

Rezension zu:

Stockmann, Reinhard/Meyer, Wolfgang (Hg.): Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven. Band 13 der Reihe Sozialwissenschaftliche Evaluationsforschung. Münster: Waxmann 2017. 260 Seiten, 34,90 EUR, ISBN 978-3-8309-3708-1

*Stefanie Krapp*¹

Angesichts neuartiger Phänomene wie „Alternative Fakten“ und „Fake News“ und der politischen Ratlosigkeit hinsichtlich globaler Entwicklungen und Krisen sollte angenommen werden können, dass Evaluation als Instrument und Profession auch in Zukunft weiter an Bedeutung gewinnen wird. Ist das wirklich so? Oder gibt es aus praktischer oder theoretischer Perspektive Herausforderungen, die dem entgegenstehen? Gibt es einen gemeinsamen globalen Pfad der Evaluation oder wird die kulturelle Vielfalt zu einer Vielzahl von unterschiedlichen Richtungen führen? Die Herausgeber und Autorinnen und Autoren der Publikation wagen einen Blick mit insgesamt 12 Artikeln in die Zukunft der Evaluation. 10 dieser Artikel sind bereits in der englischen Ausgabe „The Future of Evaluation: Global Trends, New Challenges, Shared Perspectives“ bei Palgrave Macmillan (2016) erschienen und wurden für diese Publikation ins Deutsche übersetzt und teilweise aktualisiert.

Um diesen Blick nicht zu weit schweifen zu lassen, schauen die 15 Autorinnen und Autoren auf Basis einer Betrachtung vergangener Entwicklungstendenzen aus drei Perspektiven in die Zukunft und stellen die jeweilige prognostizierte Weiterentwicklung der Evaluation zur Diskussion: 1. Institutionalisierung von Evaluation in der Gesellschaft, 2. Institutionalisierung von Evaluation im Wissenschaftssystem, 3. Institutionalisation von Evaluation als eigenständige Profession.

Aufgrund der Vielfalt von Informationen, die diese Publikation enthält, wird im Folgenden nur auf einige Beiträge eingegangen.

Welche Rolle Evaluation in der Gesellschaft spielen kann und wie sich die Funktionen und Aufgaben von Evaluation in der Zukunft

verändern werden, sind die Themen der ersten eingenommenen Perspektive „Institutionalisierung von Evaluation in der Gesellschaft“. Die zentrale Fragestellung lautet: Gibt es einen Globalisierungsprozess, der zu einer gemeinsamen Evaluationskultur und vergleichbaren Institutionalisationstendenzen in den meisten Ländern führt? Hierzu finden sich die Artikel von Peter Dahler-Larsen („Die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft“), von Reinhard Stockmann („Die Zukunft der Evaluation in modernen Wissensgesellschaften“), von Pablo Rodriguez-Bilella und Maria Alejandra Lucero („Evaluation als globales Phänomen: Die Entwicklung transnationaler Netzwerke“), von Andre Martinuzzi und Wolfgang Meyer („Evaluation nachhaltiger Entwicklung in einer globalen Gesellschaft“) und von Jan-Eric Furubo („Evaluation – ihr Erbe und ihre Zukunft“).

Peter Dahler-Larsen untersucht sehr anschaulich die sich verändernde Rolle der Evaluation in einer sich verändernden Gesellschaft aus konstruktivistischer Perspektive. Er sieht Evaluation und ihre Entwicklung in die jeweiligen Gesellschaftsprozesse eingebunden bzw. ihre Konstruktionen eingebettet. Für dieses Verhältnis sind fünf Themen zentral, die er zur Diskussion stellt: Popularisierung, Systematisierung, Outcome-Orientierung, Beziehungen zwischen Evaluation und Forschung, Nutzungsmuster. Mit Popularisierung beschreibt er die Entwicklung von Evaluation als ein Teil der öffentlichen Kultur, deren große Versprechungen aber nicht alle eingelöst werden. Systematisierung betrifft die Integration von Evaluation in Organisationen und Managementabläufe, die wenig flexible „Evaluationsmaschinen“ hervor-

