

Lena Ulbricht, Michael Huch, Ferdinand Gens

Smart und Big Data für den öffentlichen Sektor

Einführung

Daten gelten als *der* Rohstoff des 21. Jahrhunderts. Besonders von „Big Data“ erwarten Unternehmen (und Staaten) sich neue Impulse für wirtschaftliches Wachstum. Big Data charakterisiert sich durch drei zentrale Merkmale: Datenmenge, Datenvielfalt und Datengeschwindigkeit.¹ Durch die fortschreitende Digitalisierung nahezu aller gesellschaftlichen Bereiche nimmt auch die *Menge* der grundsätzlich zur Verfügung stehenden Daten immens zu. So erheben z. B. immer mehr technische Geräte, die mit dem Internet verbunden sind („Internet of Things“), eine Vielzahl an Daten. Diese *Datenvielfalt* lässt sich grob in strukturierte und unstrukturierte Daten kategorisieren, dabei zählen Text, Bild oder Videoaufnahmen (wie sie z. B. in sozialen Netzwerken zu finden sind) zu den unstrukturierten und Zahlen im Datensatzformat zu den strukturierten Daten (z. B. in gängigen Datenbanken). *Datengeschwindigkeit* bezieht sich schließlich auf die Generierung, Übertragung und Verarbeitung großer Datenmengen. So lassen sich Daten bereits „in Echtzeit“ erheben und in entsprechende Systeme zur Verarbeitung einspielen.

Würden entsprechende Daten lange Zeit nur gespeichert, können sie dank des Cloud Computing nun auch ausgewertet werden. Beim Cloud Computing stellen IT-Dienstleister ihren Kunden Speicherplatz und Softwarelösungen zum Verwalten und Bearbeiten von Daten in einer externen IT-Umgebung, der Cloud, an. Diese neuen datenanalytischen Werkzeuge ermöglichen es einer wachsenden Anzahl von Nutzern, schier unvorstellbar große Datenmengen in Echtzeit zu analysieren.

Die Nutzung von Big Data wurde bislang zuvorderst aus der Sicht der Wirtschaft diskutiert. So erstellen beispielsweise Handelsketten aus ihren Datenbeständen Prognosemodelle zur

Optimierung ihrer Geschäftsabläufe, etwa zur Absenkung des Lagerbedarfs², oder entwickeln neue Marketingstrategien auf Grundlage von Kundenfeedbacks und Informationen aus sozialen Netzwerken. Aber was ist eigentlich der Stand der Anwendung von Big Data in Politik und Verwaltung? Seit vielen Jahren ist ein allgemeiner Trend hin zu evidenzinformaten oder gar -basierten politischen Entscheidungen zu beobachten. Eine wissenschaftliche Fundierung von politischen Programmen wird zunehmend als Qualitätsmerkmal guter Politik angesehen; Forschungsprojekte, Auftragsstudien, Gutachten und Expertengespräche sollen Politikentscheidungen vorbereiten und begründen.³

In diesem Zusammenhang könnten neue Datenverarbeitungstechnologien die Evidenzgrundlage politischer Entscheidungsträger deutlich erweitern und ergänzen. Entsprechend stellt sich die Frage, welche neuen Einblicke Big Data politischen Entscheidungsträgern für Agenda-Setting ermöglicht, z. B. in Bezug auf die Definition von Problemen und die (wahrgenommene) Notwendigkeit staatlicher Intervention. Einen weiteren Mehrwert verspricht Big Data mit Blick auf die Evaluation staatlicher Interventionen, die ein wesentlicher Aspekt evidenzbasierter Politik sind: Sie erfüllen eine staatliche Rechenschaftspflicht, ermöglichen politische Kontrolle und sind häufig Grundlage für die Entscheidung, bestimmte Politikmaßnahmen weiterzuführen, anzupassen oder zu beenden.

Die Potenziale von Big Data für den öffentlichen Sektor sind Thema dieses Papiers. Dabei steht nicht allein die schiere Größe von Datensätzen im Vordergrund, sondern vielmehr eine bessere Nutzung bestehender (großer wie kleiner) Datenvolumina, die als „Smart Data“ bekannt ist. Zunächst wirft diese Publikation einen Blick auf relevante technische Grundlagen und ak-

1 Im Englischen werden in der Regel die Begriffe *Volume*, *Variety* und *Velocity* verwendet.

2 Vgl. Jüngling, Thomas (2013): Wie die Sammler von Big Data uns durchleuchten. In: Die Welt, 4.3.2013. Online unter: www.welt.de/wirtschaft/webwelt/article114121023/Wie-die-Sammler-von-Big-Data-uns-durchleuchten.html [22.10.2014].

3 Vgl. hierzu auch Kind, Sonja und Wessels, Jan (2014): Evidenzbasierte Innovationspolitik – Evaluation im Wandel. iit-perspektive Nr. 18. Online unter: www.iit-berlin.de/de/publikationen/evidenzbasierte-innovationspolitik-evaluation-im-wandel [28.01.2015].

tuelle Forschungsvorhaben, die diese weiterentwickeln sollen. In dem zweiten Kapitel werden ausgewählte Anwendungsfälle von Big und Smart Data vorgestellt, die für den öffentlichen Sektor von Nutzen sein können. Diese finden sich mehrheitlich, aber nicht ausschließlich, im Ausland. Besondere Aufmerksamkeit wird dabei dem Ansatz der „Business Intelligence“ gewidmet, der Berücksichtigung von Feedback und Echtzeitanalysen in laufenden Prozessen. Das dritte Kapitel zeigt sodann gelungene Beispiele der Visualisierung großer Datenbestände auf, die diese für politische Entscheider erst praktisch verwertbar machen. Schließlich widmet sich das vierte Kapitel wichtigen nicht-technologischen Bedingungen für die Nutzung von Big und Smart Data: und zwar rechtlichen Rahmenbedingungen, Datenschutzfragen und ethischen Implikationen. Das Arbeitspapier schließt mit einem Set an Handlungsempfehlungen für die Politik. Nicht adressiert wird dagegen die Debatte um öffentlich bereitgestellte Daten im Kontext von „Open Data“ und „Open Government“.

