

# **Zukunft der Aus- und Weiterbildung in der Markt- und Sozialforschung\***

Florian Keusch & Frauke Kreuter

## **Zusammenfassung**

Die Nachfrage nach gut ausgebildeten DatenwissenschaftlerInnen, die sowohl die Fähigkeiten besitzen, Daten auf „traditionellem Weg“ zu erheben und auszuwerten und ebenso mit großen semi- oder gar unstrukturierten Datensätzen zu arbeiten, steigt kontinuierlich an. In diesem Beitrag besprechen wir, welche Kompetenzen Sozial- und MarktforscherInnen heutzutage benötigen, um am Arbeitsmarkt erfolgreich zu sein. Wir diskutieren Herausforderungen und Chancen im Bereich der Lehre dieser neuen Inhalte und deren Potential, den steigenden Bedarf an Fachkräften im Bereich Datenerhebung und Datenanalyse in den kommenden Jahren zu decken.

## **Markt- und Sozialforschung im Wandel**

Daten über Verhalten und Einstellungen von Personen dienen häufig als Entscheidungsgrundlage für die Privatwirtschaft, den öffentlichen Dienst und gemeinnützige Organisationen. Grob kann man diese unterscheiden in Daten, die gezielt mit einem bestimmten Erhebungsdesign gewonnen werden und solchen, die organisch als Nebenprodukt administrativer oder anderer Prozesse entstehen. Viele Organisationen, vor allem aus der Privatwirtschaft, versprechen sich von „Big Data“ – meist definiert als Daten, die in großem Umfang (Volume), in Echtzeit gewonnen werden (Veracity) und eine komplexe Struktur aufweisen, da sie in verschiedenen Formaten und aus unterschiedlichen Quellen stammen (Variety) (Beyer und Laney 2012) – Effizienzsteigerung, Erkenntnisse für die Entwicklung neuer Geschäftsmodelle und größeres Wissen über verschiedene Kundengruppen (Capgemini 2017). Viele Unternehmen verstehen sich mittlerweile als “data-driven-companies” und/oder handeln anhand einer “data-driven-culture”, in der Daten aus den unterschiedlichen Unternehmensbereichen die Grundlage für unternehmerisches Handeln darstellen (Harms und Schmidt 2019). Aber auch statistische Ämter und andere Organisationen, die bisher traditionell eher nur Befragungsdaten erhoben haben, suchen vermehrt

---

\* Preprint von Keusch, F. & Kreuter, F. (2020). Zukunft der Aus- und Weiterbildung in der Markt- und Sozialforschung [The future of education in market and social research]. In Keller, B., Klein, H.-W., Wachenfeld-Schell, A., & Wirth, Th. (Eds.) Marktforschung für die Smart Data World. Chancen, Herausforderungen und Grenzen. Wiesbaden: Springer Gabler, 3-25. DOI: 10.1007/978-3-658-28664-4\_1.

nach Möglichkeiten, diese neuen digitalen Daten in den Forschungsprozess bzw. die amtliche Statistik zu integrieren. Während die Nutzung von Daten aus administrativen Prozessen in Europa schon seit längerem Tradition hat, stecken die Versuche mittels neuer digitaler Daten aus Webscraping, von Smart Meters und anderen Internet-of-Things (IoT) Technologien, Mobilfunkdaten etc. in Deutschland noch in der Testphase (Wiengarten und Zwick 2018). Hinter diesen Bemühungen steckt die Hoffnung, mit Hilfe von neuen Daten Sachverhalte besser, schneller, oder kostengünstiger darstellen zu können und gleichzeitig den Aufwand für StudienteilnehmerInnen zu reduzieren (Oppeln-Bronikowski 2018). In den Sozialwissenschaften finden sich die Aktivitäten zur Sammlung und Analyse von großen Datenmengen sowie ihrer Analyse zunehmend unter dem Begriff Computational Social Sciences (Lazer et al. 2009; Watts 2013; Stützer et al. 2018).

Gleichzeitig hat sich in den letzten Jahren aber auch gezeigt, dass diese neuen digitalen Datenquellen traditionelle Befragungen nicht ablösen werden, da die neuen digitalen Daten häufig nicht alle Aspekte abdecken können (Schnell 2018), besonders nicht jene, die für die amtliche Statistik von Interesse sind (Wiengarten und Zwick 2018). Im Jahr 2017 wurden von den Mitgliedern des Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) insgesamt 22 Millionen Interviews durchgeführt, der Großteil davon online (ADM 2019). Dieser Wert ist relativ stabil über die letzten fünf Jahre hinweg, beinhaltet jedoch keine Interviews, die von Organisationen durchgeführt wurden, die nicht Mitglied des ADM sind. Die Zahlen von neuen Playern der Marktforschungsszene, wie beispielsweise die beiden Berliner Startups Dalia Research, die laut Medienberichten in 96 Ländern 8,2 Millionen User-Antworten pro Tag bearbeiten (Penke 2017), und Civey, die laut eigener Website ca. 40.000 Meinungen pro Stunde sammeln (Civey 2019), sind hier nicht mit einbezogen. Ebenso steigt die Nachfrage nach und die Nutzerzahlen von "Do-it-Yourself"-Befragungstools. Dies deutet darauf hin, dass auch in der Privatwirtschaft traditionelle Datenerhebungsformen alles andere als ausgedient haben. Auf der Plattform von SurveyMonkey werden pro Monat weltweit ca. 90 Millionen Online-Fragebögen ausgefüllt und Qualtrics versendet etwa eine Milliarde Befragungseinladungen jährlich (Callegaro und Yang 2018). Qualtrics, ursprünglich als universitäres Spin-Off für „Experience Management“ gegründet, wurde 2018 für 8 Milliarden US Dollar von SAP, dem größten europäischen Softwarehersteller und Spezialisten für betriebliche Daten, gekauft (SAP News 2018).

Das zunehmende Interesse an der Nutzung von neuen Daten und der Verknüpfung von Daten aus verschiedenen Quellen geht einher mit einer gesteigerten Nachfrage nach gut ausgebildeten DatenwissenschaftlerInnen, die sowohl die Fähigkeiten besitzen, Daten auf „traditionellem Weg“ zu erheben und auszuwerten und ebenso mit großen semi- oder gar unstrukturierten Datensätzen zu arbeiten. In den traditionellen Disziplinen der akademischen Sozial- oder Wirtschaftswissenschaften werden diese neuen Inhalte oft noch nicht oder nur sehr selten vermittelt. Andererseits gibt es eine schnell wachsende Zahl an Data Science Studiengängen, in denen die Ausbildung allerdings oft sehr technisch orientiert ist und die Verknüpfung zu marktforschungsrelevanten und sozialwissenschaftlichen Fragestellungen zu kurz kommt. Vor diesem Hintergrund besprechen wir in diesem Kapitel welche Fähigkeiten Sozial- und MarktforscherInnen heutzutage benötigen, um am Arbeitsmarkt erfolgreich zu sein. Wir diskutieren ebenso Herausforderungen und Chancen im Bereich der Lehre dieser neuen Inhalte, wie beispielsweise digitales Lernen und berufsbegleitende Weiterbildung, die den steigenden Bedarf an Fachkräften in den Bereichen Datenerhebung und Datenanalyse in den kommenden Jahren decken soll.

### **Welche Kompetenzen brauchen Markt- und SozialforscherInnen heute?**

Die zunehmende Nutzung neuer Datenquellen führt dazu, dass das Berufsbild von Markt- und SozialforscherInnen im Umbruch ist. Für Studieninteressierte und Studierende, die eine Karriere in diesem Bereich anstreben, aber auch für Personen mit Berufserfahrung stellt sich die Frage, welche Kompetenzen und Fähigkeiten sie heute benötigen, um am Arbeitsmarkt erfolgreich zu sein. Zeitgleich fragen sich aber auch viele Arbeitgeber, was sie von BewerberInnen heute erwarten können. In diesem Abschnitt diskutieren wir einige der aus unserer Sicht wichtigsten Kompetenzen und Fähigkeiten, die eine solide Grundlage zum Arbeiten mit Daten aus traditionellen und neuen Quellen liefern können. Dabei orientieren wir uns am Task Force-Bericht der American Association for Public Opinion Research (AAPOR) zum Thema Big Data und Umfrageforschung (Japiec et al. 2015) und den Empfehlungen von Usher (2015). Wie Abbildung 1 zeigt, gibt es demnach fünf Kompetenzbereiche, die für das Arbeiten mit Daten relevant sind.