1 International Program for Development Evaluation Training (IPDET) am Zentrum für universitäre Weiterbildung der Universität Bern

gebracht hat. Der Trend der Outcome-Orientierung bedient die zunehmende Nachfrage von Auftraggebern nach kausalen Zusammenhängen. Gleichzeitig wird Gesellschaft komplexer, was es schwieriger macht, eindeutig belegbare Evidenzen zu generieren. Eine zunehmende konzeptionelle und institutionelle Entfremdung von ihren akademischen Wurzeln kennzeichnen die Beziehung zwischen Evaluation und Forschung. Das klassische Problem der Nichtnutzung von Evaluation hat sich durch die Berücksichtigung kontextueller Unterschiede für Nutzungssituationen, die Bedeutung des Evaluationsprozesses und konstitutive Aspekte überholt. Dahler-Larsen schließt mit einem positiven Fazit, dass der Evaluationsbereich aus der Auseinandersetzung mit diesen Themen gestärkt hervorgehen kann, wie bereits frühere Evaluationsgenerationen bei anderen Problemen.

Rodriguez-Bilella und Lucero stellen gut aufbereitet und informativ, aber rein deskriptiv und recht brav transnationale Evaluierungsnetzwerke vor und schreiben ihnen eine wachsende Bedeutung zur Förderung von Good Practice, zur Zusammenarbeit und zum Wissensaustausch auf dem Gebiet der Evaluation durch die Bündelung von Kräften in bestimmten Regionen zu. Dadurch können Vernetzungsangebote, Aktivitäten zur Kapazitätsentwicklung und Aktionen, die auf die Förderung der Qualität von Evaluationen zielen, länderübergreifend grössere Wirkung entfalten. Ein wachsendes Bestreben nach Good Governance, Stärkung der Zivilgesellschaft und Unterstützung der Nachfrage nach Evaluation, sowie mehr Raum für einheimische Evaluationsansätze und Methodenvielfalt als akzeptierte Orthodoxie werden als Trends identifiziert, die den transnationalen Evaluierungsnetzwerken gemeinsam sind. Dabei stellen die Autorin und der Autor einige interessante Aspekte heraus, die jedoch nicht nur für transnationale Evaluationsnetzwerke gelten, sondern für Evaluationsgesellschaften generell. So bleibt die spezifische Rolle transnationaler Netzwerke schwammig. Spannender wäre eine kritische Reflexion der rasanten Entwicklung von Netzwerken des globalen Südens, die die globale Evaluierungscommunity zunehmend unter Druck setzen, oder Ideen dazu, was sich tun müsste, damit nationale und transnationale Netzwerke mehr Einfluss auf Entscheidungsträger gewinnen.

Furubo beginnt seinen Beitrag mit der Feststellung, dass Evaluation im Großen und Ganzen kaum Kritik ausgesetzt ist. Er hat dafür zwei Erklärungen: 1) Evaluation wird seit langem von mächtigen Strukturen wie der Weltbank, der OECD und der EU betrieben; 2) Eva-

luation repräsentiert zwei Dinge, die allseitig geschätzt werden: Sie kann mithilfe von allgemeinem Menschenverstand beschrieben werden und sie ist Teil einer rationalistischen Herangehensweise. Furubo sieht die geschützte Position der Evaluation insbesondere durch die fundamentale Frage, ob sie wirklich Nutzen bringt, wanken und versteht seinen Beitrag als Teil einer aufkommenden evaluativen Diskussion über Evaluation selbst. Um zu erklären, dass die Verbreitung der Evaluation von der Verbreitung einer bestimmten Praxis in den USA der 1960er Jahre herrührt, stellt er ausführlich die Entwicklung der Evaluation dar. Wirklich interessant wird es in den beiden letzten, leider im Verhältnis eher kurzen Kapiteln, wo er aufzeigt, dass die Welt nicht mehr in inkrementellen Prozessen verläuft und sich daher die Evaluation grundlegend ändern muss. Im inkrementellen Zeitalter wollte man mit Evaluation aus Vergangenen lernen, um Interventionen besser zu machen. Dies war eine einfache Logik. Aber bei einer Welt im Wandel, zunehmender Komplexität, Vernetzung, Welthandel etc. geht es vielmehr darum, neue Wege zu beschreiten. Furubo folgert daraus: 1) die Evaluation müsste Informationen dafür bereitstellen (eher ex-ante), 2) Evaluation ist nur eine Wissensquelle, 3) Adressaten von Evaluationen sind nicht mehr vorrangig Entscheidungsträger. Die Relevanz für die Nutzung ist, dass eine breit aufgestellte Community die Evaluationen aufnimmt und verwertet.

