

Ist Anonymisierung Fälschung? Ein Kommentar zu Stefan Kühl (2020)

Steffen M. Kühnel und Stefanie Eifler

Gibt es Situationen, in denen ein Forscher Daten fälschen darf oder sogar fälschen sollte? Stefan Kühl scheint das zu meinen, wenn er seinen Beitrag „Zwischen Präzision und Anonymisierung. Wie weit muss man bei der Verfälschung wissenschaftlicher Daten gehen?“, erschienen in der jüngsten Ausgabe der *Soziologie* (2020, Jg. 49, Heft 1: 62-71) und zusätzlich auch in der FAZ (08.01.2020) unter dem Titel „Wie weit darf die Verfälschung von Daten gehen? Die Verzerrung von Forschungsdaten ist nicht in jedem Fall verwerflich. Manchmal ist sie sogar unumgänglich“, mit dem Satz beginnt: „Ich habe in meinen wissenschaftlichen Arbeiten Daten verfälscht“ (a.a.O., 2020: 62). Natürlich weiß er als verbeamteter Hochschullehrer, dass dies eine gefährliche Aussage ist, die – wenn sie zuträfe – im schlimmsten Fall mit dem Verlust seiner Professur geahndet werden könnte. Tatsächlich will er mit dieser provozierenden Aussage auf ein Dilemma der empirischen Sozialforschung zwischen dem Fälschungsverbot von Daten und dem Anonymisierungsgebot hinweisen.

Doch besteht ein solches Dilemma überhaupt? Dieser Frage möchten wir im Folgenden nachgehen, in dem wir das Fälschungsverbot von Daten und das Anonymisierungsgebot vor dem Hintergrund des normativen Rahmens, an den empirische Sozialforschung gebunden ist, erörtern. Zentrale Elemente dieses normativen Rahmens sind zum einen das Gebot der intersubjektiven Überprüfbarkeit und Kritizierbarkeit von Aussagen, zum anderen das Gebot, die Grundrechte von Probanden zu respektieren, was sich insbesondere in der Forderung niederschlägt, die informierte Einwilligung zur und Freiwilligkeit der Teilnahme an empirischen Untersuchungen sicherzustellen, jede Art von Schaden von Probanden abzuwenden, die Privatsphäre der Probanden zu achten und die Vertraulichkeit der generierten Daten zu gewährleisten (vgl. hierzu Satzung der Ethik-Kommission der Deutschen Gesellschaft für Soziologie: <https://soziologie.de/ethik>, Abrufdatum 24.02.2020).

Das Gebot der intersubjektiven Überprüfbarkeit und Kritizierbarkeit von Aussagen mündet in die forschungsethische Norm, weder Primärdaten noch darauf basierende Analyseergebnisse – etwa Darstellungen in Tabellen oder Abbildungen zu fälschen. Ansonsten können Artefakte bzw. Fakes nicht von Fakten unterschieden werden. Das Gebot, die Grundrechte von Probanden zu respektieren, was im vorliegenden Zusammenhang insbesondere das in Deutschland einklagbare Grundrecht auf informationelle Selbstbestimmung betrifft, führt dazu, dass Forschungsdaten meist faktisch oder absolut anonymisiert werden.

Aber was bedeutet eigentlich Anonymisierung? Ist Anonymisierung gleichzusetzen mit Fälschung? Wir meinen, nein.

Meist wird unter Anonymisierung verstanden, dass die individuelle Person, über die Informationen vorliegen, weder namentlich noch geographisch als Teilnehmerin an einer Untersuchung geoutet werden kann, dass also Informationen nicht der Person, von der sie stammen, zugordnet werden können. In Deutschland wird für zugängliche Forschungsdaten rechtlich die etwas weniger strenge faktische Anonymisierung verlangt, nach der der Aufwand zur Re-Identifikation einer Person so hoch sein muss, dass er nicht lohnenswert ist.

Es ist allerdings möglich, dass ein Mensch oder eine Organisation seine oder ihre Daten freigibt. Stefan Kühl meint allerdings, dass auf Anonymisierung in der Regel nur dann verzichtet wird, wenn nichts Relevantes herauskommt (Kühl, a.a.O.: 65). Forschungsethisch problematisch wird diese Meinung, wenn „relevant“ mit „neu“ gleichgesetzt wird, da Bestätigungen „bekannter“ Sachverhalte hilfreich oder sogar notwendig sind, um deren Gültigkeit einzuschätzen. Stefan Kühl bezieht sich im Kontext seiner Organisationsforschung darauf, dass man ohne Anonymisierung „vielleicht einiges über

die Vorderbühne, aber faktisch nichts über die Hinterbühne der Organisation erfährt“ (Kühl a.a.O: 66). Offenbar geht er davon aus, dass primär die „Hinterbühne“ relevant ist und diese nicht identifizierbar sein darf. Um diese These zu rechtfertigen, bedarf es aber des Vergleichs unterschiedlicher Vorder- und Hinterbühnen. Je mehr Organisationen betrachtet werden, desto leichter sollte es fallen, die einzelne Organisation in der Menge der Organisationen untergehen zu lassen.