1. Basistechnologien und Forschungsvorhaben

Als wichtigste technologische Voraussetzungen für das Aufkommen von Big Data-basierten Anwendungen gelten die im Zuge der Einführung von Cloud Technologien drastisch gesunkenen Speicherkosten sowie neue parallele Datenverarbeitungsmöglichkeiten und neue Datenbankarchitekturen. Bei den Datenverarbeitungsmöglichkeiten erlaubte zunächst das von Google eingeführte Programmiermodell „MapReduce“ parallele und auf mehrere Rechner verteilte Berechnungen.⁴ Unter dem Namen „Hadoop“ wurde das Programmiermodell zu einem Dateisystem zur Speicherung sehr großer Datenmengen auf mehreren Rechnern weiterentwickelt und fand über die Apache Software Stiftung als Open Source große Verbreitung.⁵

Zahlreiche Forschungsprojekte beschäftigen sich mit der Anwendung und Weiterentwicklung der entsprechenden Möglichkeiten: Die Europäische Technologieplattform NESSI (Networked European Software and Services Initiative) berichtet in ihrem Ende 2012 erschienenen White Paper⁶ zum Thema Big

Data von über 80 einschlägigen Forschungsprojekten im Portfolio des 7. EU-Forschungsrahmenprogramms. Auch in Deutschland arbeiten Forscher an Big Data-Analyseplattformen, z. B. im Rahmen des „Stratosphere“-Forschungsprojekts⁷ der TU Berlin und der HU Berlin sowie dem Hasso-Plattner-Institut, das durch mehrere nationale und europäische Förderquellen sowie durch Finanzierungsbeiträge großer IT-Unternehmen gefördert wird. Das Bundesministerium für Bildung und Forschung (BMBF) wird zudem in den kommenden Jahren zwei Kompetenzzentren für Big Data aufbauen: Erstens das Berlin Big Data Center (BBDC) an der TU Berlin, das automatisch skalierbare Technologien entwickeln will, um riesige heterogene Datenmengen zu organisieren und gleichzeitig aus all diesen Datenmengen „intelligente“ Informationen zu gewinnen. Und zweitens das Competence Center for Scalable Data Services (ScaDS) an der TU Dresden, das einen eher service-orientierten Ansatz verfolgt, um ein Portfolio von Big Data-Lösungen für Wissenschaft und Industrie zu entwickeln.⁸

Deutlich anwendungsorientierter und finanziert vom Bundesministerium für Wirtschaft und Energie (BMWi) arbeitet ein Konsortium unter Leitung der TU Berlin an einem Marktplatz für Cloud-basierte Informationen und Analysen (MIA – a cloud-based Marketplace for Information and Analytics)⁹, das explizit den deutschen Mittelstand als potenziellen Anwender adressiert. Erwähnenswert ist zudem das noch junge BMWi-Förderprogramm „Smart Data – Innovationen aus Daten“, in dessen Rahmen im Sommer 2014 13 Forschungsvorhaben zur Förderung ausgewählt wurden¹⁰, und das ebenfalls vor allem den deutschen Mittelstand zu Innovationen motivieren möchte. In diesen beispielhaft erwähnten Vorhaben und den zahlreichen Äußerungen zum technologischen Forschungsbedarf – etwa in der Forschungsagenda der „Strategic Research and Innovation Agenda on Big Data Value for Europe“¹¹ – wird allerdings kaum auf politische Entscheidungsträger als Anwender eingegangen.

Dies hat die EU-Kommission erkannt und im Herbst 2014 eine Studie ausgeschrieben, die EU-weit sowie in wichtigen Wettbewerbsländern aktuelle Anwendungsfälle beschreiben soll und zusätzlich einen konkreten Demonstrator für einen frei

4 Vgl. die Definition des Begriffs „MapReduce“ auf Wikipedia. Online unter: <http://de.wikipedia.org/wiki/MapReduce> [22.10.2014].

5 Vgl. Bayer, Martin (2013): Hadoop – der kleine Elefant für die großen Daten. In: Computerwoche, 25.03.2013. Online unter: www.computerwoche.de/a/hadoop-der-kleine-elefant-fuer-die-grossen-daten,2507037 [28.01.2015].

6 NESSI White Paper (2012): Big Data – A New World of Opportunities. Online unter: http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf [28.01.2015].

7 Siehe hierzu den Online-Auftritt des Forschungsprojekts „Stratosphere“. Online unter: <http://stratosphere.eu> [28.01.2015].

8 Bundesministerium für Bildung und Forschung (2014): Big Data – Management und Analyse großer Datenmengen. Online unter: www.bmbf.de/de/23429.php [05.11.2014].

9 MIA – a cloud-based Marketplace for Information and Analytics. Online unter: https://www.dima.tu-berlin.de/menue/research/completed_projects/mia [05.11.2015].

10 Bundesministerium für Wirtschaft und Energie (2014): „Smart Data“ – Projekte des neuen BMWi-Technologieprogramms sind ausgewählt, Pressemitteilung des BMWi vom 19.06.2014. Online unter: <http://www.bmwi.de/DE/Presse/pressemitteilungen,did=642872.html> [05.11.2014].

11 Big Data Value Europe (2015): European Big Data Value Strategic Research & Innovation Agenda, version 1.0, Januar 2015, Brüssel. Online unter: http://www.bigdatavalue.eu/images/SRIA/EuropeanBigDataValuePartnership_SRIA__v1%200_final.pdf [28.01.2015].

wählbaren Anwendungsfall umfasst, der die Präsentation der daten-analytisch gewonnenen Ergebnisse in Form einer Instrumententafel („Dashboard“) verlangt.¹² Es ist erklärte Absicht, diesen Demonstrator zu nutzen, um politische Entscheidungsträger für das Potenzial der heute möglichen Datenanalysen zu sensibilisieren. Eine weitere Ausnahme stellt ein durch die EU gefördertes Projekt über die Nutzung von Informations- und Kommunikationstechnik für Politikgestaltung dar: Dieses hat eine – explizit aus Sicht der Anwender formulierte – Forschungsroadmap unter dem Namen „Towards Policy-Making 2.0“ entwickelt, die auch auf die bislang wenig genutzten Potenziale von Big Data hinweist. Für eine bessere Nutzung von Big Data durch die Politik müssen der Roadmap zufolge zum einen technologische Herausforderungen gemeistert und zum anderen ein sicherer Umgang hinsichtlich der Glaubwürdigkeit und Interpretation der Daten entwickelt werden (zu diesen Themen siehe auch Kapitel 4).¹³

2. Smart und Big Data für den öffentlichen Sektor

Big Data-Analysen ermöglichen vielseitige Einblicke. Besondere öffentliche Aufmerksamkeit haben bislang Anwendungen im Bereich der öffentlichen Sicherheit erlangt, z.B. das „prospective policing“, das auch in Deutschland bereits erprobt wird.¹⁴ Weniger bekannt sind die in den folgenden zwei Kapiteln dargelegten Anwendungen in den Bereichen Daseinsvorsorge, Bildungspolitik, Sozialpolitik und Innovationspolitik.

