

## Big Social Media Data als epistemologische Herausforderung für die Soziologie

**Zusammenfassung:** ‚Big Data‘ ist als Herausforderung für herkömmliche sozialwissenschaftliche Forschungsmethoden derzeit Gegenstand eines kritischen Diskurses in der Soziologie. Epistemologische Grundlagen sozialwissenschaftlichen Arbeitens stehen ebenso zur Debatte wie der Stellenwert der Wissenschaftstheorie als Grundlage (sozial-)wissenschaftlichen Arbeitens und die Frage, wer mit Autorität über ‚das Soziale‘ spricht und dabei Gehör findet. Ermöglicht Big Data neue Erkenntnisse? Wie können diese Erkenntnisse methodologisch und forschungsethisch abgesichert werden? Im Beitrag wird argumentiert, dass die epistemologischen Fragen zu Big Data im Zusammenhang stehen mit den verschiedenen Kontingenzen, die sich in der alltäglichen Forschungsarbeit ergeben. Die alltäglichen Praktiken von Forschenden und ihre konkreten Probleme im Umgang mit Daten (vom Datenzugang bis zur Veröffentlichung von Ergebnissen) bedürfen einer größeren Aufmerksamkeit in der Diskussion um die Epistemologie von Big Data; diese Praktiken entfalten sich im Kontext von akademischem Stellenmarkt, Publikationspolitik, hohen Kosten des interdisziplinären Arbeitens und einer zunehmend wichtigeren Rolle von Social-Media-Plattformbetreibern. Big-Data-Forschung wird oft in (neuen) Disziplinen betrieben, die keine Tradition der Hinterfragung von Erkenntnis haben. Unbequeme, umständliche und zeitintensive Arten und Weisen der Erkenntnisproduktion – die beispielsweise nicht immer zu positiven Ergebnissen führen und angesichts des hohen Veröffentlichungsdrucks wie er in immer mehr Disziplinen herrscht nur schwer aufrechterhalten werden können – werden eingefordert, um den nötigen Raum für eine Reflexion der Epistemologie von Big Data zu schaffen. Die gegenwärtige Situation bietet die Chance ein Wissen um Epistemologie einzubringen und Anknüpfungspunkte zu identifizieren, an denen sich Soziologie und Big-Data-Forschung treffen können, um das Potential von Big Data durch eine kritisch reflektierte epistemologische Fundierung zu erschließen.

*“Big data are here to grow. The only way of tackling them is to know what you are or may be looking for. At the moment, such epistemological skills are taught and applied by a black art called analytics. Not exactly your standard degree at the university. Yet, so much of our well-being depends on it that it might be time to develop a philosophical investigation of its methods. Who knows, philosophers might have something to learn, but also a couple of lessons to teach.” (Floridi 2012: 437)*

## Einleitung

Dieser Beitrag beschäftigt sich mit sogenannter ‚Big Data‘-Forschung, vor allem der auf Social-Media-Daten basierenden, die als Herausforderung für herkömmliche Methoden derzeit Gegenstand eines kritischen Diskurses in der Soziologie – und den Sozialwissenschaften im Allgemeinen – ist. Im Besonderen geht es hier um die Probleme, die sich durch die implizite Annahme, aber auch die explizite Postulierung neuer Epistemologien ergibt. Es stehen damit derzeit nicht nur die epistemologischen Grundlagen sozialwissenschaftlichen Arbeitens selbst zur Debatte, sondern es geht auch um den Stellenwert der Philosophie als Grundlage sozialwissenschaftlichen Arbeitens und darum, wer mit Autorität über ‚das Soziale‘ spricht und dabei Gehör findet. Zum einen unter Bezug auf die Literatur, die sich, zum Teil unter dem Namen ‚Critical Data Studies‘ (Iliadis/Russo 2016), in Journals wie ‚Big Data & Society‘<sup>1</sup> und international ausgerichteten sozial-, medien- und kommunikationswissenschaftlichen Fachzeitschriften finden lässt und zum anderen aufbauend auf Ergebnissen eines Forschungsprojektes, in dem u.a. Interviews mit Social-Media-Forschenden geführt wurden, wird in diesem Artikel zudem gezeigt werden, wie die Debatte über die Epistemologie von Big Data nicht unabhängig geführt werden kann von Ökonomien der Aufmerksamkeit in der Entstehung forschungsbasierten Wissens. Die epistemologischen Fragen, die Big Data aufwirft, sind nicht allein theoretisch lösbar, sondern verwoben mit verschiedensten Problemen auf der Ebene von Praktikalitäten und Kontingenzen, die sich bei der alltäglichen Forschungsarbeit in verschiedensten etablierten und sich etablierenden Feldern und Fächern ergeben.

Auf der Basis von Big Data aus dem Internet treffen derzeit meist in der Informatik verankerte Disziplinen mit ‚datenintensiven‘ oder ‚computerintensiven‘ Methoden (z.B. Big Data Analytics, Data Science) Aussagen über soziale Phänomene. Inhaltsanalysen von Tweets, Facebook-Nachrichten oder anderen nutzergenerierten Inhalten ermöglichen es, etwa über das automatisierte Zählen und Vergleichen von Begriffen, Zusammenhänge in großen Textmengen freizulegen. Bei Netzwerkanalysen wird versucht, Muster und Zusammenhänge zwischen Inhalt und Nutzern

von Social-Media-Diensten zu identifizieren und zu analysieren. Der Name ‚Big Data‘ bezeichnet damit zum einen eine spezifische Herkunft der Daten: Meist sind nutzergenerierte Inhalte, beispielsweise von Social-Media-Plattformen wie Facebook oder Twitter, oder Transaktions-, Mess-, Meta- und andere Prozessdaten gemeint; gemeinsam ist ihnen, dass es sich um Daten handelt, die nicht dezidiert zu Forschungszwecken, also zum Beispiel durch eine Umfrage, erzeugt wurden. Zum anderen bezieht sich ‚Big Data‘ auf spezifische Methoden, die in der Analyse zur Anwendung kommen, auf die Analyse großer Datenmengen ausgerichtet sind und die visualisierte Darstellung von vormalig nicht sicht- oder analysierbaren übergreifenden Zusammenhängen anstreben.

Kritiker der Big-Data-Methoden dekonstruieren dieses Versprechen allumfassender Erklärungen, verweisen auf die Heterogenität von Entstehungskontexten und auf verschiedene epistemologische, methodologische, ethische und praktische Probleme. Zum Beispiel ergeben sich Schwierigkeiten durch den oft eingeschränkten Zugang zu proprietären Daten großer Plattformbetreiber (z.B. boyd/Crawford 2012; Bruns 2013). Nutzergenerierte Inhalte können zudem in den unterschiedlichsten Situationen und aus verschiedensten Motivationen heraus entstehen (vgl. z.B. Bastos et al. 2015). Zudem sind sie stark geprägt durch die jeweilige Social-Media-Plattform, auf der sie entstehen (Tufekci 2014), so dass zum Beispiel übergreifende, inhaltsanalytische Ansätze, wie sie etwa im Text-Mining zur Anwendung kommen, sehr sorgfältiger und differenzierter Prüfung in Bezug auf ihre Aussagekraft bedürfen.

Vermeehrt fordern nun Medien-, Kommunikations- und Sozialwissenschaftler/innen angesichts des ‚Aufschwungs‘ von computerintensiven Methoden die eigenen epistemologischen Grundlagen zu überdenken (Ruppert et al. 2013). Es stellt sich die Frage: Ermöglicht Big Data neue Erkenntnisse? Und wie können diese Erkenntnisse methodologisch, forschungsethisch und ontologisch abgesichert werden? Die Beantwortung dieser Fragen erfordert eine gemeinsame Auseinandersetzung zusammen mit denjenigen, die Big-Data-Forschung betreiben. Es geht dabei darum, das Potential von Big Data für das Verständnis von Technik und Gesellschaft zu erschließen und die Anknüpfungspunkte zu finden, an denen sich Soziologie und Big-Data-Forschung treffen können. Interviews und Beobachtungen von Social-Media-Forschenden (vgl. Kinder-Kurlanda/Weller 2014; Weller/Kinder-Kurlanda 2015) haben jedoch gezeigt, dass diejenigen, die Big-Data-Forschung unternehmen, oft dem Primat der Machbarkeit und den Ökonomien der Aufmerksamkeit folgen (müssen). Auf der anderen Seite sind sie, wenn sie die Potentiale der neuen Methoden ausnutzen wollen, die ja eben neue Erkenntnisse versprechen, einer mangelnden epistemologischen Fundierung der neuen Methodologien ausgesetzt. Diese ist im Rahmen traditioneller Publikationsformate und disziplinärer Anforderungen oft nur schwer zu artikulieren. Forschende kommen zu sehr unterschiedlichen Einschätzungen des Potentials von Big Data und problematisieren die mangelnde Klarheit über deren Aussagekraft (Weller/Kinder-Kurlanda 2015). Verschiedene Strate-

gien werden gefunden, um sich den Herausforderungen durch neue Methoden und technische Fähigkeiten zu stellen (Kinder-Kurlanda/Weller 2014).

### *Adressaten und Beitrag des Artikels*

Dieser Artikel liefert einen Beitrag zur kritischen Diskussion um Big Data unter besonderer Beachtung der konkreten Situationen einzelner Forschender, aus denen heraus viele der Arbeiten im Bereich Big Data entstehen. Fokussiert werden demnach das alltägliche Datenmanagement von Social-Media-Forschungsdaten, der Einfluss individueller akademischer Karrierestrategien und fachspezifische Publikations- und Förderungsformate. Dieses ‚doing‘ von Big Data ist situiert in einer Zusammenfassung wichtiger Kritikpunkte an Big Data in der Literatur. Schließlich wird für ein Zusammenbringen von Social-Media-Big-Data mit medien- bzw. techniksoziologischen Ansätzen argumentiert, um Big Data für soziologisch relevante Erkenntnisse zu erschließen.