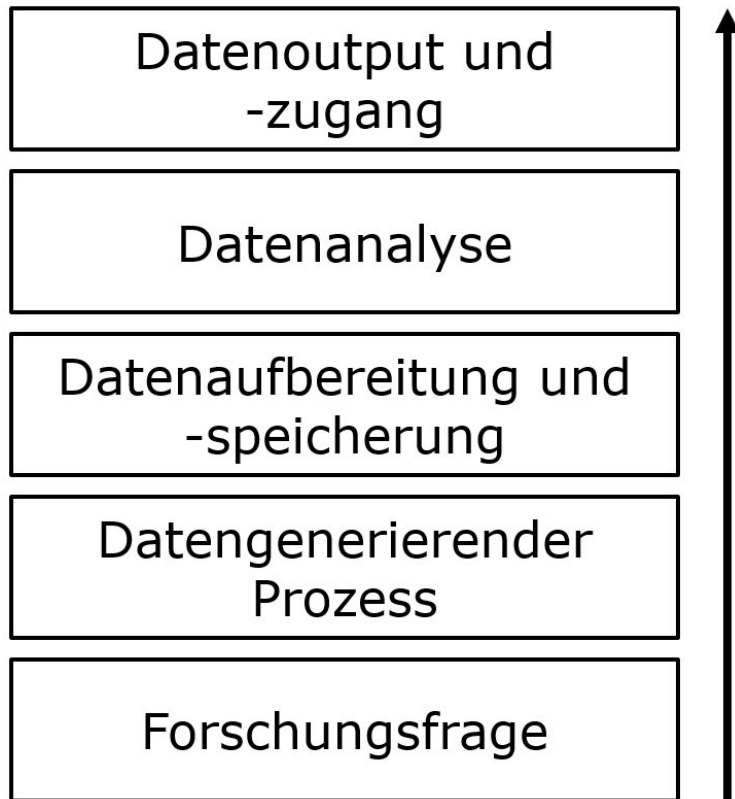


Abbildung 1. Fünf Kompetenzbereiche, notwendig zur erfolgreichen Nutzung von Daten (adaptiert und aus dem Englischen übersetzt nach Usher [2015])

### *Forschungsfrage*

Zunächst ist es essentiell einschätzen zu können, wie Forschungsfragen formuliert werden müssen und mit Hilfe welcher Daten überhaupt Antworten zu den entsprechenden Fragen gefunden werden können. Hierfür ist einerseits fundiertes Fachwissen aus den jeweils relevanten Disziplinen (z.B. Marketing, Vertrieb, Konsumentenpsychologie, Soziologie, Politikwissenschaften etc.) unerlässlich. Nur wer inhaltliche Zusammenhänge versteht, kann in weiterer Folge auch die richtigen Fragen formulieren und darauf aufbauend empirische Studiendesigns entwickeln, die valide Antworten auf diese Fragen liefern.

Darüber hinaus ist auch ein klares Verständnis darüber wichtig, welche Art von Forschungsfrage (deskriptiv, explorativ, inferentiell, prediktiv, kausal) überhaupt vorliegt und welches Forschungsdesign notwendig ist, um die entsprechende Frage zu beantworten (Leek und Peng 2015). Ob und wie diese Fragen beantwortet werden können, hängt davon ab, welches Forschungsdesign für die Erhebung der Daten gewählt wurde. In der Praxis häufig auftretende

Fehler sind beispielsweise der Versuch inferentielle Fragestellungen mit stark verzerrten Stichproben von Freiwilligen zu beantworten oder Korrelation mit Kausalität zu verwechseln.

### *Datengenerierender Prozess*

Aufbauend auf dem Wissen, welche Fragestellung überhaupt relevant ist, stellt sich als nächstes die Frage, welche Daten konkret genutzt werden können, um die Forschungsfrage zu beantworten. Wie bereits eingangs erwähnt ist die Befragung in vielen Kontexten eine weiterhin sehr wichtige Methode der Datenerhebung in der Markt- und Sozialforschung, wenn auch oft in Kombination mit anderen Datenquellen. Das Wissen darüber, wie eine Befragung durchgeführt werden muss, um reliable und valide Daten zu erheben, ist angesichts der Zunahme an Befragungen heute wichtiger denn je. Mit Stichprobenziehung und Fragebogengestaltung wollen wir hier zwei Elemente des Bewertungsprozesses herausgreifen, für die sich in den letzten Jahren die größten Veränderungen ergeben haben. Im Bereich der Stichprobenziehung kommt es verstärkt zu einer Verschiebung weg von aufwändigen und teuren zufallsgesteuerten Stichproben hin zu kostengünstigeren nicht-zufälligen Stichproben. Grundvoraussetzung für das Arbeiten mit nicht-zufallsgesteuerten, teilweise stark verzerrten Stichproben aus Freiwilligenpanels, River-Samples, sozialen Netzwerken und von digitalen Verhaltensspuren ist das Verständnis darüber, welche Aussagen mit Daten aus diesen Quellen gemacht und welche nicht gemacht werden können und unter welchen Annahmen etablierte und neue Methoden der Gewichtung von verzerrten Stichproben valide Rückschlüsse auf eine vorab definierte Grundgesamtheit zulassen (Versuche finden sich bei Elliot und Valliant [2017]; Wang et al. [2015]).

Neben der Stichprobenziehung spielt die Fragebogengestaltung, also das “Wie” gefragt wird, eine entscheidende Rolle. Wissenschaftlich fundiertes Fragebogendesign sollte zu den Grundkenntnissen all jener zählen, die Befragungsdaten erheben wollen. In den letzten Jahren hat sich mit der zunehmenden Verbreitung von Smartphones in der Bevölkerung - im Jahr 2018 haben laut Eurostat (2019) 79 Prozent der Erwachsenen in Deutschland Internet auf dem Mobiltelefon genutzt - auch der Anteil jener erhöht, die Onlinefragebögen auf ihrem Smartphone ausfüllen. Es ist heute unumgänglich, das Design von Fragebögen an die technischen Gegebenheiten von mobilen Geräten (kleiner Bildschirm, Antworteingabe mittels Touchscreen) und das spezielle Nutzungsverhalten von SmartphoneuserInnen (mobile Nutzung, kurze Zeitspannen) anzupassen (siehe zum Beispiel Couper et al. [2017] und Antoun et al. [2018]).

Pretests ermöglichen es, Probleme der Darstellung auf verschiedenen Gerätetypen rechtzeitig zu erkennen und zu beheben. Hier kann die Marktforschung noch einiges von der User Experience (UX) Forschung lernen (hilfreich ist hier der Einstieg über Geisen und Romano Bergstrom [2017]).

Neben Befragungen spielt die Nutzung von administrativen Daten aus behördlichen oder anderen Verwaltungsaktivitäten (z.B. Einkommensdaten aus Steuererklärungen, Aufnahmedaten von Krankenhäusern, Zulassungsdaten aus Bildungseinrichtungen, Sozialversicherungsmeldungen) im europäischen Raum vor allem in der offiziellen Statistik eine wichtige Rolle. Der originäre Zweck dieser Daten ist die reibungslose Abwicklung von administrativen Prozessen und die Produktion offizieller Statistiken für politische Entscheidungen, die wissenschaftliche Nutzung ist hier nachgelagert. Allerdings stellen immer mehr öffentliche Organisationen administrative Daten für Forschungszwecke zur Verfügung (siehe z.B. das Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung: <https://fdz.iab.de/>). Neben der eigenständigen Analyse dieser Daten für sozialwissenschaftliche Fragestellungen werden administrative Daten auch als Grundlage für die Stichprobenziehung für Befragungen verwendet (Brick 2011) oder mit Befragungsdaten verlinkt, um diese gemeinsam auszuwerten (Calderwood und Lessof 2009) oder Verzerrungen in den Befragungsdaten aufzuzeigen (Kreuter et al. 2010). In der Regel handelt es sich bei administrativen Daten um sehr große Datensätze ( $n = \text{alle}$ ), die sich oft auch in der Struktur und Komplexität von Befragungsdaten unterscheiden. Das Arbeiten mit diesen Daten erfordert daher besondere Kenntnisse aus den Bereichen Datenaufbereitung, -zugang und -management (Conelly et al. 2016) und für die NutzerInnen ist auch hier entscheidend, den datengenerierenden Prozess zu verstehen, um Fehlerquellen identifizieren zu können. Anders als vielleicht angenommen sind auch administrative Daten nicht frei von Fehlern (Groen 2012).

