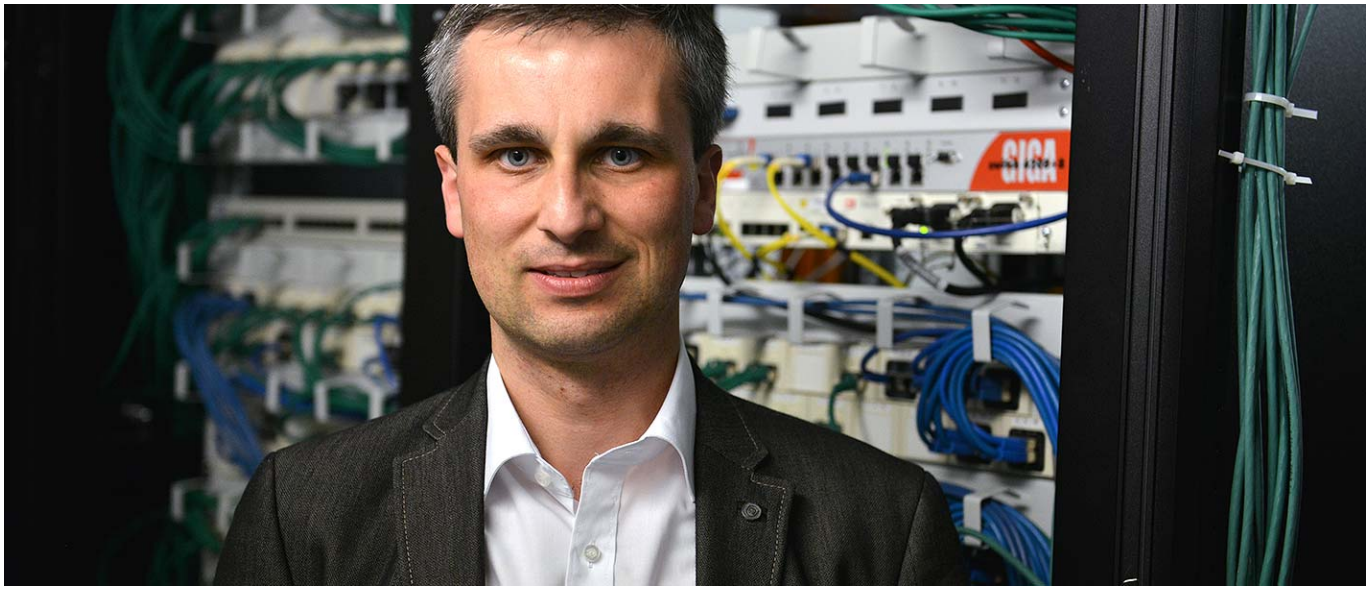
Datum: 1. November 2017



„Im Kern geht es darum, den Datensalat zu entwirren“

Christian Dickmann vom infas Institut für angewandte Sozialwissenschaft bereitet die Daten für das Panel Arbeitsmarkt und soziale Sicherung (PASS) auf und ist für den Datenschutz zuständig. Im Interview erläutert er, mit welchen Aufgaben er sich hier befasst, worauf es dabei ankommt und was die Studie für ihn persönlich bedeutet.

Herr Dickmann, Sie arbeiten bei infas, das im Auftrag des IAB die Befragungen für das PASS durchführt, in der Datenaufbereitung. Was gehört hier zu Ihrem Aufgabenbereich?
Ich bin seit 2007 bei infas in der Datenaufbereitung tätig und seit 2009 schwerpunktmäßig mit der Studie „Lebensqualität und soziale Sicherung“ betraut. Ich befasse mich daneben noch mit einigen weiteren Projekten, aber die Studie des IAB stellt den Großteil meiner Arbeit dar. Neben den Tätigkeiten in der Datenaufbereitung kommt außerdem eine Reihe von Aufgaben in meiner Funktion als Datenschutzbeauftragter hinzu.

Welche Aufgaben fallen bei der Aufbereitung der Befragungsdaten an?

Ein Zyklus beginnt im Prinzip mit dem Fragebogen, der vom IAB an uns geschickt wird und dann in ein technisches Instrument übersetzt werden muss. Im Vorfeld der Feldphase teste

ich diese Programmierung auf ihre Richtigkeit: Bekommen die Personen auch diejenigen Fragen präsentiert, die für sie bestimmt sind? Stimmen die hinterlegten Werte aus den Vorjahren noch? Läuft der Fragebogen fehlerfrei ab? Nach der Feldphase beginnt dann die Aufbereitung des Enddatensatzes, die uns hier bei infas ein ganzes Jahr lang begleitet. In erster Linie geht es dabei um die Zusammenfassung und Vereinheitlichung unterschiedlicher Angaben. Besonders zeitintensiv ist die Vercodung offener Angaben, also die Zuordnung von numerischen Codes zu den in Textform vorliegenden Antworten. Denn damit die Forscher später mit den Daten arbeiten können, müssen wir Texte auf wenige numerische Werte verdichten. Zum Abschluss steht noch die schriftliche Dokumentation der Feldarbeit an. Und dann beginnt schon wieder der neue Zyklus.

> „Wir in der Datenaufbereitung sorgen dafür, dass die Wissenschaftler später mit den Daten arbeiten können."

Warum müssen die Angaben der Befragten denn überhaupt noch einmal aufbereitet werden? Was ist der Sinn und Zweck dieses Vorgehens?