Noch stammen die meisten Anwendungsbeispiele für die Nutzung von Big Data im öffentlichen Sektor aus den USA, doch auch in Deutschland häufen sich die Vorschläge. So prognostiziert die Arbeitsgemeinschaft der Kommunalen IT-Dienstleister in Deutschland: „Ganz sicher wird Big Data in Verwaltungen Einzug halten, da dort die Datenmengen immer weiter ansteigen und die Notwendigkeit besteht, diese Daten intelligent zu analysieren.“¹⁵ Doch während praktische Erfahrungen mit Big

Data hier noch selten sind, findet eine Form von Smart Data, die sogenannte Business Intelligence, vermehrt Anwendung in der öffentlichen Verwaltung. In einer ersten Stufe werden dabei vorhandene, oder im Verlauf einer politischen Maßnahme generierte (meist strukturierte) Daten analysiert – dies wird bereits vielfach praktiziert, allerdings erfolgt die entsprechende Auswertung bislang kaum „in Echtzeit“. In einer zweiten Stufe, so ein Vorschlag eines kommunalen IT-Dienstleisters¹⁶, sollten zusätzlich unstrukturierte Daten aus verschiedenen Internetquellen hinzugezogen werden; eine dritte Stufe der Datennutzung wären dann prognostische Analysen.

Auf den Schulsektor bezogen würden z.B. für Stufe eins offizielle Schuldaten wie Abschlüsse, Anzahl der Abbrecher oder Lehrerqualifikationen analysiert; in Stufe zwei öffentliche Daten aus dem Internet hinzukommen, etwa auch Aussagen aus sozialen Plattformen – und in der dritten Stufe die prognostische Entwicklung eines Stadtteils einbezogen.¹⁷

Wie bereits angedeutet, basiert die erste Stufe zumeist auf Daten, die regelmäßig von statistischen Ämtern oder zeitgleich zur Umsetzung einer Maßnahme generiert werden. Mit einem entsprechenden Geschäftsmodell ist z. B. das US-amerikanische Unternehmen Socrata unter dem Namen „GovStat“ tätig – auch in Europa. Das zugrundeliegende Konzept „denkt“ die parallel zur Umsetzung einer politischen Maßnahme stattfindende Datensammlung und -auswertung mit. In einer dynamischen Darstellung, etwa über Instrumententafeln („Dashboards“), kann der aktuelle Umsetzungsstand einer politischen Maßnahme mit Blick auf vorab formulierte Ziele visualisiert werden.¹⁸ Auch das deutsche Unternehmen Agendo GmbH richtet ein vergleichbares Angebot an Regierungsinstitutionen, um diese bei der politischen und strategischen Planung zu unterstützen¹⁹; hier ist die Instrumententafel ebenfalls das Herzstück der Auswertung. Allerdings sind beide Anbieter bisher nicht explizit in der Datenanalyse unstrukturierter, externer Internet-Quellen tätig. Socrata positioniert sich aber bereits als Anbieter Cloud-basierter Analysetools.

12 Europäische Kommission (2014): Data Technologies for evidence-informed policy-making (including Big Data), SMART 2014/0004, Brüssel. Online unter: <http://ec.europa.eu/digital-agenda/en/news/data-technologies-evidence-informed-policy-making-including-big-data-smart-20140004> [28.01.2015].

13 Vgl. Osimo, David / Mureddu, Francesco / Onori, Riccardo / Armenia, Stefano / Misuraca, Gianluca Carlo (2013): Towards Policy-making 2.0: The International Research Roadmap on ICT for Governance and Policy Modelling, S. 64ff. Online unter: http://crossover-project.eu/Portals/0/0205F01_International%20Research%20Roadmap.pdf [05.11.2014].

14 Simon, Cathérine (2014): „Precob“: Polizei will Einbrüche per Software vorhersagen. In: Spiegel Online, 08.11.2014. Online unter: <http://www.spiegel.de/netzwelt/netzpolitik/einbruch-polizei-testet-precob-software-zur-vorhersage-a-1001816.html> [05.11.2014].

15 Sauter, Birgit (2013): Weiterdenken erlaubt. In: Vitako Aktuell, Zeitschrift der Bundes-Arbeitsgemeinschaft der Kommunalen IT-Dienstleister e.V., Ausgabe 1/2013, S. 9.

16 Vgl. Reichel, Andreas (2013): Gut vorbereitet. Big-Data-Management: wer die explodierenden Datenmengen nutzen will, sollte sich rüsten. In: Vitako Aktuell, Zeitschrift der Bundes-Arbeitsgemeinschaft der Kommunalen IT-Dienstleister e.V., Ausgabe 1/2013, S. 6–8.

17 Vgl. Ebd. S. 8.

18 Vgl. Blauer, Beth (2013): How to Run a GovStat Programm, Step-by-step guidance for building a successful government performance system based on data, S. 4. Online unter: <http://www.socrata.com/wp-content/uploads/2013/07/GovStat-how-to-guide.pdf> [02.11.2014].

19 Vgl. Agendo – Gesellschaft für politische Planung mbH. Online unter: www.agendo.de/de/content/intraplan-ng-die-software-f%C3%BCr-strategische-planung [05.11.2014].

Einen starken Fokus auf Big Data-Analysen bietet wiederum das US-Unternehmen „Palantir Technologies“, das verschiedene Softwareplattformen, in denen unterschiedliche Daten in großer Menge und Echtzeit zusammengetragen werden, miteinander in Verbindung stellt und statistisch auswertet. Palantir war zunächst im Bereich sicherheits- und finanzpolitischer Maßnahmen für Unternehmen und Regierungen tätig. Das Unternehmen hat aber auch eine schnelle Koordinierung von Katastrophenhilfe und Daseinsvorsorge²⁰ ermöglicht, indem komplexe und umfangreiche Daten für Entscheidungsträger aufbereitet wurden. So unterstützte Palantir im Oktober 2012 als eigenen Beitrag die Verteilung medizinischer Produkte in den von Hurrikan Sandy verwüsteten Gebieten. Zu diesem Zweck wurden Daten über die Vorräte von Apotheken und Krankenhäusern, den Zustand der Straßen, das Wetter, Überflutungen und Kraftstoffversorgung sowie Anfragen betroffener Bürger bei der New Yorker Stadtverwaltung zusammengeführt. Auf diese Weise konnte die Katastrophenhilfe zielgerichtet an die am stärksten bedürftigen Orte geliefert werden. Darüber hinaus wurde von den Centers for Disease Control and Prevention und Palantir Technologies ein Dashboard entwickelt, mit dem die Verbreitung ansteckender Krankheiten beobachtet werden konnte.²¹

