

Nutzung von Mikrodaten amtlicher Agrarstatistiken über das Forschungsdatenzentrum

Susanne Stricker, Hendrik Tietje

Forschungsdatenzentrum des Statistischen Ämter der Länder, Standort Kiel
susanne.stricker@statistik-nord.de, hendrik.tietje@statistik-nord.de

Abstract: Seit dem Jahr 2004 ermöglichen die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder die Umsetzung eines gesetzlich eröffneten Privilegs für die Wissenschaft: den Zugang zu Mikrodaten der amtlichen Statistik für Forschungsvorhaben. Dieser Datennutzungsweg ist mittlerweile gut etabliert und insbesondere in der empirischen Agrarstrukturforschung weit verbreitet. Zum Beispiel arbeiten derzeit rund ein Dutzend Doktoranden im Rahmen ihrer Dissertationen mit den Agrardaten des Forschungsdatenzentrums. Darüber hinaus wird der Datenbestand vielfach auch für Projekt- und Auftragsforschung genutzt.

Dieser Beitrag berichtet über die Entstehungsgeschichte sowie die heutige Ausgestaltung des Forschungsdatenzentrums, sowie über seine Ziele und Aufgaben. Darüber hinaus werden die Zugangswege zu den amtlichen Mikrodaten und erforderliche Anonymisierungsmaßnahmen dargestellt. Abschließend wird das Datenangebot mit einem Agrarfokus kurz vorgestellt.

1 Einleitung

Die Komplexität wirtschaftlichen und sozialen Wandels zusammen mit den Fortschritten in der Wissenschaft und in der Informationstechnik hat zu dem Begehren geführt, dass Wissenschaftler für Ihre empirischen Analysen die Einzeldaten öffentlicher Datenproduzenten für Forschungszwecke nutzen wollen. Vor diesem Hintergrund entstand eine intensive Diskussion, wie die Wissenschaft im Rahmen der geltenden gesetzlichen Bestimmungen die Daten der öffentlichen Datenproduzenten nutzen könnte. Die „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ (KVI) wurde eingesetzt, um die Möglichkeiten und Ausgestaltungen zu analysieren. Diese Kommission empfahl die Einrichtung von Forschungsdatenzentren bei den vier großen öffentlichen Datenproduzenten, der Rentenversicherung Bund, der Bundesagentur für Arbeit, dem Statistisches Bundesamt und bei den Statistischen Ämtern der Länder [ZUE03]. Dieser Beitrag berichtet über die Entstehungsgeschichte sowie die heutige Ausgestaltung des Forschungsdatenzentrums der Statistischen Ämter der Länder sowie über seine Ziele und Aufgaben. Auch die Datenzugangswege zu den amtlichen Mikrodaten und die erforderlichen Anonymisierungsmaßnahmen werden beschrieben. Abschließend wird das derzeitige Datenangebot kurz vorgestellt.

2 Entstehung des Forschungsdatenzentrums

Mit der Einführung eines Wissenschaftsprivilegs in das Bundesstatistikgesetz im Jahre 1987 wurde die gesetzliche Grundlage für den Zugang der Wissenschaft zu faktisch anonymen amtlichen Mikrodaten geschaffen. Mikrodaten gelten als faktisch anonym, wenn diese nur mit einem „unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“ (§16, Absatz 6 im BStatG) [MUE91]. Das bedeutet, dass es nicht gänzlich unmöglich ist, eine einzelne Untersuchungseinheit (z.B. einen Landwirt) zu identifizieren, dies aber mit einem unverhältnismäßig hohen Aufwand verbunden wäre.

Das gesetzlich verankerte Wissenschaftsprivileg führte zusammen mit dem durch Fortschritte in Wissenschaft und Informationstechnik entstandenen Begehren der Wissenschaft zu der Einrichtung des Forschungsdatenzentrums der statistischen Ämter der Länder mit 16 regionalen Standorten. Abbildung 1 beinhaltet auch die fachlichen Schwerpunkte der regionalen Standorte, auf die später noch genauer eingegangen wird. Zunächst wurde das Forschungsdatenzentrum als Arbeitsgemeinschaft aller statistischen Ämter im Jahr 2002 gegründet und ab dem Jahr 2004 als Projekt durch das Bundesministerium für Bildung und Forschung gefördert. Die Förderung endet mit dem Jahr 2009. Es wird sich derzeit um eine dauerhafte Etablierung bemüht.

