

Forschungsdatenzentrum (FDZ) der Statistischen Landesämter:

Amtliche Mikrodaten für die wissenschaftliche Forschung – Nutzerkonferenzen 2005

Sabine Haug, Nadine Bartke

Das Forschungsdatenzentrum der Statistischen Landesämter hat das Ziel, der Wissenschaft den Zugang zu Mikrodaten aus der amtlichen Statistik zu erleichtern. Damit werden Einzeldatensätze aus statistischen Erhebungen, die in Deutschland zu über 90 % dezentral in den einzelnen Bundesländern vorliegen, zentral zugänglich gemacht und On-Site an 16 Standorten in Deutschland zur Verfügung gestellt. Zudem können in zunehmendem Maße Mikrodaten aus immer mehr Statistikbereichen auch Off-Site genutzt werden. Umfangreiche Hintergrundinformationen zu den Statistiken können sich die Wissenschaftler zukünftig auch über das Internet beschaffen. Auf einer von vier Nutzerkonferenzen in Deutschland besteht für Interessenten am 9. Mai 2005 an der Universität Stuttgart die Möglichkeit, Erfahrungen auszutauschen und sich über das Datenangebot sowie die Nutzungsmöglichkeiten des Forschungsdatenzentrums der Statistischen Landesämter zu informieren. Fachliche Schwerpunkte der Veranstaltung in Stuttgart sind die Bereiche Umwelt, Hochschulbildung, Rechtspflege sowie Lohn- und Gehaltsstruktur. Weitere Informationen zum Veranstaltungsprogramm unter www.forschungsdatenzentrum.de.

Aufbereitung und Nutzung von Mikrodaten verschiedener Statistikbereiche

Mit der Gründung des Forschungsdatenzentrums (FDZ) der Landesämter im Jahr 2002 haben es sich die Statistischen Landesämter zur Aufgabe gemacht, der Wissenschaft einen besseren Zugang zu den Einzeldaten (Mikrodaten) der amtlichen Statistik zu offerieren. Seit Januar 2004 wird das FDZ der Statistischen Landesämter als Pilotprojekt für eine Zeit von 3 Jahren vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Aufgrund der Beteiligung aller Bundesländer umfasst das FDZ der Statistischen Landesämter 16 regionale Standorte. Das FDZ der Landesämter als Projekt der Statistischen Landesämter steht in einer engen Kooperation mit dem FDZ des Statistischen Bundesamtes.

Das Datenangebot des FDZ der Landesämter konzentriert sich derzeit auf Mikrodaten folgender Statistikbereiche:

- Bevölkerung
- Hochschulbildung
- Gesundheitswesen
- Rechtspflege
- Umwelt
- Agrarwirtschaft
- Produzierendes Gewerbe
- Tourismus
- Lohn- und Gehaltsstruktur
- Öffentlicher Haushalt

Dieses Mikrodatenangebot wurde durch eine vorhergehende Befragung von 700 potenziellen Nutzern ermittelt. Nach und nach wird das Angebot um weitere Statistiken und neue Fachbereiche ergänzt. Mit Beginn der Arbeiten im FDZ der Statistischen Landesämter haben die regionalen Standorte mit der Aufbereitung von 26 Statistiken aus diesen Fachbereichen für die wissenschaftliche Nutzung begonnen. Um der Wissenschaft den Zugang zu den Mikrodaten der amtlichen Statistik zu erleichtern, umfasst das Arbeitsprogramm des FDZ der Statistischen Landesämter im Wesentlichen die Bearbeitung folgender Aufgaben:

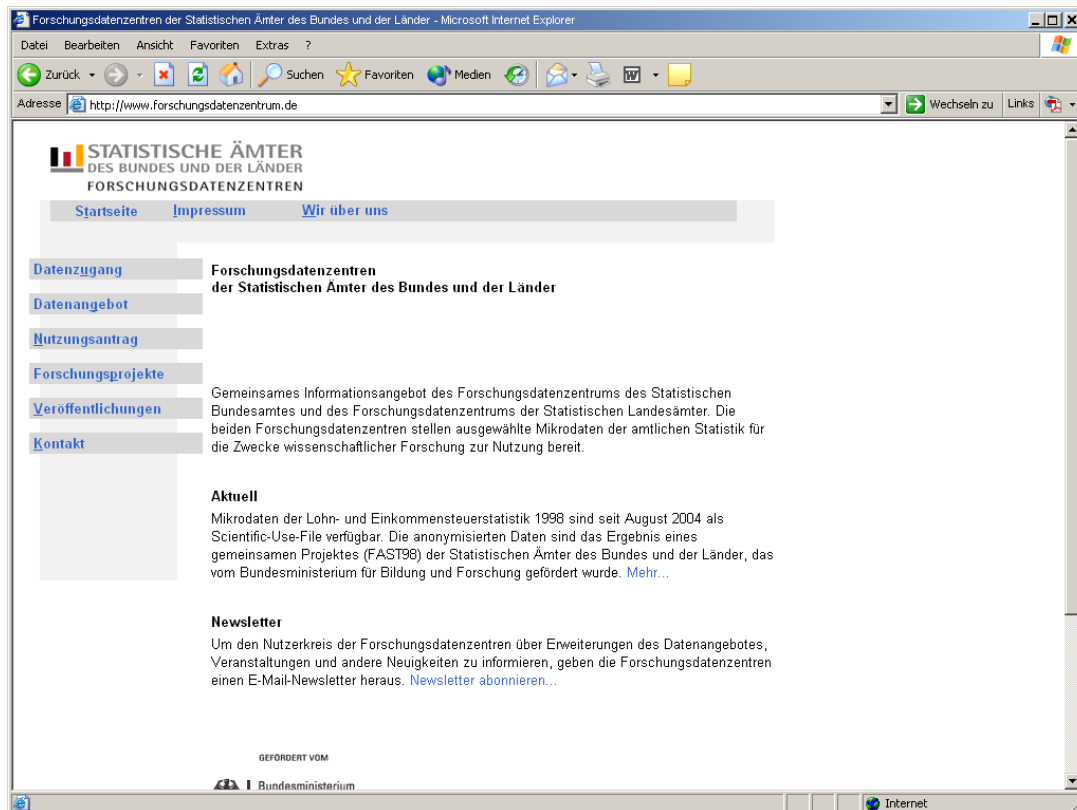
1. Zentralisierte Datenhaltung

Eine Vielzahl von wissenschaftlichen Untersuchungen bezieht sich auf eine Auswahl von Bundesländern oder gegebenenfalls auf das gesamte Bundesgebiet. In der Vergangenheit wurden derartige Forschungsprojekte dadurch erschwert, dass die Mikrodaten der jeweiligen Bundesländer zunächst relativ zeitaufwändig zusammengeführt und aufbereitet werden mussten. Dies betrifft über 90 % aller Statistiken in Deutschland, die in den jeweiligen Bundesländern erhoben, aufbereitet und gespeichert werden. Um der Wissenschaft in diesem Kontext den Zugang zu den Mikrodaten zu erleichtern, wird eine fachlich zentralisierte Datenhaltung zu den einzelnen Statistiken eingerichtet. Dabei speichern die jeweiligen regionalen Standorte aus einem oder zwei der oben auf-

Dipl.-Ing. Sabine Haug ist FDZ-Mitarbeiterin im regionalen Standort Stuttgart des Statistischen Landesamtes Baden-Württemberg.

Dipl.-Kffr. Nadine Bartke ist FDZ-Mitarbeiterin im regionalen Standort München des Bayerischen Landesamtes für Statistik und Datenverarbeitung.

Abbildung



geführten Fachbereiche Einzeldatenmaterial einer statistischen Erhebung über das gesamte Bundesgebiet. Dem Wissenschaftler ist es aber unabhängig davon möglich, länderübergreifend mit jedem der regionalen Standorte bezüglich der von ihm gewünschten Mikrodaten in Kontakt zu treten und Datenmaterial aus allen Bundesländern zu erhalten.

II. Zugangswege für die Wissenschaft

Für die jeweiligen Fachstatistiken stehen dem Forscher verschiedene Formen des Datenzugangs zur Verfügung. Zum einen besteht die Möglichkeit, dass Wissenschaftler standardisierte Mikrodatendateien außerhalb der statistischen Ämter nutzen. Für derartige Datenzugänge, die auch als **Off-Site-Nutzung** bezeichnet werden, müssen Mikrodaten zunächst anonymisiert werden. Es werden verschiedene Mikrodaten-Files angeboten, die sich im Grad der Anonymisierung und im möglichen Nutzerkreis unterscheiden. Mit den „Public Use Files“ steht allen Interessenten die Nutzung absolut anonymisierter Mikrodaten offen. Faktisch anonymisierte, das heißt weniger stark anonymisierte Mikrodaten, welche mit den „Scientific Use Files“ angeboten werden, dürfen gemäß dem Bundesstatistikgesetz lediglich unabhängigen wissenschaftlichen Institutionen bereitgestellt werden. Erste „Scientific Use Files“ existieren

beispielsweise bereits für den Mikrozensus sowie für die Lohn- und Einkommensteuerstatistik.

Sofern von Wissenschaftlern Mikrodaten gewünscht werden, die weniger stark anonymisiert wurden als die der Off-Site-Nutzung, besteht eine Möglichkeit, diese an einem Gastwissenschaftlerarbeitsplatz innerhalb des Statistischen Landesamtes zu nutzen. Solche Arbeitsplätze wurden von allen regionalen Standorten eingerichtet und werden von Wissenschaftlern zunehmend in Anspruch genommen.