Auch in Hinblick auf den Bildungsbereich gibt es zahlreiche Vorschläge für die Nutzung von Echtzeitanalysen sowie Big Data-Analysen.²² Sie sehen u. a. vor, dass die Daten, die durch statistische Ämter, Behörden, Schulen und Lehrer erhoben werden, nicht nur für jährliche Bildungsberichte, sondern auch als tägliches Arbeitsinstrument in Echtzeitanalysen genutzt werden. Dieses Ziel verfolgt z. B. das Strategic Data Project der Harvard University, das Schulen und Schulverwaltungen dazu befähigen soll, verfügbare Daten anhand fortgeschrittener Analysemethoden auszuwerten.²³ Ein europäisches Beispiel ist die digitale Plattform „E-School“ des estnischen Staates. Diese Plattform nutzen Lehrer, um Hausaufgaben zu vergeben und die Leistungen sowie das Verhalten der Schüler zu bewerten. Eltern können sich tagesaktuell über Hausaufgaben, Teilnahme und Leistungen ihrer Kinder informieren; die gleichen Informationen stehen auch den Schülern zur Verfügung. Nicht zuletzt bietet die Plattform der Schulverwaltung jederzeit maßgeschneiderte statistische Auswertungen, auch im interregionalen Vergleich.

Nach eigener Aussage des Portals nutzen 85 Prozent der Schulen diese Plattform, sodass Informationen über 95 Prozent der estnischen Schülerinnen und Schüler erfasst werden.²⁴

Darüber hinaus erheben digitale (und häufig interaktive) Lernprogramme neue Daten, die auf den Lernprozess von Schülerinnen und Schülern schließen lassen („Learning Analytics“). Ein digitales Lernsystem kann Informationen darüber erfassen, wie Schüler Aufgaben lösen, welche Strategien sie sich bedienen und wie erfolgreich sie dabei sind.²⁵ Diese Informationen können auf verschiedenen Ebenen genutzt werden: um Schülerinnen und Schüler angemessener zu fördern, um Probleme unmittelbar zu erkennen und ggf. Gegenmaßnahmen zu treffen (und so Abbrüche/Abgänge zu vermeiden). Entsprechende Analysen werden z. B. von Hochschulen dafür verwendet, um die Zahl der Studienabbrüche reduzieren zu können.²⁶

Smart und Big Data bergen somit (in Kombination, aber auch getrennt voneinander) ein großes Potenzial für verschiedene bildungspolitische Akteure: Sowohl Lehrer als auch Dozenten könnten z. B. einen detailgenauen und aktuellen Überblick über die Leistungsentwicklung eines Schülers erhalten und diese Informationen ins Verhältnis zu Referenzgruppen (Klasse, Jahrgang, Region, nationaler Durchschnitt etc.) setzen. Sie wären zudem in der Lage, Reaktionen auf erprobte Lehrinhalte und Methoden zu identifizieren sowie darauf aufbauend die Förderung einzelner Schüler anzupassen und ihr Lehrangebot zu modifizieren. Bildungseinrichtungen und Schulverwaltungen könnten die Informationen nutzen, um ihre Rekrutierungsstrategien anzupassen, ihr Lehrangebot zu optimieren und Risikostudenten eine besondere Unterstützung zukommen zu lassen. Viele weitere Anwendungsmöglichkeiten von Big bzw. Smart Data sind denkbar und könnten die Informationsgrundlage für Entscheidungsträger in Politik und Verwaltung verbessern. Hier sind strategische und langfristig orientierte Fragestellungen ebenso möglich wie Detailbetrachtungen z. B. für die konkrete Umsetzung politischer Maßnahmen auf der Ebene nachgeordneter Behörden. Dabei sind die entsprechenden Erkenntnisse als Entscheidungsgrundlage dann besonders nützlich, wenn sie sinnvoll visualisiert werden.

20 Eine entsprechende Anwendung trägt z. B. zur Optimierung des Wassermanagements in Kalifornien bei.

21 Vgl. Palantir (2014): Supporting Relief Efforts: Palantir & Direct Relief International. Online unter: <https://www.palantir.com/wp-assets/wp-content/uploads/2014/03/Impact-Study-Supporting-Relief-Efforts-DRI.pdf> [05.11.2014].

22 Für eine Vielzahl an Vorschlägen und Beispielen vgl. U.S. Department of Education (2012): Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics. An Issue Brief, Oktober 2012. Online unter: <http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf> [05.11.2014].

23 Vgl. Harvard University, Strategic Data Projekt. Online unter: <http://cepr.harvard.edu/sdp/about/index.php> [05.11.2014].

24 Vgl. e-estonia.com – The Digital Society. Online unter: <http://e-estonia.com/component/e-school> [05.11.2014].

25 Ein Beispiel für ein kostenloses „intelligentes“ Programm zur Lernunterstützung ist „ASSISTment“ das durch das Worcester Polytechnic Institute sowie OLI an der Carnegie Mellon University entwickelt wurde.

26 Diverse Beispiele auf EDUCAUSE. Online unter: <http://www.educause.edu/library/learning-analytics> [05.11.2014].

3. Visualisierung von Daten

Zur visuellen Aufbereitung der Analyseergebnisse von Big und Smart Data zählen grafische Modelle, Tabellen, Diagramme und Bilder. Idealerweise erlaubt die Visualisierung zudem eine Interaktion und Betrachtung der Daten aus verschiedenen Blickwinkeln durch den Anwender („Dashboard“). Auf diese Weise können zuvor unbekannte Zusammenhänge und Probleme für politische Entscheidungsträger sichtbar und verständlich aufbereitet werden.