Während sich der erste Teil des Buches mit den gesellschaftlichen Rahmenbedingungen von Evaluation beschäftigt, steht im zweiten die Evaluation als Element der Universitäten und Hochschulen im Zentrum der Betrachtung. Unter der Perspektive „Institutionalisierung von Evaluation im Wissenschaftssystem“ finden sich die Artikel von Verena Friedrich („Evaluationsstudiengänge an europäischen Hochschulen: Merkmale und Herausforderungen“), von Christine Nolte und Sandra Schopper („Moderne Lehr- und Lernmethoden in der Evaluation“), von Frans Leeuw („Cyber Society und die ‚Verdschungelung‘ von Regeln. Neue Herausforderungen für die Evaluation im 21. Jahrhundert“) und von Reinhard Stockmann und Wolfgang Meyer („Evaluation zwischen Wissenschaft und Praxis“).

Nolte und Schopper befassen sich in ihrem Beitrag mit digitalen Vermittlungsmöglichkeiten von Evaluation, wie e-Learning und Blended Learning. Entsprechende Angebote nehmen weltweit zu. Die Autorinnen sehen insbesondere für die fortschreitende Verbreitung und Professionalisierung von Evaluation in den Ländern des globalen Südens neue Optionen, da durch

die Bildungstechnologien das Wissen schneller auch in marginalisierte Regionen getragen werden kann. Zeitlich und räumlich ungebundene Verfügbarkeit von Informationen und die Möglichkeit aktiver zu Lernen sind klare Vorteile von Lernformaten neuer Medien und entsprechender Medienpädagogik. An ein solches Lernarrangement werden bestimmte Anforderungen gestellt, wie beispielsweise die Einfachheit der Bedienung, der wahrgenommene Nutzen des Produkts, oder die Möglichkeit des selbstorganisierten Lernens. Die Erläuterungen dazu sind jedoch sehr allgemein gehalten und den Autorinnen gelingt nicht die Verbindung mit den dargelegten Strategien zum Aufbau von Evaluation Capacity Building. Ein wichtiger informativer Teil des Beitrags ist die Aufstellung universitärer und außeruniversitärer Online-Aus- und Weiterbildungsangebote. Eine entsprechende Publikation in einer englischsprachigen Zeitschrift würde mit Sicherheit das Interesse vieler bedienen, denn eine Gesamtübersicht der inzwischen vielfältigen Angebote gibt es bisher nicht.

Der Beitrag von Stockmann und Meyer greift anregend in die fortwährende Diskussion um das Spannungsverhältnis zwischen Wissenschaftlichkeit und Nützlichkeit von Evaluationen ein. Zum einen soll Evaluation wissenschaftliche Standards einhalten, zum anderen soll sie den definierten Zwecken dienen. Dies impliziere, dass die Praxis v.a. an der Beantwortung von ‚Wie‘-Fragen und die Wissenschaft v.a. an ‚Warum‘-Fragen interessiert sei. Inwieweit dieses Verhältnis zwischen Wissenschaft und Praxis austariert ist, verdeutlichen die Autoren anhand der zentralen Evaluationsfunktionen Erkenntnis, Lernen, Legitimation und Kontrolle. Als generellen Trend erkennen sie mehr Komplexität und höhere Erwartungen an die Qualität von Evaluationen, wodurch mehr Wissenschaftlichkeit eingefordert wird. Dem gegenüber steht aber auch ein Trend zu stärkerer praktischer Verwertbarkeit. Von Evaluatorinnen und Evaluatoren wird daher erwartet, dies bei jeder Evaluation neu auszutarieren, so dass sie immer mehr Aufgaben jenseits der wissenschaftlichen Informationsproduktion erfüllen sollen bzw. müssen. Die Autoren schlussfolgern, dass entweder eine praxisorientierte Servicefunktion oder eine qualitätsorientierte Wissenschaftlichkeit die Oberhand gewinnen kann. Diese von den Autoren skizzierte Dichotomie kann zumindest mit einem Fragezeichen versehen werden. Schließt Praxisorientierung per se Wissenschaftlichkeit aus? Auch Auftraggeber sind an den Kausalmechanismen interessiert, die ihre Interventionen wirken lassen oder eben nicht (siehe auch die Beiträge von Dahler-Larsen und Müller). Und die

Beantwortung der Wie-Fragen muss ebenso nach wissenschaftlichen Standards erfolgen.