### *Das Big-Data-Versprechen*

Ein großes Datenvolumen wird nicht mehr alleine als das definierende Charakteristikum von Big Data angesehen. Vielmehr ist Big Data gekennzeichnet durch eine Zugänglichkeit der Daten, die nahezu gleichzeitig mit den Ereignissen entsteht. Des Weiteren wird Big Data durch eine variable Struktur, einen Anspruch auf Vollständigkeit, einen hohen Detailgrad, die Möglichkeit Daten mit anderen Datenquellen zusammenzuführen und durch Flexibilität charakterisiert (vgl. z.B. Manovich 2011; boyd/Crawford 2012; Kitchin 2014). Es geht also um „volume, variety and velocity“ (Housley et al. 2014: 3).<sup>2</sup> Jenseits solcher den Daten inhärenten Attributen betonen prozessorientierte Definitionen in der Literatur die Prozesse des Sammelns, Kuratierens und Nutzens der Daten und erkenntnisorientierte Definitionen die Beziehungen von Menschen mit ihren spezifischen kognitiven Kapazitäten zu den Daten (Ekbja et al. 2016). Letztere Definitionen betonen, wie computerunterstützte Big-Data-Analysen Ergebnisse aus den Daten hervorbringen können, die durch menschliche Kognition alleine nicht mehr erkennbar sind.

Besonders die Analyse von (Big-)Social-Media-Daten erschließt neue Forschungsmöglichkeiten (Savage/Burrows 2007). Social Media sind „zu gewichtigen Quellensammlungen für die statistische Massenerhebung aufgestiegen“ (Reichert 2014: 449). Von besonderem Interesse für die Kommunikations-, Medien- und Sozialwis-

2 Verschiedene Bezeichnungen konkurrieren derzeit mit dem Begriff Big Data, die auf verschiedene Aspekte der Daten oder der in ihrer Analyse angewendeten Methoden hinweisen. Die Begriffe ‚Found Data‘ oder ‚Social Data‘ verweisen darauf, dass oftmals angenommen wird, die Daten seien „naturally occurring accumulations of user communication that are ‘found’ during the research process“ (Zelenkauskaitė/Bucy 2016, S. 2). ‚Data Analytics‘, ‚Data-Intensive Science‘ oder ‚Data Science‘ verweisen auf die große Bedeutung von Daten im Forschungsprozess, ohne diese Daten selbst näher zu definieren; gemeint sind jedoch meist digital erzeugte Daten, etwa aus dem Web.

senschaften sind nutzergenerierte Inhalte, also die Daten, die Nutzer von Plattformen wie Facebook, Twitter oder WhatsApp durch ihre Interaktionen mit den Programmen erzeugen. Dies können Nachrichten sein, aber auch andere Interaktionsdaten, wie zum Beispiel die Häufigkeit, mit der mit bestimmten anderen Nutzern interagiert wird, welche Inhalte die meisten ‚Likes‘ erhalten und so weiter. Jede Plattform bietet diverse Möglichkeiten der Interaktion und kann demnach theoretisch als Quelle für verschiedenste Daten über verschiedenste Nutzer dienen. Es gibt verschiedene Möglichkeiten als Forscher/in Zugang zu Social-Media-Daten zu erlangen. Plattformbetreiber haben zum Teil Daten für die Forschung geöffnet. Zum Beispiel stellen Facebook und Twitter Programmierschnittstellen bereit, die es Forschenden ermöglichen mithilfe von Skripten und Suchanfragen Webinhalte herunterzuladen. Andere Verfahren reichen von der manuellen Datensammlung über das ‚Crawlen‘ von Webinhalten mit dem Browser bis hin zum Kauf von Datensätzen.

Social-Media-Daten versprechen Einblicke in vormalig nur schwer beobachtbares Geschehen *in situ*, ohne die mühsamen Instrumente Umfrage oder Interview (Manovich 2011: 463). Möglichkeiten für zeitnahe Analysen und Prognosen, die mit herkömmlich erhobenen Daten und Forschungsmethoden nicht möglich sind, scheinen zum Greifen nah: Aktuelle Ereignisse werden auf Social-Media-Plattformen diskutiert, können zeitnah beobachtet und interpretiert werden und versprechen somit selbst globale Zusammenhänge auf der Basis der großen Datenmengen analysierbar zu machen. Big Data steht damit im Gegensatz zu der Situation, in der Daten immer nur über kleine Bruchteile des untersuchten Feldes zur Verfügung stehen und in der deduktive Ansätze in den Sozialwissenschaften große Bedeutung erlangen. Es eröffnen sich somit derzeit zahlreiche neue Möglichkeiten zur Beforschung menschlichen Verhalten in den verschiedensten Bereichen. Als Beispiele seien hier Wahlverhalten, Informationsverhalten oder soziale (Online-) Beziehungsnetzwerke genannt. Im Gegensatz zu herkömmlichen Datensätzen, die nach oft hoch strukturierten und tradierten Verfahren speziell für die Forschung erzeugt wurden, bedeutet die Analyse von Big Data jedoch auch, sich auseinanderzusetzen mit: „abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, and the fact that much of what is generated has no specific question in mind or is a by-product of another activity“ (Kitchin 2014: 2). Es ergeben sich damit zahlreiche methodologische, praktische und auch epistemologische Probleme. Bevor diese zusammengefasst werden soll jedoch kurz auf die jüngste Kategorie von Definitionen für das Phänomen Big Data eingegangen werden, die sich mit der sozioökonomischen, kulturellen und politischen Dimension des Phänomens beschäftigen.

## Die sozioökonomische, kulturelle und politische Dimension von Big Data

Nach diesen Konzeptualisierungen liegt das entscheidend Besondere und Neue von Big Data in seiner Wirkung. Gemeint sind sowohl die Wirkung der Forschungsergebnisse (Schroeder 2014), als auch die bestimmten sozioökonomischen, kulturellen und politischen Veränderungen, die das Phänomen erst ermöglicht haben und wiederum durch es möglich oder gar verstärkt werden (z.B. Crawford et al. 2014). Eine Perspektive auf diese Veränderungen macht Big Data als Teil der andauernden Fortentwicklung von Computern fassbar. Strategische Allianzen werden sichtbar, die sich um Big Data etwa zwischen Technologiefirmen, der Open-Source-Community und Universitäten formieren und in manchen Fällen z.B. bestimmten Universitäten Zugang zu neuesten Computertechnologien verschaffen (Ekbja et al. 2016). Ebenfalls wird der Blick gelenkt auf die Vision, die solche Unternehmungen oftmals vorantreibt: nämlich, dass das Potential von Big Data darin liegt unzählige neue Bereiche zu quantifizieren und analysierbar zu machen und damit verschiedenste Aspekte unseres Lebens grundlegend zu verändern (vgl. Mayer-Schönberger/Cukier 2013). Dieses Potential entspricht jedoch, wie auch schon bei früheren technischen Entwicklungen und ‚Trends‘ in der Computertechnologie, nicht unbedingt den tatsächlichen derzeitigen Möglichkeiten: „As with previous computerization movements, however, there are significant gaps between these visions and the realities on the ground“ (Ekbja et al. 2016: 6). In den Diskursen um Big Data vermischen sich, wie auch bereits bei ähnlichen Versprechungen etwa zu Ubiquitous-Computing-Technologien, Vergangenheit, Gegenwart und Zukunft zu einer ‚misleading vision‘ (Dourish/Bell 2011: 22) auf Kosten der Beachtung der ‚delightful messiness‘ (ebd. / S. 43) tatsächlicher gegenwärtiger Erfolge. Big Data ist damit auch dadurch gekennzeichnet, dass es von einem Narrativ von Neuheit und einem Bruch mit vorangegangenen Praktiken, Epistemologien und Methoden vorangetrieben wird, obwohl es sich eigentlich um eine Kombination verschiedener Elemente mit jeweils spezifischen Entstehungskontexten handelt, die gegenwärtig aufeinandertreffen (Rieder/Simon 2016).

Das Besondere und Neue an Big Data ist damit ebenfalls der Wahrheitsanspruch, der erhoben wird: „Numbers cease to be supplements. They are couched in a rhetoric of factuality, imbued with an ethos of neutrality, and presented with an aura of certainty.“ (Rieder/Simon 2016: 4) Dieser Anspruch auf Wahrheit wird durch die Automatisierung mit begründet, durch die „idea of mechanical objectivity“ (ebd.), auf die Big Data aufbaut; je automatisierter ein Prozess desto weniger Gefahr besteht laut dieser Idee durch die Ungenauigkeit subjektiver menschlicher Wahrnehmung (vgl. auch Fleisch/Tellkamp 2001). Big Data verspricht zudem die Reichweite der Automatisierung vom Datensammeln auf das Speichern, Kuratieren und Analysieren der Daten zu erweitern und die Notwendigkeit von Theorien, Modellen und menschlicher Expertise zu reduzieren (Rieder/Simon 2016).

Viele Autoren lenken den Blick auch auf die Bedeutung neuer ‚Player‘ wie etwa proprietäre Daten haltende Social-Media-Plattformbetreiber oder andere Internet-firmen. Big Data wird auch außerhalb der akademischen Forschung eingesetzt, besonders in Unternehmen. Damit werden in bislang nicht gekanntem Umfang weitreichende technische Entwicklungen größtenteils in den Händen der Privatwirtschaft gelassen (Ekbia 2016). Das Neue an Big Data liegt damit nicht in der schieren Menge der Daten, sondern in der Beachtung, die ihnen als Ware gezollt wird (Leonelli 2014).