Ein simples Visualisierungsbeispiel ist das „Billion Prices Project“ des MIT²⁷, das einen tagesaktuellen Preisindex berechnet und von der Ausgründung PriceStats²⁸ heute weltweit vermarktet wird. Das Unternehmen setzt Web-Technologien ein, um überwiegend unstrukturierte Daten, zumeist im HTML-Format, in strukturierte Datenbanken zu überführen. Dabei nutzt das Unternehmen Preisinformationen von Online-Händlern, wobei Qualitätskriterien für die Auswahl der Händler herangezogen werden, z. B. deren Marktanteile. Tagesaktualität und Kosteneinsparungen werden dadurch möglich, dass keine Preiserhe-

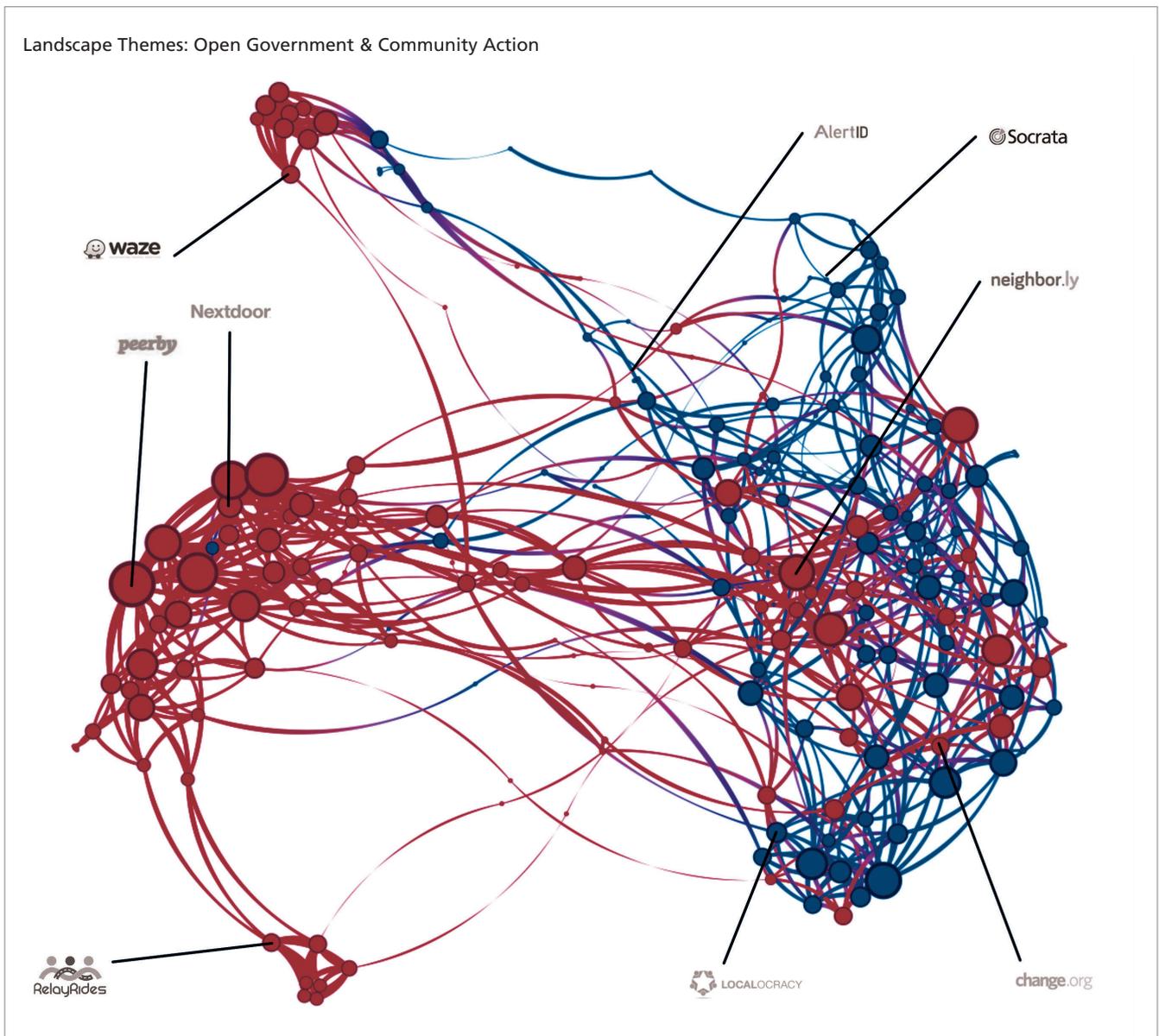


Abbildung 1: Network Map der Knight Foundation. Die Kreisgröße repräsentiert die Anzahl der Organisationen in einem Cluster. Die Stärke der Verbindungslinien steht für die Anzahl der Verbindungen der Organisationen innerhalb eines Clusters. (Quelle: Knight Foundation/farbllich angepasst)²⁹

27 Billion Prices Project @ MIT. Online unter: <http://bpp.mit.edu/usa> [05.11.2014].

28 Vgl. PriceStats. Online unter: <http://www.pricestats.com> [05.11.2014].

29 Patel, Mayur / Sotsky, Jon / Gourley, Sean / Houghton, Daniel (2013): The Emergence of Civic Tech: Investments in a Growing Field, Miami, S. 10. Online unter: www.knightfoundation.org/media/uploads/publication_pdfs/knight-civic-tech.pdf [05.11.2014].

bungen mehr vor Ort durchgeführt werden – das sind zwei besondere Vorteile dieser Methode. Berühmt wurde die Anwendung durch den „Economist“, der sich weigerte, die seiner Meinung nach geschönte offizielle Teuerungsrate Argentiniens weiter zu veröffentlichen und stattdessen – auch heute noch³⁰ – auf die Daten von PriceStats zurückgriff.

Beispielhaft für die visuelle Nutzbarmachung des Echtzeit-Monitoring für Entscheidungsträger auf höchster politischer Ebene ist das für das britische Kabinett und Premierminister David Cameron entwickelte „Number 10 Dashboard“. Das Dashboard, zu dem keine detaillierten Informationen zugänglich sind, soll über eine iPad-App abgerufen werden können und einen tagesaktuellen Überblick zu politischen Trends und Stimmungen sowie „harten Fakten“ in Bezug auf Arbeitsmarkt- und Wirtschaftsstatistiken geben. Die Quellen sollen auf hunderten Datensätzen öffentlicher, aber auch privater Unternehmen basieren – etwa wie „Adzuna“³¹, einem in London ansässigen Start-up für Stellenausschreibungen – oder sozialen Netzwerken wie Google, Twitter und Facebook. Ein Ampelsystem soll darüber hinaus den Status bestimmter Regierungsinitiativen anzeigen und so auf möglichen Handlungsbedarf hindeuten.