Im dritten Teil des Buches geht es um die „Institutionalisierung von Evaluation als eigenständige Profession“ und damit um die Ausarbeitung eigenständiger Theorien und Methoden. Dem sind die Artikel von Wolfgang Meyer, Evelyn Funk und P. Linh Nguyen („Partizipation und Bewertung: Weiterbestehende methodologische Herausforderungen“), von Christoph Müller („Kausale Wirkungsevaluation zwischen methodischem Anspruch und empirischer Praxis“) und von Brad Astbury („Von der Evaluationstheorie zum Testen von Evaluationstheorien?“) zugeordnet.

Meyer, Funk und Nguyen diskutieren die spezifischen Herausforderungen, denen Evaluationen ausgesetzt sind, wenn sie Stakeholder mit in ihre Arbeit einbeziehen und in ihren Berichten Bewertungen vornehmen. Ihre umfassende Literaturanalyse zeigt, dass es kaum Publikationen zu den Einflüssen persönlicher Sichtweisen auf den Bewertungsprozess in einer Evaluation gibt. Demgegenüber gibt es eine lebhaftige Debatte über partizipatorische Ansätze. Allerdings fehlt es immer noch an einer systematischen Forschung über Partizipation in Evaluationen und deren Auswirkungen, ebenso an wissenschaftlichen Methoden, die einen Bewertungsprozess korrekt anleiten können. Die dem Artikel zugrunde liegende Frage, wie Beteiligungs- und Bewertungsprozesse organisiert sein müssen, damit die Wissenschaftlichkeit einer Evaluation nicht verloren geht, ist sicherlich relevant, greift aber zu kurz. Hier hätte sich durchaus angeboten, darzulegen, inwiefern partizipative Ansätze oder zumindest Vorgehensweisen durchaus adäquater als herkömmliche ‚wissenschaftliche‘ Verfahren im Rahmen von Evaluationen sein können. Andererseits werden Grenzen der Beteiligung nicht deutlich.

Müller widmet sich in seinem Beitrag dem Spannungsfeld zwischen den methodischen Ansprüchen, die an Evaluierende bei der Anwendung komplexer Untersuchungsdesigns gestellt werden und der praktischen Umsetzung dieser Designs. In der Diskussion um die unterschiedlichen Ansätze kausaler Wirkungsevaluationen macht er deutlich, dass die valide Schätzung kausaler Interventionswirkungen eine äußerst anspruchsvolle und ressourcenintensive Aufgabe ist, die in der Evaluationspraxis oft nicht geleistet werden kann. Dennoch sieht Müller folgende Trends: steigende Anzahl von Wirkungsevaluationen, zunehmende Orientierung von Evaluationen an Outcomes, Wirkungsevaluationen tragen zum Verständnis der Wirkungsweise von Interventionen und zur Ermittlung von

Optimierungspotenzialen bei, Erweiterung der entsprechend notwendigen Kompetenzen von Evaluierenden, Verbesserung der Evaluations- und Methodenkompetenzen von Auftraggebern, Weiterentwicklung der informationstechnischen Rahmenbedingungen. Dem Autor gelingt auf wenigen Seiten, gut verständlich Ansätze von Wirkungsevaluationen zu vermitteln, deren Grenzen und Möglichkeiten zu verdeutlichen, Bezüge zu anderen Artikeln in dem Buch herzustellen und konkrete Zukunftsperspektiven aufzuzeigen.

Liest man erstmal nur das Inhaltsverzeichnis der Publikation wird der Eindruck eines bunten Sammelsuriums an Artikeln zu Evaluation geweckt, wie das eben bei Bänden, die viele Artikel unterschiedlicher Autorinnen und Autoren enthalten, oft der Fall ist. Aber bereits nach der Einleitung wird deutlich, hier haben sich die Herausgeber Gedanken gemacht, unter welchen Perspektiven sie die Zukunft der Evaluation betrachten. Natürlich könnte man geneigt sein, zu konstatieren, dass die Zukunft der Evaluation von vielen weiteren oder gar ganz anderen Themen bestimmt sein wird. Sicherlich. Das kann und sollte in weiteren Publikationen nachgeholt werden. Neben den einzelnen sehr interessanten und lesenswerten Artikeln erfährt der Leser/die Leserin das große Plus der Publikation im letzten Kapitel. Es ist beeindruckend, wie Stockmann und Meyer eine Synopse zu den zentralen Befunden zur Zukunft der Evaluation unter Ausdifferenzierung der in den Artikeln dargelegten gegenwärtigen Situation, der prognostizierten Trends und der Bedeutung für die jeweilige Institutionalisierung von Evaluation gelingt. Dabei stellen sie fest, dass die meisten Diagnosen inhaltlich stark übereinstimmen.