### Neue Epistemologien?

Big Data stellt jedoch auch die epistemologischen Grundlagen von Forschung, davon wie wir Erkenntnisse gewinnen, grundsätzlich in Frage. Big Data „creates a radical shift in how we think about research“ (boyd/Crawford 2012: 665). Berry (2011) spricht vom Entstehen eines neuen ontologischen Zeitalters. Big Data bedeute damit die Entstehung eines neuen Forschungsparadigmas über verschiedene Disziplinen hinweg, einhergehend mit einer grundlegenden Änderung auf der Ebene von Epistemologie (Lazer et al. 2009). Big Data ermöglicht einen neuen epistemologischen Zugang zum Verstehen der Welt, denn es wird versucht aus den Daten selbst heraus Einsichten zu entwickeln (Kitchin 2014). Ruppert et al. (2013) wehren sich zwar gegen das Ausrufen eines ‚epochal change‘, fordern jedoch ebenfalls ein Umdenken in Bezug auf theoretische, ontologische und epistemologische Grundlagen sozialwissenschaftlicher Methoden angesichts von Big Data. Es gibt daher einen dringenden Bedarf an einer umfassenden und kritischen Reflexion der epistemologischen Folgen von Big Data und den mit diesen Daten verbundenen Methoden.

Kritische Stimmen weisen auf verschiedene methodologische, ethische, praktische und besonders epistemologische Probleme hin, die Big-Data-Methoden unzuverlässig und Ergebnisse fragwürdig erscheinen lassen (z.B. Lazer et al. 2014; Frické 2014; Kitchin 2014). Es wird bezweifelt, dass das Versprechen von Big Data, gerade für die Sozialwissenschaften, überhaupt eingelöst werden kann.

### *Ein Ende der Theorie?*

Besonders wird das von Anderson (2008) postulierte ‚Ende der Theorie‘ als Überhöhung kritisiert. Nach Anderson sei es demnach das Potential von Big Data, wissenschaftliche Theorie durch die schiere Menge der für Analysen verfügbaren Daten ersetzen zu können. Die Autorinnen boyd und Crawford warnten jedoch bereits 2012 vor einer Überhöhung des Potentials von Big-Data-Forschung um soziale Zusammenhänge zu beleuchten und zu verstehen; es sei ein Mythos, dass Big Data eine „higher form of intelligence and knowledge“ ermögliche (boyd/Crawford 2012: 663). Besonders boyd und Crawfords Interventionen waren „instrumental in moving scholarly engagement with ‘big data‘ from an unquestioned acceptance of

the data evangelism of interested parties towards a more critical, considered stance” (Bruns 2013). Ansätze, wie die von Anderson werden inzwischen oft verworfen: Big Data ist keinesfalls an sich bedeutungsvoll und Muster in den Daten bedürfen sehr wohl der Interpretation. Auch wenn bei Big Data die Samples größer, die Möglichkeiten zum Testen besser und die Ergebnisse kontinuierlich statt diskret verfügbar seien, bedarf wissenschaftliches Arbeiten trotz allem der Problemstellungen, Gedanken und Theorien: „If anything, science needs more theories and less data“ (Frické 2014: 10).

### Repräsentativität und Bias in den Daten

Vielfach wird auch auf Probleme von Verzerrungen oder Lücken in den Daten hingewiesen. Weder eine Repräsentativität der ‚Gesamtbevölkerung‘, noch derjenigen, die die Plattform nicht nutzen, ist gegeben; üblicherweise ist unklar, wie sich die Nutzer einer bestimmten Plattform zusammensetzen und zu anderen Bevölkerungsschnitten verhalten (Ruths/Pfeffer 2014). Oft ist jedoch auch fraglich, ob verfügbare Datensätze repräsentativ für die Nutzer der Plattform selbst sind. Zum Beispiel können automatisierte Posts oder Spam-Bots einen großen Teil der Inhalte ausmachen und zudem ist aufgrund der Einschränkungen der Programmierschnittstellen oft unklar, wie vollständig ein vorliegender Datensatz ist, oder welche Samplingmethoden der Plattformbetreiber bereits vorgenommen hat, je nachdem aus welcher für Forschende verfügbaren Quelle die Daten stammen (Driscoll/Walker 2014). Tufekci (2014) hat zudem die weit verbreitete Praxis des Samplings über abhängige Variablen kritisiert, wie sie zum Beispiel über das Zusammenstellen eines Datensatzes auf der Basis von verwendeten Twitter-Hashtags stattfindet.<sup>3</sup> Busch (2014) weist auf die Probleme hin, die sich aus der weitverbreiteten Praxis des ‚convenience sampling‘ und der in der Interpretation notwendigen Simplifizierung großer Datensätze zugunsten einiger und zu Ungunsten anderer Aspekte ergeben.

Ferner wird in vielen Ansätzen der Big-Data-Forschung nicht beachtet, dass das Interface vorstrukturiert, welche Möglichkeiten für Interaktionen es gibt. Systeme werden jedoch entworfen und implementiert, um ganz bestimmte Daten zu erzeugen und festzuhalten (Kitchin 2014). Basierend auf der These ‚raw data is an oxymoron‘ (Gitelman 2013) haben verschiedene Autoren darauf hingewiesen, dass die Daten, die für Big-Data-Analysen verwendet werden, nicht ‚neutral‘ oder eben ‚Rohdaten‘ sind, sondern immer geprägt sind durch den spezifischen Kontext ihrer Entstehung und damit verschiedenen Ballast mit sich führen.

3 Viele Programmierschnittstellen, über die Social-Media-Plattformen Inhalte für die Nutzung in der Forschung zur Verfügung stellen, machen es zum Beispiel relativ leicht, nach allen Nachrichten, die bestimmte Begriffe enthalten, zu suchen. Im Falle einer Suche nach Twitter-Hashtags führt damit eine von Nutzern selbst vorgenommene Kategorisierung der Tweets zur Inklusion in das Sample.

Social-Media-Daten bieten zudem keinesfalls unbedingt unmittelbaren Einblick in die Motivationen und Gedankenwelten der Nutzenden. Bereits Sherry Turkle (1997) wies auf die komplexe Rolle hin, die Selbstdarstellung im Internet spielt: Nutzer und Avatar sind oft sehr verschieden und unterschiedliche Spielarten der Performanz kommen zur Anwendung. Die Daten offenbaren damit nicht unbedingt die „Fantasien, Absichten, Motive, Meinungen und Gedanken der Menschen“ (Manovich 2014: 72). Social-Media-Daten sind vielmehr Schnittstellen zwischen den Menschen und der Welt, die nur einige Aspekte ihres wirklichen Lebens und ihrer Fantasien zeigen und auch Daten enthalten, die ein bestimmtes Image produzieren sollen (ebd.). Big Data kann damit nur begrenzt abbilden, dass menschliche Systeme komplex und paradox sind und Menschen sich auf unvorhergesehene Arten und Weisen verhalten (Kitchin 2014).

Die Analyse gesammelter Social-Media-Daten selbst ist ebenfalls immer durch philosophische oder epistemologische Vorannahmen beeinflusst, d.h. die Daten werden durch Messinstrumente, infrastrukturelle Gegebenheiten, Gewohnheiten in den Sammelpraktiken, oder sogar absichtliche Verzerrungen geprägt: „...a researcher’s subjective judgments can become deeply infused into a data set through sampling, data cleaning, and creative manipulations such as data masking“ (Ekbia 2016: 15). Auch automatisierte Prozesse und Algorithmen sind „imbued with particular values and contextualized within a particular scientific approach“ (Kitchin 2014: 5). Durch Sammeln und Speichern vorbereitete Big Data müssen dann erst noch bereinigt werden, üblicherweise wiederum mit automatisierten Tools und Skripten, um Analysen überhaupt zu ermöglichen. Dabei werden Entscheidungen darüber getroffen, welche Teile und welche Eigenschaften der Daten betrachtet werden. Diese Prozesse der Datengenerierung werden wenig dokumentiert und bleiben oft opak (Helles/Jensen 2013), so dass es zu einer “imperious opacity of data-driven approaches to science” (Ekbia 2016) kommt.

In den Daten gefundene Muster sind zudem nicht, wie in einigen Ansätzen der Big-Data-Forschung angenommen, inhärent bedeutsam: Korrelationen zwischen Variablen können zufällig sein oder kaum eine kausale Assoziation aufweisen. Die schiere Menge der Daten macht das Erkennen verschiedenster (oder gar beliebiger) Muster möglich und es besteht die Besorgnis, dass Muster erkannt werden, wo gar keine sind (boyd/Crawford 2012). Diese Gefahr wird verstärkt, wenn angenommen wird, dass die Daten für sich selbst sprechen können und ohne spezifische Kenntnis des Fachgebiets interpretierbar sind: “Put simply, whilst data can be interpreted free of context and domain-specific expertise, such an epistemological interpretation is likely to be anaemic or unhelpful as it lacks embedding in wider debates and knowledge“ (Kitchin 2014: 5). Die soziologisch interessanten Fragen, nämlich danach, warum bestimmte beobachtbare Muster auftreten, wer oder was dafür verantwortlich ist und was die ökonomischen oder sozialen Konsequenzen sind, werden daher oft entweder erst gar nicht bearbeitet oder die Bearbeitung findet ohne Rückgriff auf existierende Theorien und Erkenntnisse statt. Während Big-Data-Analysen

wichtige Erkenntnisse bieten können, bedeutet dies jedoch, dass sie immer nur von begrenzter Aussagekraft sind, bestimmte Arten von Wissen erzeugen und immer der Interpretation im Kontext existierender Theorien bedürfen (Kitchin 2014). Big-Data-Forschung muss folglich mit anderen Ansätzen komplementiert werden. Ethnologaf/innen und Informatiker/innen beispielsweise haben Zugang zu unterschiedlichen Arten von Daten, stellen daher unterschiedliche Fragen, nehmen unterschiedliche Muster wahr und gelangen zu unterschiedlichen Erkenntnissen (Manovich 2014).