Ein weiteres Beispiel gelungener Visualisierung entstammt der Innovationspolitik: Das in San-Francisco gegründete Big Data-Start-up „Quid“ hilft Investoren und Regierungsorganisationen bei der Analyse von Technologieentwicklungen, Innovationen sowie Forschung und Entwicklung. Hierzu werden mathematische Muster anhand von Patentanmeldungen, wissenschaftlichen Artikeln oder Kapitalzuwendungen analysiert und visualisiert. Außerdem erlaubt das Tool, zwischen etablierten Sektoren und „technology whitespaces“, die Entwicklungspotenziale abzubilden, zu unterscheiden, und Möglichkeiten der Vermarktung neuer Produkte zu identifizieren. Hier bietet sich öffentlichen Akteuren die Chance, Forschungsmittel (noch) effektiver einzusetzen. Die Knight Foundation hat z. B. ein solches von „Quid“ entwickeltes Analyseinstrument genutzt, um die Entwicklung von Innovationen im Bereich ziviler Technologien nachzuvollziehen. Hierfür wurden Medien-, Presse- und Investmentdaten zusammengeführt; auf der Grundlage von Ähnlichkeiten in der Beschreibung der jeweiligen Technologien konnte dann eine Netzwerkkarte erstellt.³² Auf diese Weise

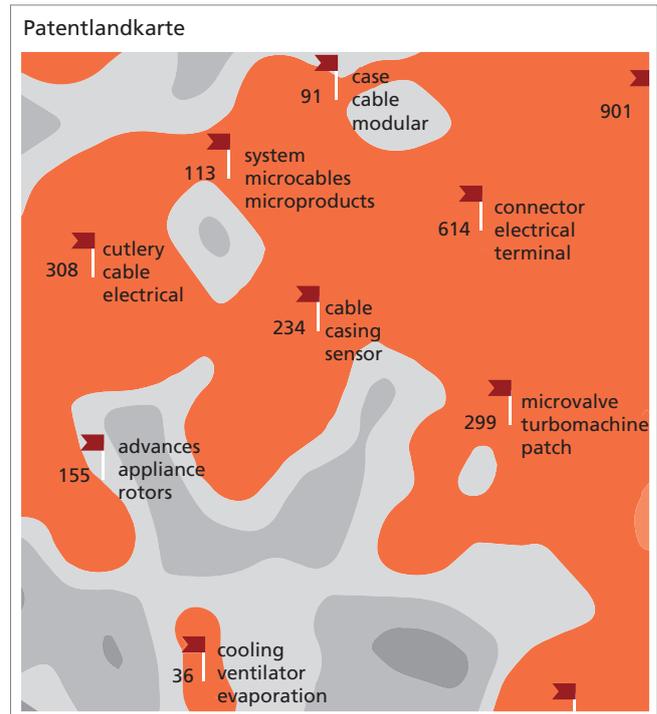


Abbildung 2: Ausschnitt aus der Patentlandkarte Berlin. (Quelle: Technologiestiftung Berlin/farbllich angepasst)³³

konnten die Cluster „Open Government“ und „Community Action“ identifiziert werden (siehe Abbildung 1, S. 5).

Eine weitere Form der Visualisierung bietet das Berliner Unternehmen mapegy³⁴, das im Auftrag der Technologiestiftung Berlin eine interaktive Patentlandkarte Berlins entwickelt hat. Das Analyseinstrument wertet relevante Patentaktivitäten³⁵ in ausgewählten Zeiträumen aus (für die Berliner Patentlandkarte z. B. von 1993 bis 2012); als Datengrundlage dient die Patentdatenbank PATSTAT, die weltweite Patentanmeldungen erfasst, ergänzt durch Daten zur Beurteilung der Marktdeckung und der Technologierelevanz. Die Patentlandkarte gibt Gründern, Förderern und Investoren auf einer skalierbaren Karte entsprechend Auskunft über die Berliner Technologielandschaft und ihre Dynamiken. Die Technologiestiftung verspricht: „Mit dem Patentradar lässt sich potenziell auch ermitteln, woran Berliner Unternehmen heute forschen und entwickeln und womit sie morgen voraussichtlich Umsatz machen werden.“³⁶

30 Vgl. The Economist. Online unter: <http://www.economist.com/node/21604509> [05.11.2014].

31 Vgl. adzuna. Online unter: <http://www.adzuna.co.uk/> [05.11.2014].

32 Vgl. Patel, Mayur / Sotsky, Jon / Gourley, Sean / Houghton, Daniel (2013): The Emergence of Civic Tech: Investments in a Growing Field, Miami. Online unter: www.knightfoundation.org/media/uploads/publication_pdfs/knight-civic-tech.pdf [05.11.2014].

33 Siehe hierzu das Online-Tool der Technologiestiftung Berlin „Technologieradar Berlin“. Online unter: www.technologiestiftung-berlin.de/de/top-themen/innovationsstandort-berlin/technologieradar [28.01.2015].

34 Vgl. mapegy. Online unter: <http://www.mapegy.com/> [20.01.2015].

35 Neben Patentanmeldungen werden auch die Gewährung, Übertragung und das Erlöschen von Patenten berücksichtigt.

36 Technologiestiftung Berlin (2014): Technologieradar Berlin. Report 2014, Berlin, S. 10. Online unter: http://www.technologiestiftung-berlin.de/fileadmin/daten/media/publikationen/140730_Studie_Technologieradar_2014.pdf [20.01.2015].

Wie in den letzten beiden Kapiteln illustriert, werden beständig neue Smart und Big Data-Anwendungen entwickelt. Dass diese in Deutschland häufig (noch) nicht umgesetzt werden, liegt nicht allein an technologischen Herausforderungen oder an den hohen Kosten. Auch rechtliche Vorgaben und ethische Erwägungen prägen den bislang eher moderaten Erfolg von Big Data-Anwendungen in Deutschland.

4. Rahmenbedingungen für Data Analytics

Viele der kursierenden Ideen für Big Data-Lösungen müssten auf Anwendungen zurückgreifen, in deren Erhebung, Speicherung und Verwendung die Nutzer unwissend oder gar nicht eingewilligt haben. Das ist vor allem ein datenschutzrechtliches Problem. Denn häufig ist der Prozess der Datenerhebung für Nutzer schlichtweg nicht erkennbar, z.B. bei der Eingabe von Suchanfragen in Suchmaschinen. Das Bundesdatenschutzgesetz gibt allerdings vor, dass personenbezogene Daten nur erhoben werden dürfen, wenn die Betroffenen eingewilligt haben (oder wenn eine Norm dies ausdrücklich gestattet). Diese Vorgabe gilt zwar nicht für technische Daten, die beispielsweise über Maschinen oder Geräte erfasst werden. Allerdings treffen technische Daten auch Aussagen über ihre Nutzer, sodass zuweilen der Vorschlag gemacht wird, dass Daten von besonderer „Sensibilität“ definiert und gesondert geschützt werden sollten. Selbst allgemein zugängliche Daten sind nicht uneingeschränkt für Big Data-Anwendungen nutzbar.