Die Beiträge zur Institutionalisierung von Evaluation in der Gesellschaft deuten darauf hin, dass v.a. auf transnationaler Ebene die Institutionalisierung recht weit fortgeschritten ist, was v.a. auf Aktivitäten großer internationaler Organisationen zurückzuführen ist. Die Herausgeber erkennen jedoch zwei wesentliche Probleme: Zum einen die Popularisierung der Evaluation und die daraus resultierenden Erwartungshaltungen, die die Evaluation in ihrer derzeitigen Verfassung enttäuschen muss, was z.T. auf die fehlende Professionalität von Evaluationen zurückzuführen ist. Und zum anderen die zunehmende Ritualisierung von Evaluation, die damit zum verwaltungstechnischen Vorgang mutiert und die Gefahr birgt, den eigentlichen Zweck des Lernens aus dem Blick zu verlieren.

Eine Institutionalisierung der Evaluation im Wissenschaftssystem können die Herausge-

ber nur ansatzweise konstatieren. Folglich ergeben sich entsprechende Defizite in der Aus- und Weiterbildung und hinsichtlich des Aufgreifens neuer Entwicklungstrends in der Evaluationsforschung und -praxis. Zwar gibt es Aus- und Weiterbildungskurse an Universitäten und Hochschulen, aber Evaluation ist keine ausreichend fundierte Wissenschaftsdisziplin geworden. In den verschiedenen Beiträgen werden bestimmte Gründe für die gebremste Entwicklung im akademischen Bereich angesprochen, so z.B. die zunehmende Distanzierung von Praktiker(innen) von der Wissenschaft und ihrer Arbeitsweise. Diejenigen, die Qualitätssteigerungen der Evaluationen und ein professionelleres Arbeiten einfordern, sind in der Minderheit. Noch viel zu oft werden immer noch erfahrene Gutachter(innen) ohne professionellen evaluatorischen Hintergrund von Auftraggebern vorgezogen.

Dass die Institutionalisierung als wissenschaftliche Disziplin noch nicht gelungen ist, schließen die Herausgeber aus der geringen Anzahl an Forscher(inne)n, die sich mit Evaluation als Forschungsgegenstand beschäftigen. Zudem ist die Forschung über Evaluation interdisziplinär, so dass Publikationen nicht den gleichen Nutzwert wie die der eigenen Fachdisziplin haben. Dies führt dazu, dass sich eine Auseinandersetzung mit Evaluation stärker an Grundlagen einzelner sozialwissenschaftlicher Teilbereiche orientiert. Die Institutionalisierung der Evaluation als eigenständige Profession wird von den Herausgebern trotz bestehender Schwachstellen mit einem positiv stimmenden Ausblick belegt. Steigende Publikationen und der dadurch wachsende Wissensbestand lassen die Evaluation als Profession einen Reifegrad erreichen, der die Entwicklung eigenständiger Forschungsprogramme und deren Verankerung in den bestehenden wissenschaftlichen Forschungssystemen ermöglicht.

Abschließend konstatieren Stockmann und Meyer Defizite einer mangelhaften Institutionalisierung von Evaluation, die die kontinuierliche Weiterentwicklung von Evaluation angesichts neuer Herausforderungen beeinträchtigen. Obwohl einzelne Autorinnen und Autoren des Bandes skeptisch sind, dass diesbezüglich absehbar Fortschritte erzielt werden können, schlussfolgern die Herausgeber eine „erstaunliche Leistungsfähigkeit der Evaluation“ und schließen mit einem hoffnungsvollen Ausblick: „Bisher hat sich die Evaluation besser entwickelt als es die überkritischen Evaluierungen zu hoffen wagten. Es spricht nichts dagegen, dass dieser Trend erhalten bleibt“ (S. 257). Möge es so kommen ...