Zusätzlich zu den etablierten Methoden der Datenerhebung sind in letzter Zeit auch andere, vermeintlich kostengünstigere und schneller zugängliche digitale Datenquellen interessant geworden. Um diese Daten für die Forschung nutzbar zu machen, benötigt es meist spezielle technische Fähigkeiten (z.B. Programmierkenntnisse). Grob lässt sich der Zugang zu digitalen Daten in drei Gruppen kategorisieren (Jünger 2018). Erstens können Rohdaten direkt aus bestehenden, meist proprietären Plattformen (z.B. Clickdaten auf einer Website, Call Record Daten von Telefonanbietern, Kreditkartentransaktionsdaten) oder eigens dafür entwickelten

Tools (z.B. Forschungsapps und Browsertracker) generiert werden. Letzteres setzt die technische Umsetzung (z.B. Programmierung einer Forschungsapp), die Rekrutierung von StudienteilnehmerInnen und die datenschutzrechtliche und ethische Abklärung der Datenerhebung voraus. Zweites lassen sich auf Webseiten verfügbaren Informationen via Web Scraping sammeln. Um Web Scraping effizient durchführen zu können, muss man einerseits die hierarchische Struktur von Webseiten und des dahinter liegenden Codes (HTML oder XML) verstehen und andererseits mit Softwareprogrammen arbeiten, um die entsprechenden Informationen automatisch zu extrahieren und aufzubereiten (Munzert et al. 2015). Dies ist bereits mit relativ einfachen Programmierkenntnissen (z.B. in R oder Python) möglich. Zusätzlich ist Wissen über technische Begrenzungen (das Scrapen von Webseiten mit dynamischen Inhalten ist äußerst komplex) und rechtliche Rahmenbedingungen (nicht alle Betreiber erlauben das Scrapen der Inhalte ihrer Webseiten) relevant. Drittens stellen immer mehr Organisationen digitale Daten durch sogenannte Application Programming Interfaces (APIs) zur Verfügung (Neylon 2017). Diese ermöglichen einen relativ einfachen und strukturierten Zugang zu ganz bestimmten Daten. Während viele APIs nicht primär für Forschungszwecke bereitgestellt werden (meist eher zur Kommunikation zwischen verschiedenen Websites und Anwendungen für Marketing- und Werbezwecke) nutzen ForscherInnen regelmäßig z.B. die Twitter API, um Daten aus dem sozialen Netzwerk zu sammeln und analysieren (Jungherr 2016). Um mit Daten aus APIs arbeiten zu können benötigt es einerseits grundlegende Programmierkenntnisse (für R und Python sind Pakete vorhanden, um verschiedene APIs direkt “anzusprechen”) und andererseits ein Verständnis dafür, welche Daten über die API überhaupt zur Verfügung gestellt werden.

Da die oben beschriebenen Datentypen individuelle Stärken und Schwächen aufweisen, besteht zunehmendes Interesse daran die Daten aus den unterschiedlichen Quellen gemeinsam nutzbar zu machen. Während die Verlinkung von Befragungsdaten und Daten aus administrativen Prozessen zumindest im europäischen Raum eine gewisse Tradition haben, wird auch die Verlinkung zwischen Befragungsdaten und Daten aus anderen Quellen in den Wirtschafts- und Sozialwissenschaften immer wichtiger (Schnell 2018; Stier et al. 2019). Das Verlinken von individuellen Daten erfordert Wissen über verschiedene deterministische (mittels eindeutigem Identifier, z.B. Sozialversicherungsnummer, Kundennummer) oder probabilistische Matchingtechniken (mittels mehrere nicht-eindeutiger Identifier, z.B. Name, Geschlecht,

Geburtsdatum) und wie diese technisch implementiert werden können (Tokle und Bender 2016). Auch hier ist die Abschätzung von Fehlern, die beim Verlinken auftreten, und wie diese reduziert werden können essentiell (Sakshaug und Antoni 2017). Besonderes Augenmerk ist auch auf das Thema Datenschutz zu legen, da in vielen Ländern die aktive Zustimmung einer Person erforderlich ist, um ihre Daten aus verschiedenen Quellen verlinken zu dürfen. Egal ob Daten mittels Befragung neu erhoben werden oder auf bestehende administrative oder digitale Daten zugegriffen wird, Grundlage für das Arbeiten mit allen Arten von Daten ist ein Verständnis über den datengenerierenden Prozess. Dazu müssen Nutzer verstehen und einschätzen können, welche Fehlerquellen bei der Entstehung der Daten zu einer Verringerung der Datenqualität führen. Das Total Survey Error (TSE) Framework und seine Varianten (für einen Überblick, siehe Groves und Lyberg 2010) liefert ForscherInnen einen Rahmen, um die Güte von Befragungsdaten anhand von verschiedenen potentiellen Fehlerquellen (Abdeckung der Grundgesamtheit, Stichprobenziehung, Nonresponse, Messung, Datenaufbereitung) abzuschätzen. Ähnliche Qualitätsmodelle wurden auch für neue Datenquellen vorgeschlagen (Japiec et al. 2015).

### *Datenaufbereitung und -speicherung*

Sind die entsprechenden Daten zur Beantwortung einer Forschungsfrage einmal identifiziert bzw. gesammelt, müssen diese vor der Analyse aufbereitet, organisiert und gespeichert werden. Für große und komplexe Datenmengen hat sich, aus der Informatik kommend für diesen Schritt der Terminus Extract, Transform, Load (ETL) eingebürgert. Dieser Aufbereitungsprozess nimmt in der Regel oft deutlich mehr Zeit in Anspruch als die eigentliche Analyse der Daten (oft gehandelte Daumenregel: 80% Datenaufbereitung, 20% Datenanalyse). Je größer und komplexer die Daten sind, umso wichtiger ist es, dass die beteiligten Personen über die entsprechenden technischen Kenntnisse verfügen, um diesen Prozess möglichst effizient zu gestalten. Markt- und SozialforscherInnen sind es traditionell gewohnt mit flachen, rechteckigen Datenfiles (z.B. im CSV-Format) zu arbeiten, in denen in der Regel jede Zeile einem Individuum, Haushalt, einer Organisation oder anderen Erhebungseinheiten entspricht und in den Spalten die Variablen der Messung dieser Einheiten abgebildet sind. Werden die Daten allerdings kontinuierlich gesammelt und ergänzt oder erneuert (z.B. Logfiledaten aus einem Onlineshop) und/oder haben die Daten eine komplexe, hierarchische Struktur (z.B. Posts aus sozialen Netzwerken inkl.



Metadaten über Autor, Verlinkungen, Likes etc.) ist die Nutzung eines Datenbankmanagementsystems (DBMS) empfehlenswert (Foster und Heus 2016). Bei extrem großen Datenmengen, die nicht mehr auf nur einem Rechner verarbeitet werden können, muss zusätzlich ein System implementiert werden, das eine Verteilung der Rechenleistung auf mehrere Server ermöglicht ("Parallel Computing"; Vo und Silver [2016]). In größeren Organisationen werden diese Aufgaben mit Hilfe der IT-Abteilung und den dort arbeitenden SpezialistInnen gelöst (Harms und Schmidt 2019). Hier ist für ForscherInnen, die später die Analysen durchführen werden, nicht nur eine gute Zusammenarbeit mit den Anbietern der Infrastruktur wichtig, sondern es sollten auch SQL-Grundkenntnisse vorhanden sein, um möglichst rasch und gezielt Daten aus einer Datenbank in die Analysesoftware laden zu können.

Selbst wenn die Entscheidung gegen ein dezidiertes DBMS gefallen ist, so müssen die Daten vor der Analyse in der Regel aufbereitet werden (Data Wrangling/Munging). Hierfür ist ein klares Verständnis zur Struktur der Daten notwendig. Kohler (2016) spricht in diesem Zusammenhang von Datability, der Fähigkeit im Umgang mit Daten, als eine der Grundanforderungen für HochschulabsolventInnen in den Sozialwissenschaften. Bei Befragungsdaten handelt es sich dabei um klassische Dateneditierungsaufgaben, wie z.B. Konsistenz- und Bereichschecks (Fellegi und Holdt 1976), die Imputation von fehlenden Werten (Carpenter und Kenward 2012) und die Bildung neuer Variablen und Indexes. Daten aus digitalen Quellen werden häufig in nicht-tabellarischen Datenformaten (z.B. JSON oder XML) extrahiert und beinhalten sogenannte "Tags", die Informationen über die Eigenschaften von Daten liefern. Diese Strukturierung erlaubt eine einfachere Darstellung von hierarchisch angeordneten Daten und spart Speicherplatz. Vor der Analyse müssen die Daten aber zuerst so umstrukturiert werden, damit sie in die Analysesoftware eingelesen und weiterverarbeitet werden können.