Vermehrt wird inzwischen auch auf die Tendenz zu Profiling und zu Vereinfachungen auf der Basis etwa ethnischer oder sozialer Stereotypen (Lyon 2007) in Big-Data-Arbeiten hingewiesen. Oft reflektieren die Analysen verbreitete Vorurteile (Barocas/Selbst 2016). Zurückführen lässt sich dies sicher zu einem großen Teil darauf, dass es intuitiv nahe liegt, einen großen Datensatz anhand beispielweise der Herkunft oder des Geschlechts der Social-Media-Nutzenden zu analysieren wenn diese Informationen zur Verfügung stehen – ohne dass dies beim jeweiligen Thema unbedingt sinnvoll ist oder überprüfbar wäre, ob die Daten in Bezug auf diese Merkmale repräsentativ sind. Es steigt damit die Gefahr der Verstärkung von Essentialismen – und eine Kritik oder Überprüfung der Methoden ist aufgrund weiter unten ausgeführter Probleme mit dem Zugang zu den verwendeten Daten zu Replikationszwecken oft schwierig bis unmöglich.

Das epistemologische Problem von Big Data ist damit nicht, dass es zu viele oder zu große Datenmengen gibt und kann auch nicht durch schnellere oder machtvollere Technologien gelöst werden (Floridi 2012). Probleme liegen vielmehr darin, herauszufinden, wo in den Daten tatsächlich neue, interessante Muster existieren und wie diese am besten genutzt werden können (Floridi 2012). Diese Muster können zwar ohne computergestützte Methoden nicht erkannt werden, jedoch nicht alleine durch diese: „...the game will be won by those who “know how to ask and answer questions” (Plato, Cratylus, 390 c) and therefore know which data may be useful and relevant...” (Floridi 2012: 437). Der Wert der Theorie liegt dann wiederum in der Möglichkeit, Informationen über widerspenstiges und anders nicht zu erfassendes Beobachtbares zu erlangen und zu strukturieren. Einzelne Beobachtungen können so auf höheren Abstraktionsbenen interpretiert werden (Ekbia 2016).

### Konkurrierende epistemologische Ansätze

Jenseits des ‘already tired binary of big data—is it good or bad?’ (Crawford et al. 2014: 1665) haben sich verschiedenste oft gegensätzliche Ansätze der Big-Data-Forschung entwickelt. Besonders stehen sich (meist medien-, kommunikations- oder sozialwissenschaftliche) Ansätze, die Big Data verwenden, um neue Fragen zu stellen und darauf drängen, Begriffe und Techniken zu hinterfragen und weiterzuentwickeln (Crawford et al. 2014: 1665) und diejenigen Ansätze gegenüber, die, oft eher ergebnisorientiert mit informatischen oder anderen technischen bzw. naturwis-

senschaftlichen Ansätzen operieren und für die die Beschäftigung mit menschlichem Verhalten nicht Teil des tradierten disziplinären Repertoires ist. Zu unterscheiden ist zudem die akademische Big-Data-Forschung, deren Ziel das Erzeugen von (generalisierbaren) Forschungsergebnissen ist, von der angewandten Forschung mit kommerziellen Zielen, die Big Data profitabel und für die Beeinflussung menschlichen Verhaltens nutzbar machen möchte (Schroeder 2014). Für angewandte Forschung mag es in vielen Fällen ausreichen, dass Korrelationen zumindest auf eine profitable Art und Weise funktionieren, es stellt sich jedoch die Frage, wie erfolgreich solche Strategien tatsächlich sind, denn „...while data can be used to target specific individuals, it may not be possible to change the behaviour of individuals (even if it is possible to ‘nudge’ them)” (Schroeder 2014: 8).

Die Debatte um die Bedeutung von Big Data für die Wissenschaft kann historisch eingeordnet werden: Die Wissenschaftsgeschichte ist gekennzeichnet durch Umbrüche; immer wieder wurden in konflikthaftern Prozessen frühere Kriterien für die Erzeugung und Vollständigkeit wissenschaftlicher Erkenntnis durch neue abgelöst. Ekbia et al. (2016) sehen die Unterscheidung zwischen Kausalzusammenhang und Korrelation im Zentrum gegenwärtiger Diskussionen um Big Data. Diese Debatte knüpft an eine langjährige Auseinandersetzung zwischen theorie- und datengeleiteter Wissenschaft an. Besonders problematisch zeigt sich dabei eine Big-Data-Forschung, die auf Ideen des ‚understanding by building‘ operiert (Ekbia et al. 2016). In den Ingenieurs- und Technikwissenschaften und zu einem großen Teil auch in der Informatik werden Systeme nach vorher festgelegten Spezifikationen gebaut. Wenn diese Systeme auch oft funktionieren, so bedeutet dies jedoch nicht, dass umgekehrt jedes funktionierende System notwendigerweise auf akkuraten Theorien beruht oder akkurate Repräsentationen bereitstellt. Ekbia et al. (2016) bezeichnen dies als ein zentrales Dilemma des 20. Jahrhunderts: „Letting go of mechanical explanations, some radical perspectives on Big Data seek to save the phenomenon by simply saving the appearance. In so doing, they collapse the distinction between the two: phenomenon *becomes* appearance” (ebd. / S. 7). Big Data scheint eine Verschiebung im Sinne der wissenschaftlichen Erfolgskriterien von kausalen Erklärungen zu vorhersagenden Modellen und Simulationen bewirkt zu haben. Diese Verschiebung wird hauptsächlich von Kräften außerhalb der Wissenschaft betrieben und setzt den im 20. Jahrhundert begonnenen Trend des Entkoppelns von Phänomen und Erscheinung (appearance) fort (Ekbia 2016) – es geht nur noch um die Vorhersage der Erscheinungen (nicht mehr um ein Explorieren der Gründe) und wir werden mit ‚structuralism run amok‘ (Ekbia 2016: 8) konfrontiert.

### *Lösungsansätze – Philosophische Big Data?*

Um die methodologischen Probleme von Big-Data-Analysen zu lösen, wird vorgeschlagen, dem Explorieren Raum zu geben. „Multi-method, multi-platform analyses“ (Tufekci 2014), „experimentation“ (Ruppert 2013) und „kludginess“ (Karpf 2012) werden gefordert. Die Sozialwissenschaften sollen sich den Herausforderun-

gen durch die neuen Daten und Methoden stellen und zudem interdisziplinäre Zusammenarbeiten eingehen (Savage/Burrows 2007; Ruppert 2013). Die technischen Fähigkeiten und die Programmierkenntnisse der Informatiker/innen könnten sich so mit fundierten Kenntnissen sozialwissenschaftlicher Methoden und analytischer Interpretation ergänzen (Tinati et al. 2014). Die epistemologische Strategie, die in der Big-Data-Forschung zur Anwendung kommen könnte, wäre damit eine, die wissens- und theoriegeleitet potentielle Fragen oder Hypothesen identifiziert, die einer weitergehenden Untersuchung wert sind.

Neben explorativen, interdisziplinären Epistemologien werden auch neue Modelle notwendig, um Partizipation abzubilden. Sieber und Haklay (2015) weisen am Beispiel von Geoinformationssystemen darauf hin, dass in der Forschung mit nutzergenerierten Internetdaten häufig auf der Basis einer Epistemologie gearbeitet wird, die von der Erzeugung der Daten durch Experten ausgeht; etwa verschiedene Grade gewollter und ungewollter Partizipation an der Erzeugung dieser Daten werden oft nicht berücksichtigt.

Eine Epistemologie für Big-Data-Forschung mit Social-Media-Daten muss sich zudem der spezifischen Blindheiten der Forschung bewusst sein, sich auf existierende, auch kritische soziale Theorien beziehen und anerkennen, dass Forschung nicht neutral oder objektiv ist, sondern immer von einem bestimmten Standpunkt aus unternommen wird, und zudem, dass Datensätze, Forschung und Interpretationen immer auch von politischen oder strategischen Überlegungen durchzogen sind (Kitchin 2014). Eine solche Epistemologie für Big Data ermöglicht die Komplementierung von Big-Data-Analysen beispielsweise mit Ergebnissen aus qualitativen Analysen.

Tatsächliche Forschungsprojekte, die einer solchen Epistemologie folgen, sind jedoch derzeit selten. Interdisziplinarität, Exploration und die Kombination von kritischen, soziologischen Ansätzen mit computerintensiven Methoden sind zwar wünschenswert, im Rahmen der tatsächlichen Arbeitssituationen, in denen sich viele Social-Media-Forschende derzeit befinden, jedoch nur selten umsetzbar. Im Folgenden sollen auf Basis einiger Ergebnisse aus Interviews und Beobachtungen von Social-Media-Forschenden konkrete epistemologische Praktiken beleuchtet werden. Diese Praktiken werden in die Literatur eingeordnet, die die verschiedenen methodischen, ethischen und praktischen Probleme der Big-Data-Forschung bereits thematisiert. Ich werde jedoch zeigen, dass die Probleme über diese Thematisierung hinaus nicht allein auf der Basis individueller Anstrengungen einzelner Forschender lösbar sind.