Rezension zu:

Margit Seckelmann: *Evaluation und Recht. Strukturen, Prozesse und Legitimationsfragen staatlicher Wissensgewinnung durch (Wissenschafts-)Evaluationen.* Tübingen: Mohr Siebeck 2018. 685 Seiten, 129,00 EUR, ISBN 978-3-16-154390-6

*Wolfgang Meyer*¹

Die Evaluation von Gesetzesfolgen gehört in den letzten Jahren zu den Gebieten, die am meisten an Bedeutung gewonnen haben. Gleichwohl fehlte es an einem Kompendium, welches sich ausführlich diesem Thema widmet. Margit Seckelmann hat mit einem knapp 700 Seiten starken Werk versucht, diese Lücke zumindest teilweise zu schließen. Es handelt sich dabei um die Publikation ihrer Habilitationsschrift an der Universität für Verwaltungswissenschaften in Speyer, die sich zum führenden Standort der Evaluation in den deutschen Verwaltungen entwickelt hat.

Das Buch gliedert sich in vier Hauptabschnitte. Unter der Rubrik „Grundsätzliche Überlegungen“ diskutiert die Autorin Evaluation als zentrales Instrument der Wissensgenerierung für den Staat im Rahmen des „New Public Management“ und die damit verbundenen Schwierigkeiten. Einige der Aussagen muten dabei sehr gewagt und nur bedingt belegt an. So heißt es z.B. gleich auf Seite 5: „Auch international ist in jüngerer Zeit eine Wende zu evidenzbasiertem staatlichem Handeln zu erkennen“. Angesichts eines US-amerikanischen Präsidenten, welcher die Ignoranz gegenüber Fakten sozusagen zum Politikprinzip erhoben hat, und einer Reihe mehr oder weniger gut nachgewiesener „Fake News“ nicht nur in den USA erscheint diese Aussage mehr als fraglich. Die angeführten Beispiele aus der Schweiz und Schweden belegen dabei eher eine langandauernde Tradition rationalen, wissensbasierten Regierens denn eines neuen Trends. Auch in diesen beiden Ländern kann gegenwärtig bezweifelt werden, ob tatsächlich evidenzbasiertes staatliches Handeln auf dem Vormarsch ist oder ob dieses nicht durch politische Vorurteile, populistische Hass-

tiraden und eine Emotionalisierung öffentlicher Debatten zunehmend in Frage gestellt wird.

Dieses Beispiel offenbart eine der größten Schwächen – und in mancher Sicht auch eine Stärke – der Arbeit: sie fokussiert sich auf die Perspektive der Verwaltung auf Evaluation und hebt die Entwicklungstrends im deutschen Rechtssystem einseitig hervor. Viele Wahrheiten des Buchs sind deshalb vor allem Wahrnehmungen der staatlichen Verwaltung. So ist z.B. die Einordnung der Evaluation in den Kontext der Diskussionen zur Folgenabschätzung von Gesetzen durchaus gelungen, die Reduktion der Evaluationsgeschichte auf diesen Aspekt ohne Berücksichtigung der Eigendynamik in verschiedenen Politikbereichen (z.B. in der Bildung, im Gesundheitswesen oder in der Entwicklungspolitik) ist aber als Schwachpunkt zu sehen.

Der Autorin geht es um den Ge- und Missbrauch von Evaluation als Instrument des New Public Management und nicht um die zur Gesetzesfolgenabschätzung eingesetzten Methoden und Verfahren. Dies wird vor allem in dem den „grundlegenden Überlegungen“ folgenden allgemeinen Teil B „Evaluationen in der Wissensgesellschaft“ deutlich, welcher die Bedeutung des Wissens für staatliches Handeln und „policy making“ unter juristischen Gesichtspunkten beleuchtet. Diese theoretischen Überlegungen werden dann im „Besonderen Teil C“ anhand von Praxisbeispielen evaluativer Verfahren in der Wissenschaft und Wissenschaftspolitik diskutiert. Die Arbeit endet mit „Schlussfolgerungen“ in Teil D in einer juristischen Bewertung (und Verallgemeinerung) der an den Wissenschaftsevaluationen gewonnenen Erkenntnissen, die zu einer besseren rechtlichen Einbettung von Evaluationen in dem Kontext eines Wissenschafts-

¹ Centrum für Evaluation (CEval), Saarbrücken

kooperationsrechts führen soll. Dies impliziert die in der Arbeit nicht in Frage gestellte Vermutung, dass Wissenschaftsevaluation repräsentativ für die gesamte Evaluationslandschaft ist und die dort auftretenden (juristischen) Probleme in anderen Anwendungsfeldern (z.B. an den Schulen) in gleicher oder zumindest sehr ähnlicher Form auftreten.