Des Weiteren wird die Nutzung von Mikrodaten in Form der kontrollierten Datenfernverarbeitung angeboten. Dabei übernimmt der Wissenschaftler die Programmierung seines Analysevorhabens in den Auswertungsprogrammen SPSS, SAS oder STATA. Von den FDZ-Mitarbeitern wird dieses Auswertungsprogramm an den Originaldatensätzen ausgeführt, die lediglich durch Weglassen von Personen- und Hilfsmerkmalen (formal) anonymisiert wurden. Nach Prüfung der statistischen Geheimhaltung durch die betreuenden FDZ-Mitarbeiter werden die erzeugten Ergebnisse letztlich an den Wissenschaftler übermittelt. Die beiden zuletzt beschriebenen Zugangswege, die Nutzung des Gastwissenschaftlerarbeitsplatzes und der kontrollierten Datenfernverarbeitung, werden auch als **On-Site-Nutzung** bezeichnet.

III. Beschaffung von Hintergrundinformationen

Zur zielgerichteten Bearbeitung von Mikrodaten benötigt der Wissenschaftler umfangreiche Hintergrundinformationen (Metadaten) über die Statistiken. Von besonderem Interesse sind dabei Informationen über die Datensatzstruktur, die Erhebung, die Aufbereitung und die Qualität des Datenmaterials. Zu diesem Zweck arbeiten die Statistischen Landesämter gemeinsam mit dem Statistischen Bundesamt an der Implementierung eines **Metadateninformationssystem**s. Mit diesem werden Forscher zukünftig in der Lage sein, die benötigten Informationen über die einzelnen Statistiken über das Internet abzufragen.

Nutzerkonferenzen 2005 des FDZ der Statistischen Landesämter

Um interessierte Wissenschaftler über die Arbeiten im FDZ der Landesämter zu informieren, sind im Frühjahr 2005 vier regionale Nutzerkonferenzen geplant. Ziel ist es, potenziellen Datennutzern das Dienstleistungsangebot des FDZ der Statistischen Landesämter näher zu bringen und realisierte Forschungsprojekte von Nutzern zu präsentieren. Um jedem Wissenschaftler die Möglichkeit zu geben, eine Konferenz in seiner nahen räumlichen Umgebung zu besuchen, finden die vier regionalen Nutzerkonferenzen an Universitäten in Stuttgart, Köln, Berlin und Kiel statt.

Die vier Konferenzen haben jeweils unterschiedliche fachliche Schwerpunkte. Im Mittelpunkt der Konferenz an der Universität Stuttgart stehen Vorträge aus dem Bereich der Umwelt- und Sozialstatistiken. Dagegen werden an der Freien Universität Berlin in erster Linie Forschungsvorhaben im Bereich der Wirtschafts- und Sozialstatistiken vorgestellt, die Nutzerkonferenz an der Universität Köln bildet den Schwerpunkt in der Vorstellung von Forschungsvorhaben aus dem Bereich Bevölkerungs- und Sozial-

statistiken und die Konferenz an der Universität Kiel konzentriert sich auf die Vorstellung von Agrar- und Wirtschaftsstatistiken.

Nutzerkonferenz an der Universität Stuttgart

An der Nutzerkonferenz in Stuttgart sind seitens der amtlichen Statistik die Landesämter aus Baden-Württemberg, Bayern, Hessen und dem Saarland beteiligt. Im Mittelpunkt der Veranstaltung stehen Vorträge, die sich mit Umweltstatistiken, insbesondere wasserwirtschaftlichen Erhebungen, der Gehalts- und Lohnstrukturerhebung sowie Hochschulbildungs- und Rechtspflegestatistiken beschäftigen. Die Nutzerkonferenz findet am 9. Mai 2005 an der Universität Stuttgart statt und wird durch den Umweltminister des Landes Baden-Württemberg, Herrn Mappus, eröffnet. Mitveranstalter ist die „Gemeinsame Kommission Umweltschutztechnik“ der Universität Stuttgart.

Das Programm der Tagung ist vielfältig. Dem Nutzer werden neben allgemeinen Darstellungen der Ziele und Nutzungsmöglichkeiten des FDZ anhand einzelner Statistiken insbesondere die Zugangswege, Auswertungsmöglichkeiten und die Metadaten vorgestellt. Außerdem berichten Wissenschaftler, die mithilfe von Einzeldaten aus dem FDZ in der Vergangenheit bereits Forschungsprojekte bearbeitet haben, über ihre Erfahrungen.

Weitere Informationen zu den geplanten Nutzerkonferenzen und das Tagungsprogramm inklusive der Anmeldeunterlagen stehen Interessenten in Kürze im Internet unter www.forschungsdatenzentrum.de zur Verfügung. Zudem findet sich unter der angegebenen http-Adresse eine Aufstellung der regionalen Ansprechpartner. ■

Weitere Auskünfte erteilt
Sabine Haug, Telefon 0711/641-2002
E-Mail: Sabine.Haug@stala.bwl.de