Einerseits geht es darum, die Informationen inhaltlich so zu lassen wie sie sind, sie andererseits aber etwas handhabbarer zu machen. So, wie die Daten abgefragt werden, entsprechen sie nicht der Form, die Forscherinnen und Forscher für ihre Analysen benötigen. Wir in der Datenaufbereitung sind also für die Überführung zuständig und sorgen dafür, dass die Wissenschaftler später mit den Daten arbeiten können. Im Kern geht es darum, den Datensalat zu entwirren und den Zahlen einen inhaltlichen Kontext zu geben, der für alle verständlich ist.

> „Wir haben ein ganzes Bündel an Maßnahmen, um Rückschlüsse auf die wahre Identität eines Befragten ausschließen zu können."

Ein wichtiger Teil Ihrer Tätigkeit besteht in der Anonymisierung der Daten. Welche Schritte werden hierzu bei der Studie unternommen?

Wir haben ein ganzes Bündel an Maßnahmen, um Rückschlüsse auf die wahre Identität eines Befragten ausschließen zu können. Zunächst einmal geben wir die Namen, Adressen und Telefonnummern der Befragten nicht an die Forscher weiter. Diese Informationen werden bei uns lediglich zur Kontaktaufnahme genutzt. Im Datensatz selbst wird nur eine achtstellige Codenummer vermerkt. Auch viele weitere erfragte Angaben werden gar nicht oder nur in

sehr stark vergröberter Form zur Verfügung gestellt. Beispielsweise wird als Wohnort nicht Bonn hinterlegt, sondern nur die Information, dass der Befragte in einer Stadt mit einer Bevölkerungsgröße zwischen 100.000 und 500.000 Einwohnern im Bundesland Nordrhein-Westfalen wohnt. Ein weiteres Beispiel ist die Berufsangabe: Selbst wenn jemand im Interview erzählt, dass er Kurierfahrer bei Hermes in Bonn ist, werden bei der Datenaufbereitung der Ort und die Firma gestrichen und lediglich „Kurierfahrer“ hinterlegt und in eine Zahl übersetzt. Wir schauen alle Angaben noch einmal gründlich durch, bevor wir sie herausgeben. Auch biografische Daten werden von uns vergröbert und in größeren Gruppen zusammengefasst. So stellen wir beispielsweise nur das Geburtsjahr zur Verfügung – und nicht das genaue Geburtsdatum. Ähnliches gilt für den Migrationshintergrund, die Herkunftsländer oder die Haushaltszusammensetzung.

> „Damit die Forscher später mit den Daten arbeiten können, müssen wir Texte auf wenige numerische Werte verdichten.“

Sie haben sich mehrere Jahre lang intensiv mit der Datenstruktur der Studie auseinandergesetzt. Worin liegt für Sie die größte Herausforderung?

Auch wenn es trivial klingen mag: Man muss den Überblick darüber behalten, was in der aktuellen Welle gefragt wurde. Gerade beim Panel Arbeitsmarkt und soziale Sicherung kommen immer wieder neue Fragen und Module hinzu, so dass der Datensatz mit jedem Jahr gewachsen ist und wir bald schon mehr als 40.000 unterschiedliche Personen und mehr als 1.500 verschiedene Variablen haben. Um diesen großen Umfang managen zu können, bedarf es ganz spezieller Prüfmethoden und Instrumente, die wir oft erst selbst entwickeln mussten. Für mich persönlich war die größte Herausforderung, mich in die Arbeiten meiner Vorgänger einzulesen, als das gesamte Paket im Herbst 2009 an mich übergeben wurde. Es ist immer schwer, sich nachträglich in eine fremde Gedankenwelt einzuarbeiten und das Ganze dann noch möglichst bruchlos fortzuführen. Der Aufwand war damals intellektuell schon sehr groß. Es war einfach nicht der eigene Programmierstil, und man musste damit irgendwie heimisch werden.

> „Ich habe da ein Kind vor Augen, das einen sehr großen Turm baut und immer noch ein Klötzchen oben draufsetzt.“

Und sind Sie das geworden?

Ja, jetzt sind wir im angenehmen Teil der Studie angekommen. Es hat sich allmählich alles eingespielt, und ich hoffe, dass es noch viele Jahre oder gar Jahrzehnte weitergeht. Das ist jetzt eine schöne Herausforderung. Man hat gewisse Routinen entwickelt und weiß, wie man zum nächsten Ergebnis kommt. Gleichzeitig gibt es immer wieder Änderungen im Fragebogen, die für Abwechslung sorgen.

Was bedeutet Ihnen das Panel ganz persönlich?

Für mich ist es eine sehr interessante Studie, die beständig wächst und viel Potenzial bietet. Ich habe da immer ein Kind vor Augen, das einen sehr großen Turm baut und dann immer noch ein Klötzchen oben draufsetzt. Und das ist wirklich spannend. Es macht mir einfach Spaß, hierbei mitzuwirken. Das motiviert einen dann schon aus sich heraus. Deshalb würde ich sagen, dass die Bindung an das Projekt sehr eng ist. Man will einfach sehen, wie es weiterwächst.

Zur Person

Christian Dickmann, Magister Artium, ist seit 2007 für infas als Senior-Projektleiter im Bereich des komplexen Datenmanagements tätig. Sein Arbeitsschwerpunkt besteht in der Datenprüfung und -aufbereitung des Panels Arbeitsmarkt und soziale Sicherung des Instituts für Arbeitsmarkt- und Berufsforschung. Er ist zudem verantwortlich für Datenprüfungen weiterer kontinuierlicher Haushaltspanelstudien und Querschnittbefragungen sowie für die Organisation der Datensätze. Seit 2014 nimmt Christian Dickmann außerdem die Aufgaben des betrieblichen Datenschutzbeauftragten bei infas wahr.

Die Fragen stellte Daniel Meyer, derzeit als Doktorand am Max-Planck-Institut für Gesellschaftsforschung in Köln tätig.