Zudem haben Versuchsteilnehmer das Recht, die Zustimmung zur Nutzung ihrer Daten zurückzunehmen. Erst wenn ihre Daten nicht mehr identifizierbar sind, geht dieses Recht verloren. Das Datennutzungsrecht kann auch für einen bestimmten Zweck, etwa eine konkrete Studie eingeschränkt werden und darf dann nicht für andere Studien genutzt werden, woran sich auch der Primärforscher zu halten hat.

Schwierig ist es, wenn der Datenzugang nur konkreten Personen gewährt wird. Dies kann etwa bei Abschlussarbeiten in einer Firma der Fall sein, wenn Angst vor dem Verlust von Betriebsgeheimnissen besteht. Dann ist die Norm der intersubjektiven Überprüfbarkeit nicht gegeben. Solche Studien arbeiten mit wissenschaftlichen Methoden, können aber streng genommen nicht als Beitrag zur wissenschaftlichen Forschung gewertet werden. Für Bachelor- oder Masterarbeiten, bei denen die Beherrschung des Handwerks gezeigt werden muss, ist dies akzeptabel. Dissertationen müssen demgegenüber publiziert werden. Wenn eine beforschte Organisation dies untersagt, kann die Annahme der Dissertation und damit das gesamte Promotionsverfahren in Frage gestellt sein. Bei der Bitte um Freigabe von Informationen für wissenschaftliche Forschung sollte daher explizit darauf hingewiesen werden, dass die Freigabe für wissenschaftliche Forschung erfolgt und Wissenschaftler die Daten auch nur hierfür nutzen dürfen.

Daten werden aber auch in der Lehre eingesetzt. Public-Use-Files sind sogar für jedermann nutzbar. Hier ist hinreichende Anonymisierung grundsätzlich notwendig. Es gibt dazu unterschiedliche Methoden. So können solche freien Daten Teilstichproben sein, bei denen sichergestellt ist, dass eine Re-Identifikation nicht funktioniert. Zudem können Variablenausprägungen aggregiert werden, also etwa Altersangaben in Altersgruppen zusammengefasst werden. Auch ist die Addition bzw. Substitution durch Zufallszahlen möglich, wobei dann die Regeln der Datenveränderung bekannt sein müssen, um die Daten korrekt analysieren zu können.

Stefan Kühl spricht in seinem Beitrag (2020: 62) davon, dass er in seinen Publikationen die geografische Lage von Organisationen verändert habe, die Zahl der tatsächlich untersuchten Organisationen verändert habe und von Interviews in anderen Organisationen als den beschriebenen. Er sieht hierin eine mangelnde Präzision der Darstellung zugunsten von Anonymisierung.

Präzisionsverzicht ist jedoch etwas grundlegend Anderes als Fälschung. Fälschung von Forschungsdaten liegt vor, wenn die berichteten Forschungsergebnisse darauf basieren, dass anstelle empirischer Daten fiktive Daten analysiert oder berichtet werden. Dies meint Kollege Kühl vermutlich nicht, denn das wäre tatsächlich ein Verstoß gegen die Regeln guter wissenschaftlicher Praxis. Aber es gibt einen Graubereich. Forschungsergebnisse sind nicht kontextfrei. Wenn Anonymisierung im Köhlschen Sinne zu einer Änderung des Kontextes führt, in dem die berichteten Ergebnisse tatsächlich nicht zutreffen, dann handelt es sich nicht mehr um unzureichende Präzision, sondern um falsche Ergebnisse. Eine so verstandene Anonymisierung ist Fälschung. Soweit darf es nicht gehen.

Und was soll eine Forscherin tun, wenn sie ohne solche Verfälschungen keine Freigabe erhält? Dann kann sie im Zweifelsfall nicht publizieren, was also die Arbeit umsonst. Reduzieren lässt sich dieses Risiko, wenn vor der Forschung den Probanden bzw. beforschten Organisationen zutreffende und hinreichende Informationen gegeben werden, mit anderen Worten die informierte Einwilligung zur Teilnahme an einer Untersuchung und zur (Nach-)nutzung der generierten Daten eingeholt wird.

Wenn dies die Erreichung des Forschungsziels in Frage stellt oder erst im Verlaufe der Forschung der Informationsbedarf der Forschung klar wird, muss diese Einwilligung frühestmöglich eingeholt werden. Ausschließen lässt sich nicht, dass die Nutzung von Daten für die Forschung nicht gegeben wird oder zurückgezogen wird.

Ein solcher Normkonflikt ist allerdings kein Dilemma, d.h., eine ausweglose Situation, in der jede Alternative gegen grundlegende Normen verstößt. Vielmehr ist der Normkonflikt aufgrund des normativen Rahmens, an den empirische Sozialforschung gebunden ist, immer schon gelöst. In diesem Sinne besteht folglich auch kein Dilemma zwischen Anonymisierung und Fälschung, sondern ein Primat zugunsten der Beforschten und zuungunsten der Forscher. Damit müssen wir leben.