

Empirische Sozialforschung, Computational Social Science und Big Data. Im Spannungsfeld zwischen neuen Möglichkeiten und alten Grenzen

Marc Hannappel, Universität Koblenz-Landau
Viola Dombrowski, Universität Koblenz-Landau

Die klassische empirische Sozialforschung mit ihrem umfangreichen Repertoire an quantitativen und qualitativen Methoden sieht sich sukzessive von einer neuen Form von Datenerhebungs- und -auswertungsverfahren herausgefordert. Auch die soziologische Theorie bleibt von dieser Entwicklung nicht unberührt, wurde doch bereits mit dem Verweis auf die ‚unbegrenzten‘ Möglichkeiten, die mit diesem neuen ‚Paradigma‘ der *Computational Social Science* (CSS) verbunden seien, das Ende der Theorie proklamiert. *Computational Social Science* sowie auch der Diskurs um *Big Data* sind mit der Hoffnung verbunden, in komplexen Gesellschaften mit unüberschaubaren sozialen Strukturen, Muster sozialen Handelns sichtbar und somit verstehbar zu machen. Ohne Übertreibung kann hier ein Spannungsfeld zwischen klassischen Formen der Sozialforschung und neuen Formen digitaler Datenerhebung und -auswertung diagnostiziert werden.

Unklar ist über weite Strecken hingegen immer noch, welche Art von Daten und Analysen unter diese Begriffe fallen. Während vor wenigen Jahren *Computational Social Science* und *Big Data* mehrheitlich noch mit der Analyse von Daten aus dem Internet assoziiert wurde (so auch die Beschreibung des Arbeitsschwerpunktes der Arbeitsgruppe CSS des GESIS-Instituts im Jahr 2017), ist in der jüngsten Vergangenheit eine Expansion in zahlreiche Anwendungsgebiete festzustellen. Das betrifft beispielsweise die Analyse von geographischen Informationen (GPS-Tracking) bis hin zu Ansätzen, die Fernerkundungsdaten mit Zensusdaten matchen und somit eine Verbindung zwischen objektiven Daten wie Luftverschmutzung und sozialstrukturellen Personendaten herstellen können (Data Linkage). Aber auch Simulationen wie agentenbasierte Verfahren und Mikrosimulationen, sollen, so ein viel beachtetes Plädoyer von Claudio Cioffi auf der „8th International ACM Web Science Conference“ 2016, als Verfahren der CSS aufgefasst werden.

Big Data und *Computational Social Science* befinden sich aktuell in einer ähnlichen Lage wie die Umfrageforschung in der ersten Hälfte des letzten Jahrhunderts. Die Etablierung der – auf Stichproben basierenden – Umfrageforschung im universitären Forschungsalltag ist eng mit dem Namen George Gallup verbunden, der mit seiner Wahlprognose während der amerikanischen Präsidentschaftswahl 1936 den Stichprobenverfahren zum Durchbruch verhalf

(gleichwohl Gallup eher eine Quotenstichprobe denn eine Zufallsstichprobe erhoben hatte). Eine solche Anwendungsorientierung überwiegt auch in den Verfahren, die dem *Computational-Social-Science-Ansatz* zugeschrieben und aktuell hauptsächlich außerhalb universitärer Forschung, etwa zum Zwecke von Markt- und Meinungsforschung innerhalb großer privatwirtschaftlicher Unternehmen, eingesetzt werden. Erst in den letzten Jahren haben sich diese Ansätze, vorwiegend noch innerhalb der Informatik, an deutschen Universitäten etabliert. Im Zuge der Integration dieser neuen Verfahren in den akademischen Forschungsalltag, finden nun methodologische und methodische Reflexionsprozesse und Diskurse statt. Nach einer anfänglichen Euphorie mehren sich daher nun auch immer mehr kritische Stimmen, die aus Perspektive der klassischen Sozialforschung die Aussagekraft der ‚neuen‘ Daten im Hinblick auf Gütekriterien und inferenzstatistischer Probleme in Frage stellen. Dass große Datenmengen kein Garant für Repräsentativität sind, ist spätestens seit dem ‚Literary Digest Disaster‘ bekannt. So ist in vielen Forschungsansätzen, in denen internetbasierte Analysen durchgeführt werden, ungeklärt, über wen die Ergebnisse eine Aussage treffen sollen (Grundgesamtheit vs. Stichprobe), welche Personen in dem Sample sind und, was neuerdings erschwerend hinzukommt, ob es sich dabei überhaupt um einzelne (reale) Personen handelt.

Folglich scheinen vor allem solche Ansätze methodologisch erfolgsversprechend, die im Sinne der klassischen empirischen Sozialforschung klare Begriffsdefinitionen, theoriegeleitete Operationalisierung und Indikatorenbildung sowie eine präzise Definition der Zielpopulation erlauben und diese mit neuen Verfahren der Datenerhebung und -auswertung kombinieren.

Im Rahmen der Ad-Hoc-Gruppe sollen daher solche Ansätze vorgestellt werden, die diese Probleme aufgreifen und Lösungsansätze zur Diskussion stellen. Es wird daher um Beiträge gebeten, die aus methodologischer Perspektive dieses Spannungsfeld zwischen neuen Möglichkeiten und alten Grenzen diskutieren.

Im Rahmen der Ad-hoc-Gruppe sollen fünf thematisch passende Beiträge mit einer Länge von je 30 Minuten (20 Minuten Vortrag, 10 Minuten Diskussion) zugelassen werden. Bitte reichen Sie dazu Ihre **einseitigen** Abstracts (max. 2400 Zeichen inkl. Leerzeichen) bis zum **30.04.2020** an marchannappel@uni-koblenz.de oder vdombrowski@uni-koblenz.de ein.