Je nach Quelle kommen digitale Daten in einer Vielzahl an unterschiedlichen Formaten vor, wobei besonders häufig Textdaten (z.B. aus Web Scraping oder von Social Media APIs) vorkommen. Das rohe Textmaterial muss zuerst bereinigt (z.B. Zerlegen des Textes in logisch zusammengehörigen Einheiten, sogenannte Tokens, Identifikation von irrelevanten Stop Words, Reduktion auf Wortstämme) und dann in eine numerische Matrix umgewandelt werden (für einen Überblick zu den Arbeitsschritten siehe Klochikhin und Boyd-Graber [2016]). Konkret ist für das Arbeiten mit Text als Daten Wissen über den Umgang mit regulären Ausdrücken zur Beschreibung von komplexen Suchmustern, generelle Kenntnisse der linguistischen

Datenverarbeitung (Natural language processing - NLP) und in weiterer Folge auch Topic Modeling zur Identifikation häufiger gemeinsam vorkommender Themen, Phrasen und Wörter gefragt.

Für die Bearbeitung und Manipulation von jeglicher Form von strukturierten und unstrukturierten Daten sind heutzutage grundlegende Computer- und Programmierkenntnisse unerlässlich. Kohler (2016) spricht hier von "Computability". Objekt-orientierte, Open Source Programme wie R und Python haben in den letzten Jahren vor allem in der universitären Ausbildung und Forschung aber auch den in den angewandten Datenwissenschaften bis dahin weit verbreiteten Datenverarbeitungsprogrammen wie SPSS, Stata und SAS den Rang abgelaufen. Welche Software letztendlich genutzt werden sollte, kann von unterschiedlichen Faktoren abhängen (siehe z.B. Willems [2015]). Am Wichtigsten ist jedoch, dass NutzerInnen so mit der Programmiersprache vertraut sind, dass sie problemlos Daten aus den gewünschten Quellen extrahieren, in die für die Analyse notwendige Form bringen und analysieren können. Dazu gehört unter anderem auch das Schreiben von Funktionen und Loops, um repetitive Arbeitsschritte, wie beispielsweise das Umcodieren von einer größeren Anzahl an Variablen, möglichst effizient und fehlerfrei zu gestalten.

### *Datenanalyse*

Das Arbeiten mit den Daten findet in der Datenanalyse seinen Höhepunkt. Die Datenanalyse selbst ist als iterativer Prozess zu verstehen, bei dem von einer Forschungsfrage ausgehende die Daten zuerst einer explorativen Begutachtung unterzogen werden (Lässt sich die Forschungsfrage mit diesen Daten überhaupt beantworten?) und danach formale Modelle gebildet und überprüft werden, deren Ergebnisse schließlich interpretiert werden müssen (für eine einfache Einführung in diese Prozesse, siehe Peng und Matsui [2015]). Hierfür sind fundierte Kenntnisse über verschiedene statistische Verfahren, den Einsatz von Hypothesentests und Verständnis von Inferenzstatistik sowie statistische Kontrolle von Drittvariablen Voraussetzung (Kohler [2016] nennt diese Fähigkeit "Statability").

Durch die zunehmende Verfügbarkeit großer und unstrukturierter Daten gepaart mit einer gestiegenen Rechenleistung von Computern haben sich Methoden im Bereich des Machine Learnings im Forschungsalltag etabliert. Viele dieser Methoden sind nicht wirklich neu und basieren auf Ideen von Clusteranalysen und logistischer Regression (siehe Ghani und Schierholz

[2016] für eine Einführung in die Methode), konnten aber aufgrund ihrer Komplexität bis vor kurzem nicht mit handelsüblichen Computern durchgeführt werden. Erweiterungen daraus, wie z.B. Random Forests and Support Vector Machines, erfordern, wie andere vertiefende statistische Verfahren auch, spezielles Training. Doch nicht jedes Markt- oder Sozialforschungsprojekt verlangt nach dem Einsatz dieser Methoden, die vorrangig für Vorhersage und Klassifikation eingesetzt werden, und nicht alle ForscherInnen müssen Machine Learning Experten werden. Ein Grundverständnis darüber, was unter Machine Learning zu verstehen ist, wodurch sich supervised und unsupervised Methoden unterscheiden und unter welchen Voraussetzungen diese sinnvoll eingesetzt werden können, schadet aber nicht, wenn man gemeinsam mit Statistikexperten an Projekten arbeitet.

Die Zunahme an Datenquellen, auch kleinräumiger Natur, hat zu einer gesteigerten Popularität von Methoden zur Small Area Estimation geführt (Marchetti et al. 2015). Ziel ist hier die Produktion reliabler sozio- und gesundheitsökonomischer Statistiken auf regionalem und subregionalem Niveau, die alleine durch Befragungen nur sehr schwer und unter extrem hohem Kosteneinsatz möglich wäre. Generell nimmt die Bedeutung von Positionsdaten und mit anderen geographischen Informationen angereicherten Daten zu (siehe z.B. Rieder und Kühne [2018] für das Arbeiten mit geo-getaggten Postings auf Twitter).

### *Datenoutput und -zugang*

Sind die Daten analysiert, müssen die Ergebnisse für Entscheidungsträger aufbereitet und für NachnutzerInnen zur Verfügung gestellt werden. Hierbei sind vor allem drei Aspekte relevant. Erstens hat die Bedeutung an guter Visualisierung, die es den BetrachterInnen ermöglicht unter möglichst geringem kognitiven und analytischen Aufwand Zusammenhänge zu erkennen und zu verstehen und darauf aufbauend datengestützte Entscheidungen zu treffen, durch die schiere Größe der Daten zugenommen (Yalcin und Plaisant 2016). Neben dem inhaltlichen Wissen über relevante Zusammenhänge und den schon zuvor erwähnten statistischen Kenntnissen kommen hier künstlerisch-ästhetische Skills hinzu. Grundprinzipien von analytischem Design und das Wissen, welche Visualisierungstechniken am besten für die Präsentation von uni- und multivariaten Daten eingesetzt werden sollten, sind besonders gefragt (Few 2012). Auch hier gilt, dass nicht so sehr die verwendeten Tools (z.B. ggplot2 in R, Tableau oder Shiny) im

Vordergrund stehen sollten sondern die generelle Fähigkeit, Ergebnisse klar und für die relevante Zielgruppe verständlich darzustellen (Nussbaumer Knafllic 2015).

Neben der Aufbereitung der Ergebnisse spielt auch die nachvollziehbare Dokumentation aller Arbeitsschritte in einem Datenprojekt eine immer wichtigere Rolle. Einerseits hat die sogenannte Reproduktionskrise in den Natur- und Sozialwissenschaften klar aufgezeigt, wo die Schwächen in der Nachvollziehbarkeit und Replizierbarkeit wissenschaftlicher Ergebnisse liegen (Pashler und Wagenmakers 2012). Andererseits hat die Komplexität vieler Projekte zugenommen, so dass in den seltensten Fällen heute alleine im stillen Kämmerchen gearbeitet wird. Sauber annotierter Code erlaubt es auch anderen ForscherInnen das bisher Getane nachzuvollziehen und darauf aufzubauen und ist daher Grundvoraussetzung für Kollaboration (Peng 2015). Tools wie git und GitHub erleichtern die Versionsverwaltung von Code, verringern die Fehleranfälligkeit und erhöhen die Transparenz (Perkel 2018). Das Teilen von Code und, falls rechtlich und ethisch möglich, von Daten ermöglicht es der wissenschaftlichen Community voneinander zu lernen und zum weiteren wissenschaftlichen Fortschritt beizutragen.

Eng mit der Datenweitergabe ist natürlich das Thema Datenschutz und der ethische Umgang mit den Daten verbunden. Ein Grundverständnis über die 2018 in Kraft getretene EU-Datenschutzgrundverordnung (EU-DSGVO) zur Sammlung, Nutzung, Speicherung und Weitergabe personenbezogener Daten ist heute in vielen datenintensiven Projekten notwendig. Dazu zählt das Wissen, unter welchen Rahmenbedingungen die EU-DSGVO greift und wer im Rahmen eines Forschungsprojektes für die Einhaltung dieser Regeln zuständig ist. Wichtig ist hier zu verstehen, dass die herkömmlich als “personally identifiable information” bezeichneten Variablen (Name, Adresse, Geburtsdatum) oft nicht die einzigen Informationen sind, die dazu führen können, dass eine Person (oder ein Unternehmen) im Datensatz identifiziert werden kann. Gerade in der Zusammenführung vieler umfangreicher Datensätze kann es zu uneindeutigen Merkmalskombinationen kommen. Der Sammelband von Lane et al. (2014) zu Big Data, Privacy and the Public Good gibt einen Überblick über viele der wichtigsten Konzepte und Diskussionen.