## Konkrete Praktiken

In einer interviewbasierten Studie von Forschenden, die Social-Media-Daten verwenden, wurden die alltäglichen Praktiken des akademischen Arbeitens, vom Konzeptionieren eines Forschungsprojektes über die Erhebung von Daten bis zur Veröf-

fentlichung und die Zusammenarbeit mit anderen Forschenden betrachtet. Einige Ergebnisse der Interviewstudie wurden in detaillierterer Form bereits anderswo dargestellt (Kinder-Kurlanda/Weller 2014; Weller/Kinder-Kurlanda 2015). Ziel der Studie war eine Exploration der Herausforderungen, die sich für Forschende derzeit in der Arbeit mit Social-Media-Daten ergeben. Durchgeführt wurden 42 Interviews mit und zahlreiche Beobachtungen von Forschenden auf verschiedenen internationalen Konferenzen der Internetforschung. Es wurden nur Forschende befragt, die bereits mit von Nutzenden auf Social-Media-Plattformen wie Twitter oder Facebook erzeugten Daten gearbeitet hatten. Fast alle Befragten arbeiteten an Universitäten oder öffentlich geförderten Forschungseinrichtungen in Europa, den USA oder Australien. Es wurden Forschende verschiedenster Karrierestufen befragt (Master-Studenten, Doktoranden, Postdocs, Professoren etc.). Die Interviews stützten sich auf einen Leitfaden, der möglichst offen gehalten war entsprechend eines "thematic guide with probes and invitations to expand on issues raised" (May 2011) und während der Befragung iterativ angepasst wurde. Neben theoretischen Überlegungen zur Bedeutung von Big Data für epistemologische Diskussionen innerhalb der Sozialwissenschaften sollten die tatsächlichen Praktiken und Situationen einzelner Forschender betrachtet werden, die diese zur Anwendung bringen.

Forschende waren derzeit aus verschiedenen, zum Teil hier dargestellten Gründen oft nicht in der Lage, auf der Basis von epistemologischen Grundlagen zu arbeiten, wie sie etwa Kitchin (2014), Schroeder (2014), Ekbja et al. (2016) oder die Critical Data Studies einfordern, nämlich situativ, reflektiert, interdisziplinär, theoriegeleitet, explorativ und gleichzeitig neueste computerintensive Methoden nutzend. Dies soll hier keineswegs entschuldigt werden, es ist jedoch zu fragen, wie Praktiken des wissenschaftlichen Arbeitens geschaffen werden können, die den dargelegten epistemologischen Überlegungen entsprechen und die auch das Nachdenken über Epistemologie an sich ermöglichen. Die Forschenden, die Big-Data-Analysen durchführten, problematisierten oft, dass sie ihre eigenen Methoden als unzureichend empfanden. Methoden zu finden, die dem veränderlichen Feld angepasst sind, ist auf der Ebene von Big-Data-Analysen schwierig, erfordert es doch gerade Partikularität und Beschäftigung mit Details und keine umfassenden, übergreifenden Ansätze. Am häufigsten wurden jedoch verschiedene Probleme, die sich aus anderen Eigenschaften der Daten als deren Größe ergaben thematisiert: 1.) der oft schwierige Zugang zu den Daten, 2.) die ethischen Probleme in deren Nutzung, 3.) das notwendige technische Spezialwissen. Über diese datenpraktische Dimension hinaus zeigte sich, dass 4.) Herausforderungen des interdisziplinären Arbeitens, 5.) derzeitige Publikationsformate und Praktiken und 6.) die Probleme der akademischen Karriereplanung in einem neuen Forschungsfeld besonderen Einfluss auf die Herangehensweisen hatten.

## Datenzugang

Die für Nachvollziehbarkeit und Validität der Forschung notwendige Transparenz über die Art und Weise der Datenerhebung herzustellen ist aus verschiedenen Gründen schwierig, allen voran aus den Problemen, die sich beim Zugang zu Social-Media-Daten ergeben. Bei den Interviewten existierten starke Ungleichheiten in Bezug auf die Menge und Qualität an Social-Media-Daten zu denen sie Zugang hatten: Forschende die über genügend finanzielle Mittel verfügten, um Daten zu kaufen<sup>4</sup> oder deren Institutionen Kollaborationen mit der Industrie ermöglichen konnten, hatten Zugang zu großen Mengen oft auch proprietärer Daten, manchmal auch zu solchen Daten sozialer Netzwerke, die üblicherweise nicht für die akademische Forschung zur Verfügung standen. Diese Daten wurden als qualitativ hochwertiger angesehen als solche, die über von Plattformanbietern für die Forschung bereitgestellten Programmierschnittstellen (sogenannten APIs) oder anders händisch gesammelt worden waren. Gerade Forschende außerhalb der USA oder an kleineren Institutionen waren eher darauf angewiesen, Daten selbst zu sammeln und mussten die hierfür notwendigen Kenntnisse erwerben. Beim Thema Zugang wurden so etablierte Ökonomien der Aufmerksamkeit, durch die der internationale Wissenschaftsbetrieb ohnehin gekennzeichnet ist, reproduziert oder sogar verstärkt (Weller/Kinder-Kurlanda 2015). Die Ungleichheiten wurden, wie auch von Ruths und Pfeffer (2014) beschrieben, als Bedrohung angesehen: "The rise of 'embedded researchers' (researchers who have special relationships with providers that give them elevated access to platform specific data, algorithms, and resources) is creating a divided social media research community." (ebd. / S. 1063)

Auf der anderen Seite existierte jedoch auch das Argument, dass der Datenzugang zu einfach sei. Social-Media-Daten seien "very available compared to ethnographic research" sagten etwa Befragte in der Interviewstudie. Oftmals wurden Studien sogar, wie auch von Tufekci (2014) kritisiert, durch Datenverfügbarkeit motiviert. Viele Forschende äußerten daher die Sorge, dass opportunistische Ansätze zur Anwendung kämen (vgl. Bruns 2013) und Forschende mit Daten arbeiteten, die verfügbar waren, und nicht mit solchen, die am geeignetsten waren um eine bestimmtem Forschungsfrage zu beantworten.<sup>5</sup>

## Ethische und rechtliche Fragen

Privacy und Datensicherheit (Ohm 2010; Zimmer 2010) werden im Allgemeinen als die hauptsächlich ethischen Herausforderungen in der Verwendung von

4 Twitter-Daten zum Beispiel können über verschiedene Anbieter, die mit Twitter zusammenarbeiten, gekauft werden.

5 Vgl. hierzu auch Tufekcis (2014) Konzeptualisierung von Twitter als „Modellorganismus“ der Social-Media-Forschung. Die Plattform ist in der Forschung lange Zeit im Vergleich zu ihren tatsächlichen Nutzungszahlen stark überrepräsentiert gewesen und hat sich daher trotz eingeschränkter Bedeutung für Nutzende zu einer Art prototypischer Plattform in der Forschung entwickelt.

Social-Media-Daten angesehen: Social-Media-Nutzende haben keine informierte Einwilligung im klassischen Sinne unterzeichnet, um zuzustimmen, dass ihre Daten für die Forschung verwendet werden können und die Anonymisierung der Daten ist schwierig oder unmöglich. Es zeigte sich in der Befragung, dass sich Forschende in einer äußerst unklaren forschungsethischen Situation sahen – so war die rechtliche Lage in Bezug auf das Teilen von Daten oft ebenso unklar wie die Ansichten von Social-Media-Nutzenden zur Forschung über die von ihnen erzeugten Inhalte (Weller/Kinder-Kurlanda 2014; Kinder-Kurlanda/Ehrwein Nihan 2013).

### Fehlende technische Fähigkeiten?

Wie oben angesprochen schränkt auch die Notwendigkeit, bestimmte technische Fähigkeiten zu besitzen, um Zugang zu Big Data zu bekommen, die Anzahl der Personen ein, die Zugang zu Big Data haben. Die Bedienung von APIs oder das Crawlen von Webinhalten erfordern zumindest grundlegende Programmierkenntnisse. Forschende bestätigten, dass ein Mangel an technischen Kenntnissen und Programmierkenntnissen die Möglichkeiten zur Durchführung von Social-Media-Daten basierter Forschung wesentlich bestimmte: "My questions are limited to what I can do." Da die Analyse von Big Data spezialisierte technische Fähigkeiten erfordert gehen manche Autoren, zusätzlich zu den Schwierigkeiten mit der Verfügbarkeit proprietärer Social-Media-Daten für die Forschung, von einer ‚Scholarly Divide‘ (Zelenkauskaite/Bucy 2016) aus. Demnach kann das Potential von Big Data bisher nur von denjenigen realisiert werden, die über ausreichend technisches Know-How, finanzielle Möglichkeiten oder institutionelle Verbindungen verfügen. Die große Menge der Forschenden bliebe demnach außen vor. Diese Sichtweise negiert jedoch zum einen die zahlreichen Forschungsprojekte, die wenn auch nicht Big Data so zumindest Medium- oder Small-Data-Forschung (um bei der Metapher zu bleiben) mit Internetdaten betreiben. Zum anderen zeigte die Interviewstudie, dass es statt einer ‚Divide‘ vielmehr graduelle Unterschiede im Zugang zu Social-Media-Daten gibt, abhängig von den von Zelenkauskaite und Bucy (2016) genannten Faktoren (technisches Know-How, finanzielle Möglichkeiten oder institutionelle Verbindungen) in Kombination mit dem Einfluss weiterer Faktoren wie etwa persönlicher Netzwerke (z.B. wenn Daten befreundeter Forschender genutzt werden können) oder geografischer Verortung (z.B. wenn Daten aufgrund von Datenschutzgesetzgebungen mehr oder weniger verfügbar sind). Drittens verdeckt der Fokus auf eine Scholarly Divide auch die Frage danach, ob das Versprechen von Big Data wirklich so groß und unwälzend ist, wie angenommen. Die Qualität der Big-Data-Forschung wird zugunsten der Forderung nach einem besseren Zugang zu den Daten nicht in Frage gestellt. Das heißt auch, die Fragen, ob bestimmte Daten überhaupt für die Forschung verwendet werden sollten oder ob Ergebnisse tatsächlich aussagekräftig im Sinne neuer Forschungserkenntnisse sind, werden nicht gestellt. Zudem erlaubt ein technisches Skillset, wie es für den Abruf von Daten über APIs oder durch das Crawlen von Webseiten notwendig ist, keine Kompeten-

zen in der Interpretation oder Theoriebildung, die genauso wichtig sind. Denn während Programmierkenntnisse als begehrenswert angesehen wurden, sahen viele der im Projekt Befragten das Problem, dass sie diese nur auf Kosten anderer Tätigkeiten wie etwa Literaturrecherche oder detaillierter Betrachtungen des Datenmaterials erwerben konnten. Ein Befragter fasste es so zusammen: "I think you would be able also as a social science researcher to learn that [Python programming] stuff, but then you don't have time to theorize." Kompromisse mussten geschlossen werden. Fortgeschrittene Forschende konnten das Sammeln, Bereinigen und Analysieren etwa durch die Mithilfe von Doktoranden oder Hilfskräften bewältigen – waren damit aber auch wiederum auf deren Spezialwissen zu den spezifischen Datensätzen angewiesen.