Für nicht speziell an Wissenschaftsevaluation interessierte Leser(innen) ist sicher der allgemeine Teil B „Evaluationen in der Wissensgesellschaft“ am spannendsten, weil hier einige grundlegende und sehr tiefgehende Analysen zum verwaltungstechnischen Umgang mit Wissen vorgestellt werden. Angesichts der Nutzung neuer Medien wird die Wissensproduktion und -nutzung stark vereinfacht, wobei die politische Wissenssteuerung mit der technischen Entwicklung kaum schritthalten kann. Eine neue „Informationsordnung“ ist erst in verwaltungswie verfassungsrechtlichen Konturen zu erkennen (S. 25). Aus juristischer Sicht ergeben sich Probleme besonders im Zusammenspiel von privaten und staatlichen Anbietern in Form einer „Wissensgovernance“ (S. 31), die bisher nicht oder nur unzureichend reguliert ist. Hier gilt das Diktum im Zwischenfazit (S. 66), dass mit steigender Wissensproduktion und Vielfalt der Quellen die Möglichkeiten einer staatlichen Steuerung und Regulierung sinken.

Die Notwendigkeit einer Wissensproduktion durch den Staat selbst ergibt sich aus den Gesetzgebungsverfahren. Hier kommt der Verwaltung im Rahmen ihrer „Informationsfunktion“ (S. 36) die Aufgabe zu, während der Politikvorbereitung über die entsprechenden Folgen der Gesetzgebung zu informieren und sie damit einer öffentlichen Diskussion zu öffnen. Verwaltungstätigkeit ist in dieser Hinsicht immer auch ‚Wissensarbeit‘, welche eine rechtliche Regelung des Zugangs zu verwaltungstechnisch produziertem Wissen erfordert. Es entsteht ein „Informationsverwaltungsrecht“ (S. 38), in dessen Kontext auch das „Evaluationsrecht“ einzuordnen wäre. Evaluationen sind dieser Argumentation zu Folge „Instrumente legislativer Wissensgenerierung und -verarbeitung“ (S. 45), die „gut begründete Plausibilitäten“ (S. 68) der Legislative bereitstellen.

Ferner leitet die Autorin aus dem Grundgesetz eine Verpflichtung zur Rationalität bei Gesetzgebungsverfahren ab, die angesichts der Rechtsentwicklung den Gesetzgeber zu einer Überprüfung der Gesetze verpflichtet. Diese lässt sich am ehesten in einem institutionalisierten Lernprozess gewährleisten („policy cycle“) und mündet in eine „politische Kyber-

netik“ (S. 156) mit Feedbackschleifen und Kontrollinstrumenten, bei denen Evaluationen dann eine zentrale Aufgabe übernehmen. Die juristische Verankerung von Evaluationen im Gesetzgebungsverfahren kann auf dreierlei Wegen erfolgen. Erstens können Gesetze mit „sunset clauses“ (S. 179) versehen werden und zu einem bestimmten Zeitpunkt außer Kraft treten. Eine Überprüfung des Gesetzes wird dann zwangsläufig im Rahmen eines neuen Gesetzgebungsverfahrens notwendig. Zweitens besteht die Möglichkeit, eine zeitlich fixierte Berichtspflicht im Gesetz zu verankern, die praktisch einer Wirkungsprüfung gleichkommt. Allerdings gibt es in der Bundesrepublik nicht wie in der Schweiz die Tradition einer „parlamentarischen Wirkungskontrolle“ (S. 181), welche dem Evaluationsbericht ein deutlich höheres politisches Gewicht zukommen lässt. Drittens schließlich besteht die Möglichkeit einer Kombination von Befristungs- und Berichtsklauseln, damit sich der Gesetzgeber mit seinen Normen turnusmäßig beschäftigen und hierfür Evaluationen als Grundlage heranziehen muss. Für jede dieser Vorgehensweisen gibt es in dem Buch Beispiele, wengleich die Autorin hier keine systematische Analyse der generellen Entwicklung vornimmt.