### *Müssen Markt- und SozialforscherInnen das wirklich alles können?*

Angesichts der doch recht langen Liste an Kompetenzen, die wir hier präsentieren, werden sich einige LeserInnen bestimmt fragen, ob sie wirklich alle das können müssen, um in Zukunft auch noch erfolgreich in der Markt- und Sozialforschung arbeiten zu können. Doch ist es überhaupt

realistisch, dass es solche Generalisten mit ausgewiesener methodischer Expertise in so diversen Bereichen, wie Fragebogendesign, Stichprobenziehung, Web Scraping, Data Munging, Cloud Computing, Machine Learning, Datenvisualisierung und Datenschutz gibt, die zusätzlich tiefgehendes betriebs- und sozialwissenschaftliches Fachwissen mitbringen? Die Antwort ist natürlich “Nein” und, in den meisten Fällen, ist das auch gar nicht notwendig. Projekte werden heute großteils in Teams abgewickelt in denen Experten aus den einzelnen hier diskutierten Teilgebieten ihr Know-how einbringen. Umso wichtiger ist es aber, dass die Markt- und SozialforscherInnen der Zukunft sich mit den Fachexperten austauschen können.

Als Datenmethodikerinnen bringen die meisten Markt- und SozialforscherInnen jetzt schon die Fähigkeit mit, die richtigen inhaltlichen Fragen zu stellen und entsprechende Studiendesigns zu entwickeln. Sie sind die ExpertInnen für den datengenerierenden Prozess und die Evaluation der Qualität der erhobenen Daten. Dieses Wissen lässt sich nicht einfach durch andere technisches Fähigkeiten ersetzen. Gepaart mit soliden Programmierkenntnissen und fundiertem Statistikwissen, ohne sich gleich zu InformatikerInnen und StatistikerInnen ausbilden lassen zu müssen, und der Bereitschaft, sich aktiv mit den neuen Entwicklungen rund um das Thema digitale Daten zu beschäftigen, ergibt das eine Kombination, die auch am Arbeitsmarkt der Zukunft sehr gefragt sein wird. Dies sollte all jenen Hoffnung geben, deren Studienabschluss schon mehrere Jahre (teilweise sogar Jahrzehnte) zurückliegt, die jetzt aber die Sorge haben, dass sie mit den derzeit auf den Arbeitsmarkt drängenden, technisch teilweise besser ausgebildeten Nachwuchskräften nicht mithalten können.

### **Trends in der Aus- und Weiterbildung von Sozial- und MarktforscherInnen**

Bis vor kurzem haben sich Personen, die im deutschsprachigen Raum im Bereich der empirischen Markt- und Sozialforschung arbeiten, meist aus AbsolventInnen von traditionellen sozialwissenschaftlichen (z.B. Soziologie, Psychologie, Politikwissenschaften), kommunikationswissenschaftlichen oder betriebswirtschaftlichen Studiengänge an Universitäten und Fachhochschulen rekrutiert. Seltener kamen AbsolventInnen auch aus der Statistik. Die bereits weiter oben beschriebenen Veränderungen an die Anforderungen an Sozial- und MarktforscherInnen gehen mit einem Um- und Aufbruch in der Ausbildung einher. In diesem Abschnitt beschreiben wir sowohl inhaltliche Trends in der Veränderung der Studieninhalte als

auch strukturelle Änderungen, was die Anbieter von Aus- und Weiterbildungsangeboten und die Formate der Darbietung von Lerninhalten betrifft.

### *Aktuelle Inhaltliche Ausrichtung in der Aus- und Weiterbildung*

In der akademischen Ausbildung in den Sozial- und Wirtschaftswissenschaften wird zwar, je nach Standort, mehr oder weniger viel Wert auf eine solide quantitative Grundausbildung der Studierenden gelegt, der Fokus lag und liegt hier aber meist immer noch auf dem Arbeiten mit Daten aus traditionellen Erhebungsmethoden wie Befragungen und Beobachtung. Das Arbeiten mit großen semi- und unstrukturierten Daten aus neuen digitalen Quellen findet sich hier meist noch nicht im Curriculum. Auf der anderen Seite sind in den letzten Jahren, dem internationalen Trend folgend, neue Data Science Studiengänge, sowohl auf Bachelor- als auch auf Masterebene, entstanden. Zum Zeitpunkt des Schreibens dieses Beitrags konnten wir über 50 solcher Studienprogramme an deutschen, österreichischen und schweizer Universitäten und Fachhochschulen identifizieren (Tabelle 1). Der Großteil dieser Programme sind an Fakultäten der Informatik, Mathematik oder Statistik angesiedelt. Die Curricula dieser Programme sind daher meist sehr technisch ausgerichtet, mit einem starken Fokus z.B. auf Datenmanagement, IT-Infrastruktur und Algorithmenbildung, und haben wenig Bezug zu den Sozialwissenschaften oder der Marktforschung. Zusätzlich gibt es immer mehr Studiengänge im Bereich des Business Analytics, meist mit einer klaren Fokussierung auf die Gewinnung und Analyse von betriebswirtschaftlichen Prozessdaten.

**Tabelle 1. Data Science Studienprogramme in Deutschland, Österreich und der Schweiz**

<b>Programmname</b>	<b>Standort</b>	<b>Fakultät/Fachbereich</b>	<b>Abschluss</b>	<b>Format</b>
Angewandte Data Science	Universität Göttingen	Mathematik und Informatik	BSc	Präsenz
Applied Data Analytics	HWZ (Zürich)	Informatik	CAS	Online & Präsenz
Applied Data Science	Universität Bern	Philosophie & Naturwissenschaften	CAS	Präsenz
Applied Information and Data Science	Hochschule Luzern	Wirtschaft	MSc	Präsenz
Big Data & Business Analytics	FOM Hochschule (mehrere Standorte)	-	MSc	Präsenz, berufsbegleitend

<b>Programmname</b>	<b>Standort</b>	<b>Fakultät/Fachbereich</b>	<b>Abschluss</b>	<b>Format</b>
Big Data and Business Analytics	SRH Heidelberg	Information, Medien und Design	MSc	Präsenz, Voll- und Teilzeit
Big Data Management	AKAD University	Digitalisierung und Innovation	BA	Online
Big Data Management	IUBH (Bad Honnef, Berlin)	International Business	MA/MBA	Präsenz
Business Administration (Vertiefung: Information and Data Management)	HTW Chur	-	MSc	Präsenz, berufsbegleitend
Business Intelligence & Analytics	TU Chemnitz	Wirtschaftswissenschaften	MSc	Präsenz
Business Intelligence & Data Science	ISM (Dortmund und München)	-	MSc	Präsenz
Computational and Data Science	Friedrich-Schiller-Universität Jena	Mathematik und Informatik	MSc	Präsenz
Computer Science (Track: Data Analytics)	TU Berlin	Elektrotechnik und Informatik	MSc	Präsenz
Data Analytics	Universität Hildesheim	Mathematik, Naturwissenschaften, Wirtschaft und Informatik	MSc	Präsenz
Data and Information Science	TH Köln	Informations- und Kommunikationswissenschaften	BSc	Präsenz
Data Engineering	Universität Potsdam	Digital Engineering	MSc	Präsenz
Data Engineering and Analytics	TU München	Informatik	MSc	Präsenz
Data Science	EPFL (Lausanne)	Computer and Communication Sciences	MSc	Präsenz
Data Science	Fernfachhochschule Schweiz (Zürich oder Bern)	Informatik	DAS	Online
Data Science	BFH (Bern)	Technik und Informatik	MAS	Präsenz
Data Science	FHNW (Brugg-Windisch)	Technik	BSc	Präsenz
Data Science	ETH Zürich	Informatik	MSc	Präsenz
Data Science	Hochschule Aalen	Elektronik und Informatik	BSc	Präsenz

