also being replicable and open to scholarly debate” (Mahrt & Scharnow, 2013, S. 26).

Die Ergebnisse einer solchen Aushandlung sollten bestenfalls in der Formulierung von Richtlinien münden, die zur Orientierung im Prozess ethischer Entscheidungsfindung, der wissenschaftlichen Selbstkontrolle sowie der Legitimation gegenüber Akteurinnen außerhalb der Wissenschaft dienen können. Aufgrund der Besonderheiten onlinebasierter Kommunikation stoßen Forscherinnen bei der Übertragung ethischer Standards bislang häufig an Grenzen und gängige formalisierte Verfahrensweisen erweisen sich oft als ineffektiv oder impraktikabel. Es braucht daher praxisorientierte Handlungsleitlinien, die flexibel für die jeweiligen Forschungsvorhaben adaptierbar sind. Neuhaus und Webmoor (2012) etwa schlagen eine „agile ethics for massified research and visualization“ vor, bei der es weniger um standardisierte Ethik-Protokolle, sondern um situativ anzupassende „working solutions“ und eine Sensibilität für gute Praxis geht. Letztlich tragen ein verantwortungsvoller Umgang mit den Daten sowie Transparenz und Zurechenbarkeit nicht nur zur Sicherung der Qualität und Güte wissenschaftlichen Arbeitens bei, sondern steigern auch das Vertrauen in Big-Data-Forschung.

Die im Beitragstitel aufgeworfene Frage „Big Data – small problems?“ muss angesichts der vielschichtigen forschungsethischen Herausforderungen verneint werden. Wieder einmal stehen Wissenschaftlerinnen vor der Aufgabe, technische Machbarkeit und ethische Vertretbarkeit ihres Handelns abzuwägen. Das Aufkommen einer „data-driven science“ sollte auch als Erinnerung an die besondere Verantwortung sozialwissenschaftlicher Forschung verstanden werden. Denn, wie Diener und Crandall bereits vor über dreißig Jahren anmahnten, „the social sciences can (...) have a tremendous impact on society, even to the point of revolutionizing our conceptions of human nature, society, and culture. (...) [We] must accept responsibility for its social consequences” (Diener & Crandall, 1978, S. 195).

*Nele Heise, M.A.* ist wissenschaftliche Mitarbeiterin an der Universität Hamburg und Doktorandin an der Graduate School Media and Communication Hamburg

## Quellenverzeichnis

- Baym, N. (2013). Data not seen: The uses and shortcomings of social media metrics. *First Monday*, 18(10). doi: 10.5210/2Ffm.v18i10.4873
- boyd, d. (April 2010). Privacy and Publicity in the Context of Big Data. Vortrag auf der WWW. Raleigh: North Carolina. Abgerufen von <http://www.danah.org/papers/talks/2010/WWW2010.html>
- boyd, d., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679. doi: 10.1080/1369118X.2012.678878
- Crawford, K. (1. April 2013). The Hidden Biases in Big Data. *Harvard Business Review*. Abgerufen von [http://blogs.hbr.org/cs/2013/04/the\\_hidden\\_biases\\_in\\_big\\_data.html](http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html)
- Crawford, K., & Schultz, J. (2013). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review*, 55(1), 93-128. Abgerufen von <http://lawdigitalcommons.bc.edu/bclr/vol55/iss1/4/>
- Cowls, J. (28. Januar 2014). "Twitter says..." – Can big social data tell us about public opinion? Abgerufen von <http://joshcowls.com/2014/01/28/twitter-says-can-big-social-data-tell-us-about-public-opinion/>
- Deterding, S. (9. Juli 2014). The Facebook Loophole. Abgerufen von <http://codingconduct.tumblr.com/post/91239582975/the-facebook-loophole>
- DFG (2013). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*. Weinheim: Wiley-VCH. Abgerufen von [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf)
- Diener, E., & Crandall, R. (1978). *Ethics in Social and Behavioral Research*. Chicago: University of Chicago Press.
- Eynon, R., Schroeder, R., & Fry, J. (2009). New Techniques in Online Research. Challenges for Research Ethics. *21st Century Society*, 4(2), 187-199. doi: 10.1080/17450140903000308
- Fenner, D. (2010). *Einführung in die Angewandte Ethik*. Tübingen: Francke.
- Heise, N. (2013). 'Doing it for real' – Authentizität als kommunikationsethische Voraussetzung onlinebasierter Forschung. In M. Emmer, A. Filipovic, J.-H. Schmidt, & I. Stapf (Hrsg.), *Echtheit, Wahrheit, Ehrlichkeit. Authentizität in der computervermittelten Kommunikation* (S. 88-109). Weinheim: Juventa.