### Interdisziplinarität als Lösung?

Ein Weg, um zu verwirklichen, dass Ergebnisse von Big-Data-Analysen Interpretation im Kontext existierender Theorien erfahren, ist die interdisziplinäre Zusammenarbeit. Interdisziplinarität ermöglicht zudem auch Forschenden, die wenig technische Fähigkeiten besitzen, Zugang zu Social-Media-Daten zu erhalten. Ein gleichzeitiges Ausschöpfen des Potentials rechenintensiver Methoden, Algorithmen, Modelle und fundierter Theorie-Kenntnisse scheint möglich. Die interdisziplinäre Zusammenarbeit ist jedoch in der Praxis oft schwierig bis unmöglich, z.B. wenn Grundannahmen über die Ziele der Forschung nicht in Übereinstimmung gebracht werden können (vgl. Kinder-Kurlanda 2014). Es gibt zudem nur wenige institutionalisierte Möglichkeiten zur Zusammenarbeit zwischen denjenigen, die ‚close readings‘ und denjenigen die Big-Data-Forschung unternehmen. In der Studie beobachtete Forscher/innen aus der Computational Social Science zum Beispiel, die keinen sozialwissenschaftlichen Hintergrund hatten, sahen sich oft eher als Konkurrenz zu den ‚traditionellen‘ Sozialwissenschaften. Sozialwissenschaftler/innen selbst wiederum hatten oft keine Kenntnisse in Methoden, wie etwa dem Data Mining, die in den Ausschreibungen der Stellen in der Big-Data-Forschung gefordert wurden und fanden so keinen Zugang zu entsprechenden Forschungsgruppen. Sowohl Drittmittel-Fördermöglichkeiten als auch der akademische Stellenmarkt wurden zudem Interdisziplinarität gegenüber nicht immer als aufgeschlossen wahrgenommen, so dass das Verfolgen neuer, interdisziplinärer Forschungsdesigns mit hohen beruflichen Risiken verbunden war – Forschende operierten oft, trotz großer Erfolge, aus Positionen extremer Prekarität heraus (Kinder-Kurlanda/Weller 2014). Dies galt besonders für Medien-, Kommunikations- und Sozialwissenschaftler/innen, die, anders als ihre Kolleginnen aus der Informatik, die Beschäftigung mit Social Media oder Big Data auch oft als nachteilig für ihre Wettbewerbschancen auf dem akademischen Stellenmarkt ansahen, besonders wenn neue Methoden zur Anwendung gebracht wurden, die sich außerhalb des disziplinären Mainstreams befanden.

## Probleme der Veröffentlichungspraxis

Abhängig vom Fach ergaben sich für Forschende verschiedene Schwierigkeiten bei der Publikation von Ergebnissen Social-Media-basierter Studien in den Mainstream-Publikationskanälen der jeweiligen Disziplin. Zum Beispiel berichtete eine Forscherin, dass Rezensenten, die an qualitative Analysen gewohnt waren, ihre Bedenken, wörtliche Zitate aus nutzergenerierten Inhalten zu verwenden, nicht verstehen konnten. Besonders Informatiker/innen fanden es schwierig Möglichkeiten zur Publikation detaillierter Beschreibungen von Datenerhebungsinfrastrukturen zu finden, da dies oft nicht als originäre Forschung galt. Oftmals boten herkömmliche Publikationsformate generell keine Möglichkeit, Methoden in einem solchen Detailgrad darzustellen, dass eine tatsächliche Nachvollziehbarkeit der zur Anwendung gekommenen Prozesse des Sammelns, Prozessierens und Analysierens der Daten gegeben wäre (Weller/Kinder-Kurlanda 2015). Diese Beobachtung deckt sich mit der Forderung in der Literatur nach mehr Transparenz und Nachvollziehbarkeit für Big-Data-Forschung (z.B. Pasquale 2015). Ein Forscher erklärte es wie folgt: "...if you really wanted to document how you got the data, what you did with the data, how you analyzed, how you processed, visualized the data, if you did all that then you'd already have a paper written before you even get to documenting the outcomes and findings themselves." Meist wurden daher gerade die Teile einer Studie zugunsten der Ergebnisse weggelassen, die deren Validität hätten beweisen können.

Zudem durften auch die Daten aufgrund der Bestimmungen der Plattformbetreiber oft gar nicht geteilt oder zu Reproduktionszwecken archiviert oder anderweitig weitergegeben werden. Forschende befanden sich in Bezug auf das Teilen von Daten in einem „fog of confusion“ (Thomas und Walport 2008), in dem große Unsicherheit darüber herrschte, welche Daten auf welche Weise weitergegeben werden durften. Oft wurden Daten daher eher informell mit befreundeten Forschenden geteilt (Weller/Kinder-Kurlanda 2015).

Schwierigkeiten beim Veröffentlichlichen von Forschungsergebnissen aus Big-Data-Forschung zeigten sich je nach Disziplin auch darin, dass neue, und besonders interdisziplinäre Ansätze, die erforderlich sind, um mit Big Data arbeiten zu können, Forschende automatisch an den Rand etablierter Disziplinen drängten, wo sie Publikationsmöglichkeiten wahrnahmen, die zwar thematisch passend, jedoch nicht so anerkannt waren wie traditionellere Zeitschriften oder Konferenzen. Neben gesonderten Veröffentlichungsmöglichkeiten oder neuen Genres, die die Publikation von ergänzendem Material möglich machen würden, ging es jedoch auch noch um andere Aspekte der Publikationskultur. Weniger Veröffentlichungsdruck, also weniger Artikel und weniger Fokus auf Rankings, wurde auch als Möglichkeit gesehen, Forschenden die Zeit zu geben, sich wirklich auf ein Projekt zu konzentrieren und dieses ausreichend detailliert beschreiben zu können. Diese Probleme, die sich

auf epistemologische Ansätze auswirkten, waren jedoch auf der individuellen Ebene nicht lösbar.

### Individuelle Karriereplanung, Publikationsstrategien und Machbarkeit

In den Interviews und Beobachtungen zeigte sich zudem, wie die individuelle Planung der akademischen Karriere über die gezielte Platzierung von Veröffentlichungen und den Werdegang über Promotion, Lehre und Stellensuche die Forschungsarbeit rahmte. Einzelne Forschungsvorhaben wurden immer auch in den Kontext der konkreten Situation (z.B. als Promovierende in einem Drittmittel-Projekt oder als Inhaberin einer Professur) eingeordnet. Die Situation stellte sich jedoch für Forschende aus den Sozial-, Medien- oder Kommunikationswissenschaften anders dar als für Forschende aus Informatik oder anderen technischen Wissenschaften. Erstere hatten, wie eben beschrieben, zum Beispiel oft Probleme, Ergebnisse zu publizieren: "It's difficult to publish, right. Because you don't belong anywhere (...). You go to the American Political Science Review, they were like, what is this, that is not really political science tradition." Auch die in der Informatik und verwandten Disziplinen verbreitete Art und Weise, Forschung zu betreiben, erforderte eine strategische Ressourcenplanung und Wahl der Forschungsthemen. Prozesse des Peer Review wurden als stark formalisiert angesehen: Wenige hochrangige Konferenzen mit sehr geringen Akzeptanzraten erhielten gerade deshalb sehr hohe Zahlen an Einreichungen. Die Publikation von neuen, interessanten oder gar spektakulären Ergebnissen wurde als besonders erfolversprechend gesehen und manchmal traten Inhalte gar zugunsten von strategischen Überlegungen in den Hintergrund. Individuelle Karrierewünsche, Erfolge, oder Positionen innerhalb bestimmter Institutionen wirkten sich demnach stark auf die Möglichkeiten aus, die einzelne Forschende zur kritischen Entwicklung neuer Epistemologien für Big Data hatten.

### Epistemologielose Praxis?

Aus Beobachtungen der Konferenzen und ihrer Publikationspraktiken, der Beobachtung des Felds und der wichtigsten Publikationsorgane wurde jedoch auch offensichtlich, dass diejenigen, die jenseits erkenntnistheoretischer Überlegungen operierten (d.h. gar nicht auf der Basis der Philosophie arbeiteten und z.B. eher Ansätze verfolgten, in denen lediglich Muster in den Daten gesucht wurden), es einfacher hatten mit den Anforderungen der strategischen Ergebnisorientiertheit umzugehen. Ihre Ergebnisse schienen spektakulärer, so dass sie eher die Möglichkeit hatten das Bild der Big-Data-Forschung zu prägen und Aufmerksamkeit zu erlangen.