<b>Programmname</b>	<b>Standort</b>	<b>Fakultät/Fachbereich</b>	<b>Abschluss</b>	<b>Format</b>
Data Science	Hochschule Albstadt-Sigmaringen	Informatik	MSc	Online mit Präsenzphasen, berufsbegleitend
Data Science	Hochschule Darmstadt	Mathematik und Naturwissenschaften	MSc	Präsenz
Data Science	LMU München	Mathematik, Informatik und Statistik	MSc	Präsenz
Data Science	Philipps-Universität Marburg	Mathematik und Informatik	BSc	Präsenz
Data Science	Philipps-Universität Marburg	Mathematik und Informatik	MSc	Präsenz
Data Science	RWTH Aachen	Mathematik, Informatik und Naturwissenschaften	MSc	Präsenz
Data Science	TU Chemnitz	Mathematik	MSc	Präsenz
Data Science	TU Wien	Informatik	MSc	Präsenz
Data Science	Universität Bielefeld	Wirtschaftswissenschaften	MSc	Präsenz
Data Science	Universität Potsdam	Mathematik & Naturwissenschaften	MSc	Präsenz
Data Science	Universität Salzburg	Naturwissenschaften	MSc	Präsenz
Data Science	Universität Stuttgart	Informatik	BSc	Präsenz
Data Science	ZHAW (Winterthur)	Engineering	MAS	Präsenz, berufsbegleitend
Data Science and Business Analytics	Hochschule Aalen	Wirtschaftswissenschaften	MSc	Online & Präsenz, berufsbegleitend
Data Science and Business Analytics	FH St. Pölten	Informatik & Security	BSc	Präsenz, berufsbegleitend
Data Science and Business Analytics	Hochschule der Medien Stuttgart	-	MSc	Online & Präsenz, berufsbegleitend
Data Science in Business and Economics	Eberhard Karls Universität Tübingen	Wirtschafts- und Sozialwissenschaften	MSc	Präsenz
Data Science & Intelligent Analytics	FH Kufstein Tirol	Wirtschaft und Technik	MSc	Präsenz, berufsbegleitend
Data Science in der Medizin	Technische Hochschule Ulm	Informatik	BSc	Präsenz
Datenanalyse und Date-management	TU Dortmund	Statistik	BSc	Präsenz



<b>Programmname</b>	<b>Standort</b>	<b>Fakultät/Fachbereich</b>	<b>Abschluss</b>	<b>Format</b>
Datenwissenschaft	TU Dortmund	Statistik	MSc	Präsenz
Digital Business & Data Science	University of Applied Sciences Europe (Berlin, Hamburg)	-	BSc	Präsenz
Informatik (Nebenfach: Data Science)	Universität Zürich	Wirtschaftswissenschaften	MSc	Präsenz
Informatik (Schwerpunkt: Big Data)	Universität Leipzig	Mathematik und Informatik	MSc	Präsenz
Informatik (Schwerpunkt: Data Science)	Fachhochschule Wiener Neustadt (auch am Standort Tulln)	Technik	MSc	Präsenz, berufsbegleitend
Informatik (Schwerpunkt: Data Science)	Universität Wien	Informatik	MSc	Präsenz
International Program in Survey and Data Science	Universität Mannheim & University of Maryland	Sozialwissenschaften	MSc	Online, berufsbegleitend
Machine Learning	Eberhard Karls Universität Tübingen	Mathematik & Naturwissenschaften	MSc	Präsenz
Machine Learning and Data Analytics	Hochschule Aalen	Elektronik und Informatik	MSc	Präsenz
Mannheim Master in Data Science	Universität Mannheim	Wirtschaftsinformatik & Wirtschaftsmathematik in Kooperation mit Sozialwissenschaften	MSc	Präsenz
Management & Data Science	Leuphana Universität Lüneburg	Wirtschaftswissenschaften	MSc	Präsenz
Mathematics in Data Science	TU München	Mathematik	MSc	Präsenz
Mathematical Data Science	Universität Göttingen	Mathematik und Informatik	BSc	Präsenz
Project Management and Data Science	HTW Berlin	Wirtschafts- und Rechtswissenschaften	MSc	Präsenz
Social and Economic Data Science	Universität Konstanz	Wirtschaftswissenschaften	MSc	Präsenz

Nur wenige Studiengänge verknüpfen in ihrem Curriculum technische Komponenten der Statistik und des Programmierens mit einem Fokus auf Datenerhebung und Datenqualität unter Berücksichtigung verschiedenster Datenquellen und deren Kombination, mit Anwendungsfällen, die für marktforschungs- und sozialwissenschaftliche Fragestellungen relevant sind. Wie bereits weiter oben in diesem Beitrag besprochen sehen wir die zentrale Wichtigkeit im Verstehen des datengenerierenden Prozess, insbesondere dann, wenn es sich um Daten von und über Menschen handelt. Im Vordergrund sollten aus unserer Sicht dabei das Vermitteln generalisierbarer Konzepte stehen, die unabhängig von der thematischen Fragestellung angewandt werden können, und weniger das Lernen z.B. einer bestimmten Programmiersprache.

### *Traditionelle und neue Anbieter von Aus- und Weiterbildung*

Traditionell war die Ausbildung im Bereich der Markt- und Sozialforschung im deutschsprachigen Raum eine Domäne etablierter Universitäten und Fachhochschulen und wie Tabelle 1 zeigt, haben diese auch auf die zunehmende Nachfrage nach Expertise in den Bereichen Data Science und Big Data reagiert. Neu hinzu kamen in den letzten Jahren allerdings nicht-universitäre Anbieter, die vor allem online Aus- und Weiterbildung in diesem Bereich anbieten. Online Lernplattformen wie Coursera (<https://www.coursera.org/>), edX (<https://www.edx.org/>), Khan Academy (<https://www.khanacademy.org/>) oder Udacity (<https://eu.udacity.com/>) bieten ein breites Spektrum an Kursen aus verschiedenen Fachrichtungen an, unter anderem auch Spezialisierungen zum Arbeiten mit Daten aus neuen aber auch traditionellen Quellen. DataCamp (<https://www.datacamp.com/>) ist eine weitere Online Lernplattform auf der man interaktiv Programmieren mit R, Python und SQL lernen kann. Sage Campus (<https://campus.sagepub.com/>) versucht mit seinem Kursangebot gezielt SozialwissenschaftlerInnen anzusprechen, die sich im Bereich Data Science aus- und weiterbilden wollen.

Die neuen Anbieter unterscheiden sich in ihrem Angebot von der traditionellen Ausbildung an Universitäten und Fachhochschulen in mehreren Punkten. Erstens bieten sie ihre Kurse ausschließlich online an und erreichen so ein breites internationales Publikum. Zweitens bieten einige Lernplattformen ihre Inhalte als sogenannte Massive Open Online Courses (MOOCs) an, bei denen die Studierenden ohne Studiengebühren an Vorlesungen aller Fachrichtungen teilnehmen können. Direkte Interaktion mit den Vortragenden ist dabei in der Regel nicht

vorgesehen und die Auseinandersetzung mit den Lerninhalten beschränkt sich meist auf das Ansehen von voraufgezeichneten Videos und das Lösen standardisierter Aufgaben. Dieses selbstgerichtete Lernen erfordert hohe Motivation auf Seiten der Studierenden und so kämpfen viele MOOCs mit extrem hohen Abbruchraten (siehe dazu auch Somailova et al. [2018] und Reich und Ruiperez-Valiente [2019]). Neben kostenlosen Formaten bieten viele Online Lernplattformen aber mittlerweile auch kostenpflichtige Versionen ihrer Kurse an, bei denen die TeilnehmerInnen dann auch individuell betreut werden und die Kurse mit Zertifikaten abschliessen. Häufig beinhalten solche Zertifikatskurse dann auch eine Projektkomponente, wobei die Studierenden meist anhand eines kleinen Datenprojekts das Gelernte praktisch umsetzen. Viele der neuen Anbieter haben Kooperationen mit großen internationalen Unternehmen und Organisationen und die Studierenden können in den Kursen mit echten Daten arbeiten. Darüber hinaus unterscheiden sich die neuen Online Lernplattformen von etablierten Studienangeboten in der Modularisierung und Flexibilisierung des Angebots. TeilnehmerInnen können auf der einen Seite einzelne Kurse je nach Interesse und Bedarf flexibel belegen und auf der anderen Seite thematisch zusammenhängende Kurssequenzen zu einem Schwerpunktthema belegen.

Die neuen Online Lernplattformen bieten Markt- und SozialforscherInnen die Möglichkeit sich ein den jeweiligen Ansprüchen entsprechend angepasstes Portfolio an Zusatzfähigkeiten, vor allem im Bereich der technischen Aspekte des Arbeitens mit neuen und großen Daten (z.B. Programmierung), anzueignen. Erst langsam erkennen die etablierten tertiären Ausbildungseinrichtungen im deutschsprachigen Raum, dass sie hier im Bereich der Weiterbildung und des lebenslangen Lernens Kompetenzen haben und selbst auch entsprechende Angebote schaffen müssen (Ehlers und Kellermann 2019; Wissenschaftsrat 2019).