- Heise, N., & Schmidt, J. (2014). Ethik der Onlineforschung. In M. Welker, M. Taddicken, J.-H. Schmidt, & N. Jakob (Hrsg.), *Handbuch Online-Forschung* (S. 532-553). Köln: Herbert von Halem.
- Hopf, C. (2009). Forschungsethik und qualitative Forschung. In U. Flick, E. Kardoff, & I. Steinke (Hrsg.), *Qualitative Sozialforschung* (S. 589-600). Reinbek: Rowohlt.
- Kaczmirek, L., Mayr, P., Vatrapu, R., Bleier, A., Blumenberg, M., Gummer, T., et al. (2014). Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter. *Gesis Working Papers* (31). Abgerufen von [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_arbeitsberichte/WorkingPapers\\_2014-31.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_arbeitsberichte/WorkingPapers_2014-31.pdf)
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790. doi: 10.1073/pnas.1320040111
- Mahrt, M. & Scharnow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, *57*(1), 20-33. doi: 10.1080/08838151.2012.761700
- McFarland, M. (2012). Ethical Implications of Data Aggregation. Abgerufen unter <http://www.scu.edu/ethics/practicing/focusareas/technology/internet/privacy/data-aggregation.html>
- Neuhaus, F., & Webmoor, T. (2012). Agile ethics for massified research and visualization. *Information, Communication & Society*, *15*(1), 43-65. doi: 10.1080/1369118X.2011.616519
- Nissenbaum, H. F. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*(1), 119-158.
- Pietsch, W. (2013). Big Data – über Chancen und Risiken einer neuen Wissenschaft des Komplexen. *Universitas*, *806*.
- Russel, G., & Kelly, N. (2002). Research as Interacting Dialogic Processes: Implications for Reflexivity. *Forum: Qualitative Social Research*, *3*(3). urn: urn:nbn:de:0114-fqs0203181
- Schmidt, J. (2012). Persönliche Öffentlichkeiten und informationelle Selbstbestimmung im Social Web. In J. Schmidt & T. Weichert (Hrsg.), *Datenschutz* (S. 215-225). Bonn: Bundeszentrale für politische Bildung.
- Strohm Kitchner, K., & Kitchner, R. F. (2009). Social Research Ethics. Historical and Philosophical Issues. In D. Mertens, & P. Ginsberg (Hrsg.), *The Handbook of Social Research Ethics* (S. 5-22). Thousand Oaks: Sage.

- Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big Data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology*, 48(4), 663-681. doi: 10.1177/0038038513511561
- Uprichard, E. (1. Oktober 2013). Big Data, Little Questions? *Discover Society*. Abgerufen von <http://www.discoverociety.org/focus-big-data-little-questions/>
- Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10). doi: 10.5210/2Ffm.v18i10.4878
- Wen, H. (21. Juni 2012). The Ethics of Big Data. *Forbes*. Abgerufen von <http://www.forbes.com/sites/oreillymedia/2012/06/21/the-ethics-of-big-data>
- Wolff, O. (2007). *Kommunikationsethik des Internets: eine anthropologisch-theologische Grundlegung*. Hamburg: Dr. Kovač.
- Zeller, F. (2014). Online-Forschung und Big Data. In M. Welker, M. Taddicken, J.-H. Schmidt, & N. Jakob (Hrsg.), *Handbuch Online-Forschung* (S. 453-480). Köln: Herbert von Halem.
- Ziegau, S. (2009). *Die Abhängigkeit der Sozialwissenschaften von ihren Medien. Grundlagen einer kommunikativen Sozialforschung*. Bielefeld: transcript.
- Zimmer, M. (2010). 'But the data is already public': on the ethics of research on Facebook. *Ethics and Information Technology*, 12(4), 313-325. doi: 10.1007/s10676-010-9227-5
- Zimmer, M. (26. Mai 2012). New media, new ethics. How social media-based research demands new attention to research ethics. Vortrag auf der 62. Jahrestagung der International Communication Association (ICA). Phoenix: Arizona. Abgerufen von <http://de.slideshare.net/michaelzimmer/new-media-new-ethics-ica-2012>