Eine Möglichkeit, große Datenmengen dennoch für Big Data-Anwendungen zu nutzen, ist die Entkopplung von der Person (*de-identification*); allerdings greifen die entsprechenden Verfahren z.T. zu kurz und können von den sich rasch entwickelnden Datenerhebungs- und Auswertungsmethoden überholt werden.³⁷ Denn durch eine steigende Datenmenge steigt gleichzeitig das Risiko, dass ein Personenbezug wiederhergestellt werden kann (*re-identification*).³⁸ Erschwerend kommt hinzu, dass viele der (meist privaten) Anbieter von Big Data-Datensätzen und -Anwendungen, häufig um Marktvorteile zu

bewahren, wenig transparent hinsichtlich der Erhebung und Verarbeitung der genutzten Daten sind. Und schließlich verdeutlichen Skandale, dass Big Data leicht Persönlichkeitsrechte verletzen kann: In einem Fall haben Forscher mithilfe öffentlich verfügbarer Daten, u. a. aus sozialen Netzwerken³⁹, die Sozialversicherungsnummern von Internetnutzern identifiziert; in einem anderen Fall wurde eine Nutzerin der Videoplattform Netflix durch anonymisierte Nutzerdaten, die Netflix an Forscher weitergereicht hatte, unfreiwillig als gleichgeschlechtlich geoutet.⁴⁰ Entsprechende Erfahrungen tragen zum Vertrauensverlust in der Bevölkerung bei, die zunehmend Wert auf ihre informationelle Selbstbestimmung legt.⁴¹

Darüber hinaus haben kritische Stimmen ethische Probleme aufgezeigt, die sich durch Big Data-Szenarien ergeben können. So können Personen illegitime Nachteile erfahren, wenn sie z. B. bei der Wohnungs- und Kreditvergabe benachteiligt oder in Bewerbungsverfahren sowie Polizeikontrollen diskriminiert werden.⁴² Entsprechende negative Konsequenzen sind besonders problematisch, da die betroffenen Individuen nicht wissen, auf welchen Daten diese Benachteiligungen beruhen. Auch eine Einflussnahme auf die problematische Datenerhebung seitens der betroffenen Individuen ist nicht möglich. Aus diesem Grund ist eine Einschätzung der Datenqualität und der möglichen Risiken von Big Data-Anwendungen gerade im öffentlichen Sektor ganz besonders wichtig.

Ein weiterer Vorbehalt gegenüber Big Data basiert auf der Befürchtung, dass das gesellschaftliche Verständnis von angemessenen Entscheidungsverfahren (privat, politisch und in der Wissenschaft) vom Prinzip der „Kausalitäten“ zum Dogma der „Korrelationen“ wechseln könnte. In einer Big Data-geprägten Welt wäre es nicht mehr zentral, Phänomene zu verstehen, sondern allein Wirkungen zu identifizieren. Eine Problemlösung würde nicht mehr auf mögliche Ursachen abzielen, sondern ausschließlich nach wirksamen Maßnahmen suchen.⁴³ Auf die Politik übertragen könnte ein entsprechendes Dogma bedeuten, dass wichtige Entscheidungen nicht mehr durch Menschen, sondern durch Algorithmen getroffen werden.⁴⁴

37 Vgl. Executive Office of the President (2014): Big Data and Privacy: A Technological Perspective. Report to the President by the Presidents's Council of Advisors on Science and Technology, Washington.

38 Vgl. Weichert, Thilo (2013): Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Kiel. S. 10. Online unter: www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf [28.01.2015].

39 Vgl. Acquisti, Alessandro und Gross, Ralph (2009): Predicting social security numbers from public data. In: Proceedings of the National Academy of Science, Jg. 106, H. 27, S. 10975–10980.

40 Vgl. Mayer-Schönberger, Viktor und Cukier, Kenneth (2013): Big Data: A Revolution That Will Transform How We Live, Work and Think, John Murray, S. 150ff.

41 Vgl. Europäische Kommission (2011): Special Eurobarometer 359. Attitudes on Data Protection and Electronic Identity in the European Union, Brussels. Online unter: http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf [28.01.2015].

42 Vgl. Weichert, Thilo (2013): Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Kiel. S. 10. Online unter: www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf [28.01.2015]; Morozov, Evgeny (2014): The rise of data and the death of politics, the guardian, 20.07.2014. Online unter: www.theguardian.com/technology/2014/jul/20/rise-of-data-death-of-politics-evgeny-morozov-algorithmic-regulation [05.11.2014].

43 Vgl. Mainzer, Klaus (2014): Die Berechnung der Welt. In: Forschung und Lehre, München. Online unter: www.forschung-und-lehre.de/wordpress/?p=16900 [05.11.2014].

44 Vgl. Morozov, Evgeny (2014): The rise of data and the death of politics, the guardian, 20.07.2014. Online unter: www.theguardian.com/technology/2014/jul/20/rise-of-data-death-of-politics-evgeny-morozov-algorithmic-regulation [05.11.2014].

Aus einer evidenzbasierten Politik würde dann eine evidenzbestimmte, automatisierte Politik entstehen.⁴⁵ Dabei verfügen allein die gewählten Vertreter über die Legitimation, politische Entscheidungen zu treffen; dies gilt nicht für Algorithmen und jene, die sie programmiert haben.

Die beschriebenen wahrgenommenen Risiken führen dazu, dass Befürworter von Big Data-Anwendungen deren Akzeptanz in Deutschland als gering wahrnehmen.⁴⁶ Als Beispiel hierfür wird z. B. der Versuch des Unternehmens O₂ genannt, im Jahr 2012 über seine Mobilgeräte Standortdaten zu erfassen und zu Werbezwecken (anonymisiert) zu verwenden. Aufgrund des massiven öffentlichen Protests wurde das Projekt jedoch zurückgezogen.⁴⁷ Neben Protesten kann mangelnde gesellschaftliche Akzeptanz auch dazu führen, dass Nutzer ihre Daten in abnehmendem Maße zur Verfügung stellen. Deshalb ist ein verantwortungsvoller Umgang mit Daten nicht nur als Grenze für Big Data-Anwendungen zu betrachten, sondern auch als eine wichtige Grundlage.