### *Online Lehre, Flipped Classroom und Lernen in kleinen Gruppen*

Wie bereits erwähnt bieten sich MOOCs zum Erlernen erster technischer Fähigkeiten an, die keine Interaktion zwischen den Vortragenden und den Studierenden erfordert. Als Alternative zu MOOCs und traditionellen Präsenzveranstaltungen bietet sich das sogenannte „Flipped Classroom Model“ (auch „Inverted Instruction“ oder „Inverted Classroom“) an, das ein flexibles, selbstgesteuertes Lernen unterstützt und angeleitet durch Vortragende ermöglicht. In „Flipped Classrooms“ wird das Vorlesungsmaterial auf Video aufgezeichnet oder auf anderem Wege

jederzeit online abrufbar zur Verfügung gestellt. Studierende arbeiten mit den so zur Verfügung gestellten Materialien zeit- und ortsunabhängig. Die regelmässige Kontaktzeit mit den Lehrenden wird dazu genutzt, direkt mit den Studierenden in kleineren Gruppen zu interagieren, Fragen zu beantworten, Fallbeispiele zu diskutieren und das Gelernte anhand von Beispielen zu vertiefen. Screen-capture-software, Desktop-sharing und virtuelle Meeting Rooms bilden das technische Fundament mit dem diese neuen Unterrichtsformen effektiv eingesetzt werden können. Während in traditionellen Vorlesungen Lehrende in der Regel nur mit denjenigen Studierenden interagieren, die Fragen stellen, verfolgen Lehrende in „Flipped Classrooms“ den Fortschritt und das Verständnis jedes einzelnen Studierenden in jedem Lernabschnitt. Damit sitzen alle Studierenden im virtuellen (Video-)Klassenzimmer in der ersten Reihe.

Online abgewickelte Programme haben weiters den Vorteil, dass diese nicht an die üblichen akademischen Kalenderzeiten gebunden sind und Kurse über das gesamte Jahr verteilt besucht werden können. Diese Flexibilisierung des Lernens kommt vor allem nicht-traditionellen Studierenden entgegen, die berufsbegleitend studieren wollen oder familiäre Verpflichtungen mit dem Studium vereinbaren müssen.

### **Zusammenfassung**

Im vorliegenden Beitrag haben wir die aktuellen Entwicklungen im Bereich der Aus- und Weiterbildung für Markt- und SozialforscherInnen diskutiert. Das zunehmende Interesse von Unternehmen und öffentlichen Einrichtungen, Daten aus neuen, digitalen Quellen zu nutzen und diese mit Daten aus Befragungen und administrativen Prozessen zu verknüpfen, erfordert Kenntnisse, die bisher nur teilweise von den traditionellen Methodenkompetenzen abgedeckt werden. Wir haben fünf Bereiche identifiziert, in denen Markt- und SozialforscherInnen Kompetenzen aufweisen müssen, um erfolgreich mit Daten aus verschiedenen Quellen zu arbeiten. Zunächst ist es essentiell, einschätzen zu können, wie Forschungsfragen adäquat formuliert werden müssen und welche Daten Antworten auf die entsprechenden Fragen liefern können. Hier ist Fachwissen aus den entsprechenden Disziplinen, die für Markt- und Sozialforschung relevant sind unerlässlich. Wissen über die datengenerierenden Prozesse und die damit verbundenen Fehlerquellen konstituieren den zweiten Kompetenzbereich. Ganz entscheidend ist der sichere und effiziente Umgang mit den Daten selbst, in Form von

Datenaufbereitung, Datenorganisation und Strukturierung von Daten in Datenbanken. Ebenso wichtig ist der Einsatz von Analysetechniken, die auf die jeweilige Fragestellung und das vorhandene Datenmaterial abgestimmt ist. Datenvisualisierung, Datenweitergabe und, ganz eng damit verbunden, Datenschutz stellen die abschließende Komponente dar. Markt- und SozialforscherInnen müssen nicht in all diesen Bereichen zu ausgewiesenen ExpertInnen ausgebildet oder umgeschult werden, allerdings ist ein solides Grundwissen zu den inhaltlichen, methodischen und technischen Aspekten des Arbeitens mit Daten heute die Grundvoraussetzung, um auch in der Zukunft noch am Arbeitsmarkt mithalten zu können.

Für die Aus- und Weiterbildung gibt es mittlerweile ein vielfältiges Angebot von Universitäten und Fachhochschulen sowie neuen Anbietern aus dem Bereich des Online Lernens. Die Bandbreite an Formaten reicht dabei von einzelnen, kostenlosen Onlinekursen, die zusammen mit Tausenden anderen Studierenden aus der ganzen Welt besucht werden können (MOOCs), bis hin zu traditionellen akademischen Präsenzprogrammen, die mit einem Bachelor- oder Mastertitel abschliessen. Dazwischen ist aus unserer Sicht aber auch Platz für Hybridmodelle, die die Flexibilität des Online Lernens mit intensiver Interaktion mit den Vortragenden verbindet.

Mit dem International Program in Survey and Data Science (IPSDS) haben wir an der Universität Mannheim in Kooperation mit dem Joint Program in Survey Methodology an der University of Maryland ein solches Programm geschaffen (siehe auch <https://survey-data-science.net/>). Inhaltlich richtet sich das Programm an den fünf von uns weiter oben identifizierten Kompetenzbereichen aus. Im Programm wird technisches Wissen um das Arbeiten mit neuen Daten mit einem Fokus auf Datenerhebung und Datenqualität unter Berücksichtigung verschiedenster Datenquellen und deren Kombination vereinigt. Zielgruppe des dieses Masterprogramms sind weltweit Personen, die beruflich mit Datenerhebung und Datenanalyse in den unterschiedlichsten Bereichen, wie z.B. angewandter demografischer und ökonomischer Forschung befasst sind oder einen Wiedereinstieg in diesen Bereich suchen. Dazu zählen Mitarbeitende in Statistischen Ämtern, internationalen Datenerhebungsorganisationen und Marktforschungsabteilungen großer Firmen. Fünf Elemente zeichnen das Programm aus (Kreuter et al. 2018):

- Kursmaterialien können von Studierenden asynchron, zu jeder Zeit und von jedem Standort aus per Video genutzt werden.

- Zusätzlich zu den asynchronen Vorlesungen erlauben Diskussionsforen und kleine, virtuelle Klassenräume (etwa 15 Studierende) gemeinsam bestimmte Inhalte mit Dozenten/innen und anderen Studierenden zu erarbeiten.
- Das Programm ist modular und erlaubt damit eine Schwerpunktsetzung, die das im Beruf bereits Erlernte optimal ergänzen kann.
- Das Lehrangebot wird in englischer Sprache präsentiert.
- Das Programm ist international ausgerichtet und erlaubt den Studierenden den Zugang zu einem weiten Netzwerk von Peers und die Bildung von akademischen und privatwirtschaftlichen Kontakten weltweit.

Die Veränderungen in diesem Feld sind schnell und gewaltig. Weiterbildung wird deshalb zunehmend wichtiger, ganz gleich ob diese durch intensives Selbststudium von Artikeln passiert oder durch organisierte Programme. Manche Aspekte ändern sich allerdings nicht so schnell wie die eingesetzten technischen Hilfsmittel: Das Grundverständnis dafür, welche Daten sich zur Beantwortung welcher Fragestellung eignen und vor allem welche nicht.

### **Literaturverzeichnis**

- ADM. (2019). Die Marktforschung in Zahlen. Abgerufen von <https://www.adm-ev.de/die-branche/mafo-zahlen/> am 3.6.2019.
- Antoun, C., Katz, J., Argueta, J. & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, 36, 557-574.
- Beyer, M. A., & Laney, D. (2012). The importance of “Big Data”: A definition. Stamford, CT: Gartner.
- Brick, J.M. (2011). The future of survey sampling. *Public Opinion Quarterly*, 75, 872-888.
- Calderwood, L., & Lessof, C. (2009). Enhancing longitudinal surveys by linking to administrative data. In P. Lynn (Hrsg.), *Methodology of longitudinal surveys*. (S. 55-72). Chichester, UK: Wiley.
- Callegaro, M. (2018). How can web surveys benefit from design and user experience research? Keynote präsentiert auf der GOR18, Köln.