Die Autorin geht mit der tieferegreifenden Betrachtung der „Wissenschaftsevaluation“ einen anderen Weg (Abschnitt C). Unter dem Begriff der „Wissenschaftsevaluation“ fasst sie die Hochschul-, Forschungs- und Förderpolitikevaluation zusammen und behandelt sie im Folgenden wie eine Einheit. Problematisch ist an dieser Zusammenfassung u.a. die gemischte Zuständigkeit von Bund und Ländern sowie die unterschiedliche Form der Verbindung staatlicher und privater Akteure. Die Schwierigkeit wird z.B. im Kapitel 8 (S. 275ff.) bei der Darstellung der historischen Entwicklung von Wissenschaftsevaluation deutlich, in der sehr stark auf die Bundesebene und damit den Bereich der Forschungspolitik abgehoben wird. Die Entwicklung der Hochschulevaluation wird in erster Linie als Ableitung davon begriffen („Top-Down-Perspektive“).

Trotz dieser Einschränkungen sind diese Passagen für Personen, die an Forschungs- und/oder Hochschulevaluation interessiert sind, ausgesprochen spannend und gehaltvoll. Dies gilt insbesondere für Kapitel 9 zur (rechtlichen) Funktion der Evaluationen (S. 307ff.), in dem die Verbindung unterschiedlicher Traditionen und normativer Konzepte (z.B. Autonomie, Leistungsorientierung, Qualitätssicherung, Wettbewerb) im Institutionalisierungsprozess deutlich herausgearbeitet und problematisiert wird.

Das Methodenkapitel 10 fällt dagegen sehr kurz (15 Seiten), selektiv und wenig kritisch aus. Gerade hier wäre eine ähnlich ausführliche und systematische Darstellung aufgrund des kritischen (Fach-)Diskurses einer Vielzahl neuer Vorgehensweisen, Verfahren und Methoden wünschenswert gewesen.

Dies wirft zum Abschluss die Frage auf, für wen die Lektüre eines solch umfangreichen Werks lohnenswert erscheint. Wer sich für die Vorgehensweise und Verfahren der Gesetzesfolgenabschätzung interessiert und vorrangig an einer methodischen Diskussion interessiert ist, wird durch dieses Buch enttäuscht. Weder im Allgemeinen Teil (mit Bezug auf die Bewertung von Gesetzen) noch im Besonderen Teil (mit dem Fokus auf die Forschungs- und Hochschulevaluation) werden die eingesetzten Verfahren, ihre Möglichkeiten und Probleme sowie die neusten methodischen Entwicklungen geschildert und diskutiert.

Der Autorin geht es vielmehr (in ihren eigenen Worten) „um die Rolle von Evaluationsverfahren als Instrumente einer staatlichen Wis-

senschaftsbeschaffung“ (S. 525). Nicht die Evaluation, sondern deren Nutzung im Rahmen staatlichen Handelns bildet den Fokus der Arbeit. Hierzu liefert Margit Seckelmann ein facettenreiches Werk, welches durch seinen interdisziplinären Zugang und seine Vielfältigkeit beeindruckt. Es sind vor allem die juristischen Details (wie z.B. die institutionelle Verankerung von Gesetzesfolgenabschätzungen oder Wissenschaftsfreiheit), welche aus der Sicht der Evaluation interessant und für viele neu sein dürften.

Dabei bleibt für Evaluatoreninnen und Evaluatoren zumindest ein beruhigender Aspekt im Gedächtnis: die Evaluation ist in Deutschland im Rechtssystem und im Verwaltungsdanken mittlerweile schon so gut verankert, dass eine Trendumkehr mehr als unwahrscheinlich erscheint. Margit Seckelmann kann im Gegenteil verdeutlichen, dass eine weitere Verankerung systemimmanent notwendig ist und vermutlich in absehbarer Zeit erfolgen wird. Trotz aller Kritik und Skepsis gegenüber Evaluationen bleiben sie als Instrumente zur Wissensbeschaffung und Bewertung staatlichen Handelns unabdingbar.

Rezension zu:

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

Elisabeth Nindl¹

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

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

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

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

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

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

1 KMU Forschung Austria, Wien

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

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

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

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

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

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

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

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

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

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

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

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

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