5. Empfehlungen an die Politik

Die Nutzbarmachung von Big Data-Analysen im Sinne evidenzbasierter Politikgestaltung ist kein leichtes Unterfangen. Grundsätzliche Voraussetzungen sind die Entwicklung der Technologie und Datenverarbeitungsinfrastruktur, die es erlauben, Datenmengen von immenser Größe und Vielfalt möglichst zeitnah zu verarbeiten. Politik und Verwaltung sind gefragt, diese Entwicklungen auch für die öffentliche Hand zu nutzen. Hierfür müssen nicht zuletzt Rahmenbedingungen geschaffen werden, um Unsicherheiten, aber auch die missbräuchliche Verwendung der zur Verfügung stehenden Daten zu verhindern. Im Einzelnen wären deshalb folgende Maßnahmen sinnvoll:

1. *Kompetenz aufbauen*: In Politik und Verwaltung sollten die notwendigen Kompetenzen geschaffen werden, die es erlauben, die Potenziale von Big Data für den öffentlichen Sektor fruchtbar zu machen. Um mit technologischen und gesellschaftlichen Entwicklungen Schritt zu halten und in angemessener Form regulierend eingreifen zu können,

müssen Politik und Verwaltung ihre digitale Kompetenz entwickeln. Diese kann nicht in der Distanz entstehen, sondern bedarf des Kontakts und der Anwendung.

2. *Potenzial erproben*: Für eine bessere Einschätzung des Potenzials könnten entsprechende Demonstrationsvorhaben für Regierungsbehörden gefördert werden.
3. *Vorbild sein*: Der Staat hat die Chance, sich als Vorbild zu positionieren, was die Nutzung von Big Data-Anwendungen bei gleichzeitiger Wahrung von Persönlichkeitsrechten anbelangt. Ob staatliche Akteure die Auswertungen selbst vornehmen oder kommerzielle Angebote in Anspruch nehmen: Es muss stets geprüft werden, ob die genutzten Daten rechtmäßig erhoben und gespeichert wurden. Zudem sollten die Datenerhebungsmethode, ggf. der Datensatz selbst und das Auswertungsverfahren (inklusive der verwendeten Algorithmen) offengelegt werden. Nicht zuletzt sollte auch transparent gemacht werden, in welcher Form die Ergebnisse der Anwendung in die Politikgestaltung einfließen.
4. *Leitlinien entwickeln*: Staatliche Akteure sollten Prinzipien oder Leitlinien für die Verwendung von Big Data-Analysen im öffentlichen Sektor erarbeiten. Zu diesem Zweck könnten sie auch auf Good-Practice-Beispiele und „No Gos“ verweisen.
5. *Rahmenbedingungen justieren*: Wenn der Staat das Potenzial von Big Data-Anwendungen selbst nutzen will, scheint geboten, dass er sich parallel für den Schutz seiner Bürgerinnen und Bürger vor einer illegitimen Nutzung von Daten einsetzt. Hier müssen zum einen relevante rechtliche Normen auf nationaler wie internationaler Ebene ausgearbeitet werden, z. B. durch eine Reform des Datenschutzgesetzes. Zum anderen kann der Staat weiterhin die Forschung über die Chancen und Risiken der Nutzung von Big Data fördern. In beiden Bereichen hat die Bundesregierung den Handlungsbedarf erkannt und formuliert entsprechende Ziele in ihrer Digitalen Agenda⁴⁸. Darüber hinaus hat der Bundesinnenminister Thomas de Maizière erklärt, das

45 Vgl. Urbanski, Jürgen (2013): Projekt mit Breitenwirkung. Vielfacher Ertrag aus der Datenflut. In: VITAKO AKTUELL (2013): Big Data. Enorme Ressourcen, Zeitschrift der Bundes- und Arbeitsgemeinschaft der Kommunalen IT-Dienstleister e.V., Heft 1, Februar 2013, S. 10–11. Online unter: <http://www.vitako.de/Publikationen/Documents/Vitako%20aktuell%201-2013.pdf> [05.11.2014].

46 Selbst in den USA, die sich durch einen weitaus weniger kritischer Umgang mit Big Data auszeichnen, ist in jüngster Zeit das Bewusstsein für die Bedeutung des Datenschutzes gestiegen, wie ein Expertenbericht, der im Auftrag des Präsidenten angefertigt wurde, verdeutlicht. Vgl. Executive Office of the President (2014): Big Data and Privacy: A Technological Perspective. Report to the President by the President's Council of Advisors on Science and Technology, Washington.

47 Verweise auf den O₂-Skandal und die Sensibilität, mit der Big Data-Anwendungen in Deutschland als Gefahr für den Datenschutz angesehen werden, finden sich z. B. Weichert, Thilo (2013): Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Kiel, S. 8. Online unter: www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf [28.01.2015] und Ulbricht, Carsten (2013): Schwieriges Terrain: Big Data ist eine Herausforderung für den Datenschutz. In: Vitako Aktuell, Zeitschrift der Bundes-Arbeitsgemeinschaft der Kommunalen IT-Dienstleister e.V., Ausgabe 1/2013, S. 14.

48 Vgl. Die Bundesregierung (2014): Digitale Agenda 2014 – 2017, Berlin/Bonn. Online unter: www.digitale-agenda.de/DA/Redaktion/DE/Publikation/digitale-agenda.pdf [28.01.2015].

Potenzial von Big Data in Deutschland leichter und sicherer nutzbar machen zu wollen.⁴⁹ Diesen Weg gilt es weiter zu gehen, damit Behörden, Unternehmen und Individuen sich in einem rechtssicheren Raum bewegen können.

6. Gesellschaftliche Auseinandersetzung fördern: Öffentliche Behörden sollten schließlich aktiv zu einer Sensibilisierung

für die Vor- und Nachteile, die Chancen und Risiken des Einsatzes von Big Data-Anwendungen beitragen – unter Behörden und auch in der Öffentlichkeit. Denn am Ende handelt es sich hier um Entscheidungen, die Juristen nicht allein treffen können. Vielmehr müssen gesellschaftliche Debatten und Präferenzen berücksichtigt werden.

⁴⁹ Vgl. Bundesministerium des Innern (2014): Expertenrunde „Big Data – eine Herausforderung für den Datenschutz“, Nachricht des Bundesministerium des Innern vom 22.08.2014, Berlin. Online unter: www.bmi.bund.de/SharedDocs/Kurzmeldungen/DE/2014/08/expertenrunde-big-data.html [13.11.2014].

Kontakt:

Institut für Innovation und Technik (iit)
Steinplatz 1, 10623 Berlin

Lena Ulbricht

Tel.: 030 310078-362
E-Mail: lena.ulbricht@vdivde-it.de

Michael Huch

Tel.: 030 310078-363
E-Mail: michael.huch@vdivde-it.de

Ferdinand Gens

Tel.: 030 310078-334
E-Mail: ferdinand.gens@vdivde-it.de

iit perspektive Nr. 24

Februar 2015
Layout: Jennifer Büttner,
Anne-Sophie Piehl