Forschende waren beim Durchführen der Analysen zudem oft so sehr mit der Machbarkeit beschäftigt, dass die Forschung fast zwangsläufig weitgehend unabhängig von bestehenden Traditionen und Diskursen um Epistemologie und Erkenntnisgewinn oder existierenden Theorien zum untersuchten Gegenstand (z.B. Kom-

munikation, Geschlechterungleichheit, Wahlverhalten) stattfanden – ein Zustand, der die These vom Ende der Theorie in die Realität zahlreicher problematischer Forschungsprojekte umsetzte. Die Frage, ob und wie Social-Media-Daten die Wirklichkeit überhaupt abbilden und welche Ideologien und Entscheidungen ihrer Entstehung zugrunde liegen, wurde oft nicht gestellt. Beim ‚end of theory‘ ging es also eher darum, dass nicht auf existierende Theorien oder Forschung aufgebaut wurde, wenn mit immer ausgefeilteren computerintensiven Methoden gearbeitet wurde. Der Druck, viel und gezielt zu publizieren widersprach lange ausgedehnten Deliberationsprozessen oder Auseinandersetzungen um Begriffe und Theorien. Es fand somit auch in vielen Big-Data-Studien kaum ein Aufbauen auf existierender Literatur statt. Zum Beispiel würde jedoch eine Beachtung der Rolle von Performanz im Internet oder verschiedener Theoretisierungen der Avatare (Turkle 1997) tragbarere Modelle der heterogenen Nutzer(gruppen) von Social-Media-Plattformen erlauben, die Big Data erzeugen.

## Big-Data-Forschung und Medien- und Techniksoziologie

Auch um ein Wiederaufleben des Grabenkampfes zwischen quantitativen und qualitativen Ansätzen zu vermeiden (Ruppert 2013), müssen sich die Sozialwissenschaftler den Herausforderungen durch die neuen Daten und Methoden stellen. Dies scheint jedoch derzeit oft problematisch: Vertreter/innen der Big-Data-Forschung kommen zwar aus verschiedensten disziplinären Hintergründen, gerade interdisziplinäre Kollaborationen zwischen Wissenschaftler/innen informatischer oder anderer stark technischer Prägung mit theoretischer arbeitenden Sozialwissenschaftler/innen sind selten. Dadurch bleibt das Potential einer theoretischen Fundierung von Big-Data-Analysen, beispielsweise in medien- oder techniksoziologischen Ansätzen derzeit weitestgehend unausgeschöpft. Dies mag daran liegen, dass es für Informatiker/innen nicht unbedingt notwendig ist, sich mit soziologischen Ansätzen zu beschäftigen um etwa im Bereich der Computational Social Science oder der Web Science erfolgreich zu publizieren. Die Lücke zwischen Big Data und Soziologie mag auch, „am theoretischen Zugriff der Soziologie“ liegen, der bewirkt „dass die Technik in ihrem Werken kaum in Erscheinung tritt“ (Rammert 2007: 1) bzw. daran, dass die Soziologie, zumindest in Deutschland, gerade Gefahr läuft, die Digitalisierung und neue Forschungsmethoden zu ‚verschlafen‘ (Diekmann 2016). Angeknüpft werden könnte seitens der Big-Data-Forschung jedoch zum Beispiel an die Konzeptualisierung komplexer Nutzungssituationen jenseits vereinfachter On- und Offline-Dualitäten (z.B. Faßler 1999), an die Modellierung von computervermittelten sozialen Netzwerken (z.B. Stegbauer 2001) mit menschlichen und nicht-menschlichen Akteuren (z.B. Passoth et al. 2012), oder an verschiedene Konzepte zu Interface und Mediatisierung. Auf der anderen Seite bleibt auch das Potential unausgeschöpft, Big-Data-Methoden zu verwenden, um zur soziologischen Theoriebildung beizutragen, Begriffe weiterzuentwickeln und die Veränderungen des

Entstehens und Erhaltens von Sozialität angesichts technisch-netzwerk-vermittelter Medialität zu beobachten.

Durch die Analyse von Social Media Big Data ist nur das Online-Verhalten ganz bestimmter Personen, die eine bestimmte Plattform nutzen, beobachtbar und wird aufgrund der verschiedenen oben genannten Probleme des Zugangs und der Qualität der Daten auch nur zu einem gewissen Grad sichtbar. Die Daten sind beispielsweise vergleichbar mit den Inhalten von geschriebenen Briefen und zusätzlichen Daten über deren Sender und Empfänger (Schroeder 2014), nicht jedoch mit einer Befragung der Nutzenden zu ihren Motivationen und Intentionen. Erst die Verlinkung der Daten zum Online-Verhalten mit Offline-Demographien und ähnlichem ermöglichen ein kompletteres Bild des Nutzenden (Schroeder 2014). Dieser Schritt zum ‚Offline‘ wird jedoch meist nicht gegangen – umgekehrt analysieren diejenigen, die Nutzende selbst befragen, selten deren Online-Inhalte. Ganzheitliche Konzepte, die On- und Offline überbrücken bzw. jenseits dieser Dualität operieren, sind selten. Die wichtige Frage danach, was die Nutzenden sich denn nun ‚tatsächlich‘ gedacht haben, wird von beiden Seiten nicht beantwortet. Ein Zusammenbringen von Big-Data-Ansätzen und medien- bzw. techniksoziologischen Ansätzen könnte dazu beitragen, den derzeitigen Erkenntnisstand über die Internetnutzung zu vertiefen. Welche Auswirkungen haben zum Beispiel Fake-News oder Social-Bot-Netzwerke tatsächlich auf die politische Meinungsbildung? Big-Data-Ansätze können hier in der Tat z.B. zeigen „...that emotions may be transmitted online via words, not just in face-to-face situations“ (Schroeder 2014: 3) in dem sie die Reaktionen auf positive Nachrichten nachzeichnen – inwiefern das Senden positiver Inhalte in Reaktion auf andere positive Inhalte jedoch beeinflusst ist von der Struktur des Interface, vom konkreten Handlungszusammenhang der Nutzenden oder welche Bedeutung dem beigemessen wird, bleibt jedoch unklar. Umgekehrt bleiben bei einer Befragung der Nutzenden zu ihren Motivationen ohne die Betrachtung der tatsächlich geposteten, womöglich von der eigenen Wahrnehmung abweichenden Inhalte, ebenfalls wichtige Aspekte verborgen. Big Data bietet zudem für die Soziologie das Potential mit Daten über globales Verhalten zu arbeiten, verspricht die fast zeitgleiche Beobachtung von Millionen von Menschen weltweit und ermöglicht die Untersuchung globaler Fokussierungen der Aufmerksamkeit. Eine Beschäftigung mit den epistemologischen Problemen, die die Big-Data-Forschung derzeit oft inkompatibel mit soziologischen Ansätzen macht, ist geboten.

## Fazit

Big Data redefiniert und verstärkt einige der zentralen Dilemmata der Geistes- und Sozialwissenschaften, nämlich Fragen nach „relevance (what counts and what doesn't), validity (how meaningful the findings are), generalizability (how far the findings reach), and replicability (the degree to which the results can be reproduced)“ (Ekbia 2016: 10). Aufbauend auf den Vorschlägen für eine Epistemologie von

Big Data aus der Literatur und auf den Ergebnissen aus der konkreten Praxis schlage ich vor, die verschiedenen Arten und Weisen der Produktion von Wissen in Form von Journalartikeln, Konferenzbeiträgen, White Papers, Github-Beiträgen oder Datenkollektionen zu betrachten, um die Big-Data-Forschung weiter zu analysieren. Manche dieser Praktiken sind, abhängig von der disziplinären Verortung, erfolgreicher im Sinne der Erzeugung von Aufmerksamkeit als andere und sind damit besonders erstrebenswert im Sinne persönlicher Karriereplanung einzelner Big-Data-Forschender, deren Einflussnahme und Bekanntheitsgrad in der Forschungsgemeinschaft. Gerade als erfolgreich betrachtete Praktiken können unvereinbar sein mit einer kritischen Reflexion von Theorie und Beobachtungsstandpunkt. Zusätzlich zu der Frage, welches die richtige Epistemologie für Big Data ist, geht es also um den Stellenwert der Philosophie und darum, ob erkenntnistheoretische Überlegungen über Validität und Nachvollziehbarkeit Teil des Forschungsprozesses sein müssen; ob also die Autorität, mit der Aussagen gemacht werden, kritisch zu hinterfragen ist – jenseits eines kurzen Absatzes zu ‚limits of the work‘.

Diese Situation stellt eine Chance für die Sozialwissenschaften dar: Wenn neue oder andere Daten zur Verfügung stehen, können diese auf der Basis vorhandener epistemologischer Instrumentarien methodische Anpassungen der Methodologien vornehmen. Ein Wissen um Epistemologie wird dringend benötigt. Im derzeitigen Alltag der Big-Data-Forscher/innen und in den Publikationsformaten gerade der Informatik ist allerdings derzeit oft kein Raum dafür, herausfinden wo die Grenzen der Aussagefähigkeit der tradierten Methoden liegen und diese Grenzen offen anzusprechen.

Die Praktiken im Alltag der Forschenden und ihre konkreten Probleme im Datenumgang müssen stärker in der Diskussion um die Epistemologie von Big Data eingebracht werden; dazu gehören auch die Rolle der Plattformbetreiber, der akademische Stellenmarkt, Publikationspolitik und die Kosten der Interdisziplinarität, die Big-Data-Forschung einschränken. Es geht also darum, überhaupt die Bedingungen dafür zu schaffen, dass die Auseinandersetzung mit epistemologischen Fragen in der Big-Data-Forschung umfassend stattfinden und eine explizite Auseinandersetzung mit den philosophischen Grundlagen beinhalten kann. Big-Data-Forschung wird oft in Wissenschaften betrieben, die auf eher impliziten, über die Lehre spezifischer Methoden tradierten Epistemologien basieren oder gar keine Tradition der Hinterfragung von Erkenntnis haben. Unbequeme, umständliche und zeitintensive Arten und Weisen der Wissens- und Erkenntnisproduktion, die angesichts des hohen Veröffentlichungsdrucks, wie er in immer mehr Disziplinen herrscht, nur schwer aufrechterhalten werden können, müssen eingefordert werden, um den nötigen Raum für die Reflexion der Epistemologie von Big Data zu schaffen. Diese Forderung betrifft auch die immer stärkere Quantifizierung und Messbarmachung von wissenschaftlichem Erfolg mithilfe verschiedener Publikationsmetriken: Neue Inklusionsmuster dürfen nicht als Demokratisierung der Wissenschaft rhetorisch überhöht

werden (Dickel/Franzen 2015) ohne die Wissenschaft als spezifische Struktur, die digital destabilisiert werden könnte, kritisch in den Blick zu nehmen.