- Callegaro M., & Yang, Y. (2018). The role of surveys in the era of “Big Data”. In D. Vannette & J. Krosnick (Hrsg.), *The Palgrave handbook of survey research*. (S. 175-192). Cham: Palgrave Macmillan.
- Capgemini. (2017). Studie IT-Trends 2017. Überfordert Digitalisierung etablierte Unternehmensstrukturen? Abgerufen von <https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/02/it-trends-studie-2017.pdf> am 7.6.2019.
- Carpenter, J., & Kenward, M. (2012). *Multiple imputation and its application*. New York, NY: Wiley.
- Civey. (2019). Erfahren Sie, was Deutschland denkt! Abgerufen von <https://civey.com/> am 3.6.2019.
- Connelly, R., Playford, C.J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12.
- Couper, M.P., Antoun, C., & Mavletova, A.. (2017). Mobile web surveys: A total survey error perspective. In P.P. Biemer et al. (Hrsg.), *Total survey error in practice*. (S. 133-154). Hoboken, NJ: Wiley.
- Elliott, M.R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 32, 249-264.
- Ehlers, U.-D., & Kellermann, S.A. (2019). *Future skills – The future of learning and higher education. Results of the International Future Skills Delphi Survey*. Karlsruhe.
- Eurostat. (2019). *Digital economy and society: ICT usage in households and by individuals*. Abgerufen von <http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database> am 4.6.2019.
- Fellegi, I.P., & Holt, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, 71, 17-35.
- Few, S. (2012). *Show me the numbers. Designing tables and graphs to enlighten*. Analytics Press.
- Foster, I., & Heus, P. (2016). Databases. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches*. (S. 93-124). Boca Raton, FL: Chapman and Hall/CRC Press.
- Geisen, E., & Romano Bergstrom, J. (2017). *Usability testing for survey research*. Morgan Kaufmann, 2017.

- Ghani, R., & Schierholz, M. (2016). Machine learning. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches*. (S. 147-186). Boca Raton, FL: Chapman and Hall/CRC Press.
- Groen, J.A. (2012). Sources of error in survey and administrative data: The importance of reporting procedures. *Journal of Official Statistics*, 28, 173-198.
- Groves, R.M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74, 849-879.
- Harms, C. & Schmidt, S. (1. April 2019). Marktforschung im Daten-Zeitalter: Software ist nur die halbe Miete. Abgerufen von <https://www.marktforschung.de/dossiers/themendossiers/plattformen-und-datensysteme-2019/dossier/marktforschung-im-daten-zeitalter-software-ist-nur-die-halbe-miete/> am 7.6.2019
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C. & Usher, A. (2015). Big Data in survey research. AAPOR Task Force Report. *Public Opinion Quarterly*, 79: 839–880.
- Jünger, J. (2018). Mapping the field of automated data collection on the web: Collection approaches, data types, and research logic. In C.M. Stutzer, M. Welker, & M. Egger (Hrsg.), *Computational social science in the age of Big Data. Concepts, methodologies, tools, and application*. (S. 104-130). Köln: Herbert van Halem.
- Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13, 72-91.
- Klochikhin, E., & Boyd-Graber, J. (2016). Text analysis. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches*. (S. 187-214). Boca Raton, FL: Chapman and Hall/CRC Press.
- Kohler, U. (2016). Anforderungen an Hochschulabsolventen. In C. König, M. Stahl, & E. Wiegand (Hrsg.), *Human resources*. (S. 43-77). Wiesbaden: Springer.
- Kreuter, F., Keusch, F., Somailova, E. & Frößinger, K. (2018). International Program in Survey and Data Science. In C. König, J. Schröder, E. Wiegand (Hrsg.), *Big Data – Chancen, Risiken, Entwicklungstendenzen* (S. 27-41). Wiesbaden: Springer VS.



- Kreuter, F., Müller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: Making use of administrative data. *Public Opinion Quarterly*, 74, 880-906.
- Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). *Privacy, Big Data, and the public good. Frameworks for engagement.* Cambridge University Press.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., et al. (2009). Computational social science. *Science* 323(5915), 721-723.
- Leek, J.T., & Peng, R.D. (2015). What is the question?. *Science*, 347(6228), 1314-1315.
- Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., ... & Gabrielli, L. (2015). Small area model-based estimators using big data sources. *Journal of Official Statistics*, 31, 263-281.
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining.* Wiley.
- Neylon, C. (2016). Working with web data and APIs. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches.* (S. 23-70). Boca Raton, FL: Chapman and Hall/CRC Press.
- Nussbaumer Knaflic, C. (2015). *Storytelling with data.* Wiley.
- Oppeln-Bronikowski, S.v. (2018 Begrüßung durch die Direktorin beim Statistischen Bundesamt. In C. König, J. Schröder, E. Wiegand (Hrsg.), *Big Data – Chancen, Risiken, Entwicklungstendenzen* (S. 9-12). Wiesbaden: Springer VS.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530.
- Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, 12.3, 30-32.
- Peng, R.D., & Matsui, E. (2015). *The art of data science. A guide for anyone who works with data.* leanpub.
- Penke, M. (2017, 26. Januar). Marktforschungs-Startup Dalia bekommt sieben Millionen Dollar. Abgerufen von <https://www.gruenderszene.de/allgemein/marktforschung-dalia-research-investment> am 3.6.2019.
- Perkel, J.M. (2018). A toolkit for data transparency takes shape. *Nature*, 560, 513-515.

- Reich, J., & Ruipérez-Valiente, J.A. (2019). The MOOC pivot. *Science*, 363(6423), 130-131.
- Rieder, Y., & Kühne, S. (2018). Geo-spatial analysis of social media data –A practical framework and applications using Twitter. In C.M. Stutzer, M. Welker, & M. Egger (Hrsg.), *Computational social science in the age of Big Data. Concepts, methodologies, tools, and application*. (S. 418-441). Köln: Herbert van Halem.
- Samoilova, E., Keusch, F., & Kreuter, F. (2018). Integrating survey and learning analytics data for a better understanding of engagement in MOOCs. In H. Jiao, R.W. Lissitz, & A. Van Wie (Hrsg.), *Data analytics and psychometrics: Informing assessment practices*. (S. 247-261). Charlotte, NC: Information Age Publishing.
- Sakshaug, J., & Antoni, M. (2017). Errors in linking survey and administrative data. In P.P. Biemer et al. (Hrsg.), *Total survey error in practice*. (S. 557-573). Hoboken, NJ: Wiley.
- SAP News. (2018, 11. November). SAP SE to acquire Qualtrics International Inc., Sees experience management as the future of business [Press release]. Abgerufen von <https://news.sap.com/2018/11/sap-to-acquire-qualtrics-experience-management/> am 7.6.2019.
- Schnell, R. (2018). 'Big Data' aus wissenschaftssoziologischer Sicht: Warum es kaum sozialwissenschaftliche Studien ohne Befragungen gibt. German Record Linkage Center Working Paper Series, No. WP-GRLC-2018-01.
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*. Online first publiziert am 24. April 2019 auf <https://doi.org/10.1177/0894439319843669>.
- Stützer, C. M., Welker, M., & Egger M. (2018). *Computational Social Science in the Age of Big Data*. Köln: Herbert von Harlem.
- Tokle, J., & Bender, S. (2016). Record linkage. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches*. (S. 71-92). Boca Raton, FL: Chapman and Hall/CRC Press.
- Usher, A. (2015). Skills required to integrate Big Data into public opinion research." Paper präsentiert auf der 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, Florida.
- Vo, H., & Silver, C. (2016). Scaling up through parallel and distributed computing. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research:*

- Theory and practical approaches. (S. 125-144). Boca Raton, FL: Chapman and Hall/CRC Press.
- Wang, W., Rothschild, D., Goel, S., & Gelman, A.. Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31, 980-991.
- Watts, D.J. (2013). Computational social science: Exciting progress and future directions. *The Bridge on Frontiers of Engineering*, 43(4), 5-10.
- Wiengarten, L. & Zwick, M. (2018). Integration neuer digitaler Datenquellen in der amtlichen Statistik. In C. König, J. Schröder, E. Wiegand (Hrsg.), *Big Data – Chancen, Risiken, Entwicklungstendenzen* (S. 43-60). Wiesbaden: Springer VS.
- Willems, K. (2015, 12. Mai). Choosing R or Python for data analysis? An infographic. [Blog post] Abgerufen von <https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis> am 7.6.2019.
- Wissenschaftsrat. (2019). Empfehlungen zu hochschulischer Weiterbildung als Teil des lebenslangen Lernens. Vierter Teil der Empfehlungen zur Qualifizierung von Fachkräften vor dem Hintergrund des demographischen Wandels. Drs. 7515-19. Berlin.
- Yalcin, M.A., & Plaisant, C. (2016). Information visualization. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, and J. Lane (Hrsg.), *Big Data and social science research: Theory and practical approaches*. (S. 243-264). Boca Raton, FL: Chapman and Hall/CRC Press.