## Literatur

- Anderson, C. (2008): The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June 2008. Online: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory).
- Barocas, S. / Selbst, A.D. (2016): Big Data's Disparate Impact, in: *California Law Review* 104, S. 671-732.
- Bastos, M. / Puschmann, C. / Travitzki, R. (2013): Tweeting across hashtags: Overlapping users and the importance of language, topics, and politics. *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, 1–3, Paris, S. 164-168.
- Berry, D. M. (2011): The Computational Turn: Thinking about the Digital Humanities. *Culture Machine* 12, S. 1-22.
- boyd, d. / Crawford, K. (2012): Critical Questions for Big Data. *Information, Communication & Society* 15(5), S. 662-679.
- Bruns, A. (2013): Faster than the Speed of Print: Reconciling "Big Data" Social Media Analysis and Academic Scholarship. *First Monday*, Oktober.
- Busch, L. (2014): A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets. *International Journal of Communication* 8, S.1727-1744.
- Crawford, K. / Gray, M. / Miltner, K. (2014): Critiquing Big Data: Politics, Ethics, Epistemology. *Special Section Introduction*, in: *International Journal of Communication* 8, S. 1663-1672.
- Dickel, S. / Franzen, M. (2015): Digitale Inklusion: Zur sozialen Öffnung des Wissenschaftssystems, in: *Zeitschrift für Soziologie* 44(5), S. 330-347.
- Diekmann, A. (2016): Die Gesellschaft der Daten. *Süddeutsche Zeitung*, 25. September 2016. Online: [www.sueddeutsche.de/kultur/geisteswissenschaften-die-gesellschaft-der-daten-1.3178096](http://www.sueddeutsche.de/kultur/geisteswissenschaften-die-gesellschaft-der-daten-1.3178096)
- Dourish, P. / & Bell, G. (2011): *Divining a digital future: Mess and mythology in ubiquitous computing*. Cambridge, Mass., London.
- Driscoll, K. / Walker, S. (2014): Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data, in: *International Journal of Communication* 8, S.1745-1764.
- Ekbja, H. / Mattioli, M. / Kouper, I. / Arave, G. / Ghazinejad, A. / Bowman, T. / Suri, V.R. / Tsou, A. / Weingart, S. / Sugimoto, C. R. (2015): Big data, bigger dilemmas: A critical review, in: *Journal of the Association for Information Science and Technology* 66(8), S. 1523-1545.
- Faßler, M. (Hrsg.) (1999): *Alle möglichen Welten: virtuelle Realität, Wahrnehmung, Ethik der Kommunikation*. München.
- Fleisch, E. / Tellkamp, C. (2001): Business perspectives on ubiquitous computing. *M-Lab working paper* (4), S. 83-87.
- Floridi, L. (2012): Big Data and Their Epistemological Challenge, in: *Philosophy & Technology* 25, S. 435-437.
- Frické, M. (2014): Big Data and Its Epistemology, in: *Journal of the Association for Information Science and Technology* 66(4), S. 651-661.

- Gitelman, L. (2015) (ed.): "Raw Data" Is an Oxymoron. Cambridge, Mass., London.
- Helles, R. / Jensen, K. B. (2013): Introduction to the special issue 'Making data – Big data and beyond'. *First Monday*, September 2013.
- Housley, W. / Procter, R. / Edwards, A. / Burnap, P. / Williams, M. / Sloan, L. / Rana, O. / Morgan, J. / Voss, A. / Greenhill, A. (2014): Big and broad social data and the sociological imagination: A collaborative response, in: *Big Data & Society* 1(2), S. 1-15.
- Iliadis, A. / Russo, F. (2016): Critical data studies: An introduction, in: *Big Data & Society* 3(2), S. 1-7.
- Karpf, D. (2012): Social Science Research Methods in Internet Time, in: *Information, Communication & Society* 15(5), S. 639-661.
- Kinder-Kurlanda, K. (2014): Ethnography in a Computer Science Centered Project, in: ICC'14 Workshop, ACM WebSci'14 Conference, Bloomington, Indiana.
- Kinder-Kurlanda, K.E. / Weller, K. (2014): 'I always feel it must be great to be a hacker!': The role of interdisciplinary work in social media research, in: *Proceedings of the 2014 ACM conference on Web Science*, S. 91-98. New York: ACM.
- Kinder-Kurlanda, K.E. / Ehrwein Nihan, C. (2013): Ethically Intelligent? A Framework for Exploring Human Resource Management Challenges of Intelligent Working Environments, in: van Berlo, A. et al (Hrsg.). *Ambient Intelligence – Software and Applications. 4th International Symposium on Ambient Intelligence (ISAmI 2013). Advances in Intelligent and Soft Computing*, Wiesbaden.
- Kitchin, R. (2014): Big Data, new epistemologies and paradigm shifts, in: *Big Data & Society* Juni 2014, S. 1-12.
- Lazer, D. / Kennedy, R. / King, G. / Vespignani, A. (2014): The Parable of Google Flu : Traps in Big Data Analysis, in: *Science* 343, S. 1203-1205.
- Lazer, D. / Brewer, D. / Christakis, N. / Fowler, J. / King, G. (2009): Computational Social Science, in: *Science*, 323(5915), S. 721-723.
- Leonelli, S. (2014): What difference does quantity make? On the epistemology of Big Data in biology, in: *Big Data & Society* 1(1), S. 1-11.
- Lyon, D. (2007): *Surveillance studies: An overview*. Cambridge, Malden, Mass.
- Manovich, L. (2011): Trending: The promises and the challenges of big social data, in: *Debates in the digital humanities*, 2, S. 460-475.
- Manovich, L. (2014): Trending: Verheißungen und Herausforderungen der Big Social Data, in: Reichert, R. (Hrsg.). *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*, Bielefeld, S. 65-83.
- May, T. (2011): *Social research: Issues, methods and process*. Maidenhead, New York.
- Mayer-Schönberger, V. / Cukier, K. (2013): *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Ohm, P. (2010): Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, Vol. 57, p. 1701, 2010; University of Colorado Law Legal Studies Research Paper No. 9-12.
- Pasquale, F. (2015): *The black box society: The secret algorithms that control money and information*. Cambridge, Mass., London.
- Passoth, J.-H. / Peuker, B. / Schillmeier, M. (Hrsg.) (2012): *Agency without Actors? New Approaches to Collective Action*, London.

- Rammert, W. (2007): Technik – Handeln – Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie, Wiesbaden.
- Reichert, R. (2014): Facebooks Big Data. Die Medien- und Wissenstechniken kollektiver Verdauung, in: Reichert, R. (Hrsg.). Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie, Bielefeld, S. 437-452.
- Rieder, G. / Simon, J. (2016): Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data, in: Big Data & Society, 3(1), S. 1-6.
- Ruppert, E. / Law, J. / Savage, M. (2013): Reassembling Social Science Methods: The Challenge of Digital Devices, in: Theory, Culture & Society, 30(4), S. 22-46.
- Ruppert, E. (2013): Rethinking empirical social sciences, in: Dialogues in Human Geography, 3(3), S. 268-273.
- Ruths, D. / Pfeffer, J. (2014): Social media for large studies of behavior, in: Science 346(621), S. 1063-1064.
- Savage, M. / Burrows, R. (2007): The Coming Crisis of Empirical Sociology, in: Sociology, 41(5), S. 885-899.
- Sieber, R.E. / Haklay, M. (2015): The epistemology(s) of volunteered geographic information: a critique, in: Geo: Geography and Environment, 2(2), S. 122-136.
- Schroeder, R. (2014): Big Data and the brave new world of social media research, in: Big Data & Society 1(2), S. 1-11.
- Stegbauer, C. (2001): Von den Online Communities zu den computervermittelten sozialen Netzwerken: eine Reinterpretation klassischer Studien, in: Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung 2(2), S. 151-173.
- Thomas, R. / Walport, M. (2008): Data Sharing Review Report, 11. Juli 2008. Online: <http://amberhawk.typepad.com/files/thomas-walport-datasharingreview2008.pdf>
- Tinati, R. / Philippe, O. / Pope, C. / Carr, L. / Halford, S. (2014): Challenging Social Media Analytics: Web Science Perspectives, S. 177-181.
- Tufekci, Z. (2014): Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In ICWSM'14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media.
- Turkle, S. (1997): Playing in the MUDs. Konstruktionen und Rekonstruktionen des Ich in der virtuellen Realität, in: Schneider, I. / Thomsen, C. (Hrsg.) (1997): Hybridkultur: Medien, Netze, Künste. Köln, S. 324-337.
- Weller, K. / Kinder-Kurlanda, K. (2015): Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research? In: Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop. Proceedings Ninth International AAAI Conference on Web and Social Media Oxford University, May 26, 2015 – May 29, 2015, S. 28-37. Ann Arbor.
- Zelenkauskaite, A. / Bucy, E. P. (2016): A scholarly divide: Social media, Big Data, and unattainable scholarship. First Monday, 21 (5). Online: <http://journals.uic.edu/ojs/index.php/fm/article/view/6358>.

Dr. Katharina Kinder-Kurlanda  
GESIS Leibniz-Institut für Sozialwissenschaften  
Unter Sachsenhausen 6-8  
50667 Köln

Institute for Web Science and Technologies  
Universität Koblenz-Landau  
Universitätsstraße 1  
56070 Koblenz  
katharina.kinder-kurlanda@gesis.org