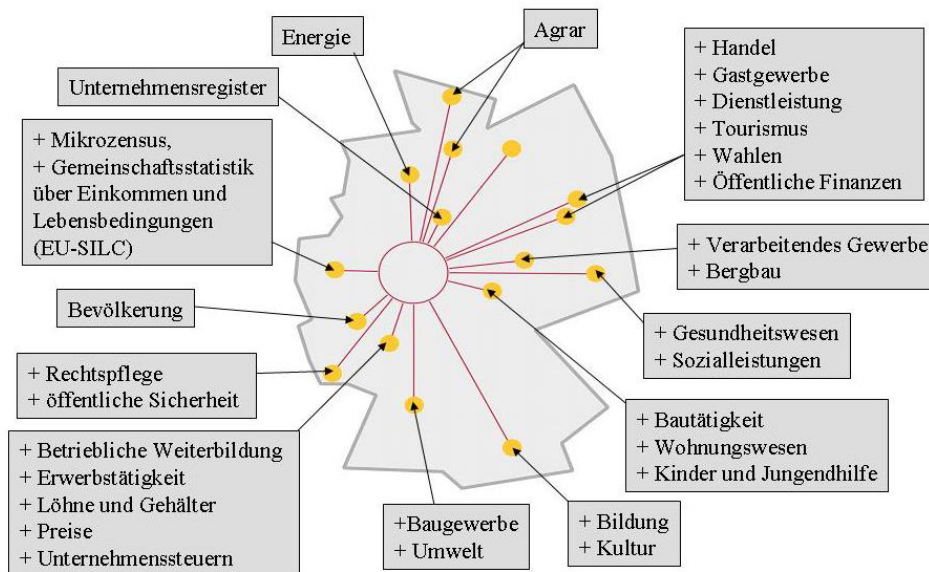


Abbildung 1: Fachlich zentralisierte Datenhaltung an den 16 Standorten des Forschungsdatenzentrums der statistischen Ämter der Länder

3 Ziele und Aufgaben des Forschungsdatenzentrums

Die Ziele und Aufgaben des Forschungsdatenzentrums lassen sich anhand folgender vier Punkte zusammenfassen:

1. Verbesserung der Dateninfrastruktur für wissenschaftliche Analysen
2. Erleichterung des Zugangs der Wissenschaft zu amtlichen Mikrodaten
3. Beratung und Service für die Nutzer
4. Durchführung gemeinsamer Forschungsprojekte mit der Wissenschaft

Das wesentliche Ziel des Forschungsdatenzentrums besteht darin, eine Dateninfrastruktur zu schaffen, die den Zugang der Wissenschaft zu amtlichen Mikrodaten erleichtert. Über 90% der Statistiken in Deutschland werden dezentral durch die Statistischen Landesämter erhoben, aufbereitet und gespeichert. Die eingerichtete fachlich zentralisierte Datenhaltung ermöglicht es, an einem Standort die Mikrodaten von dezentral erhobenen Statistiken vorzuhalten und gemäß den Anforderungen eines Forschungsprojektes aufzubereiten und zeitnah zur Verfügung zu stellen. Um der Wissenschaft den Datenzugang zu erleichtern gibt es verschiedene Datenzugangswege, die in dem folgenden Abschnitt näher beschrieben werden.

Beratung und Service sowohl fachlicher als auch methodischer Art gehört ebenso zum Leistungsspektrum des Forschungsdatenzentrums wie die Durchführung gemeinsamer empirischer Forschungsprojekte mit der Wissenschaft. Diese Forschungsprojekte dienen entweder dazu das Analysepotential der Daten zu erforschen oder aber der Entwicklung neuer statistischer Methoden und Verfahren der Anonymisierung.

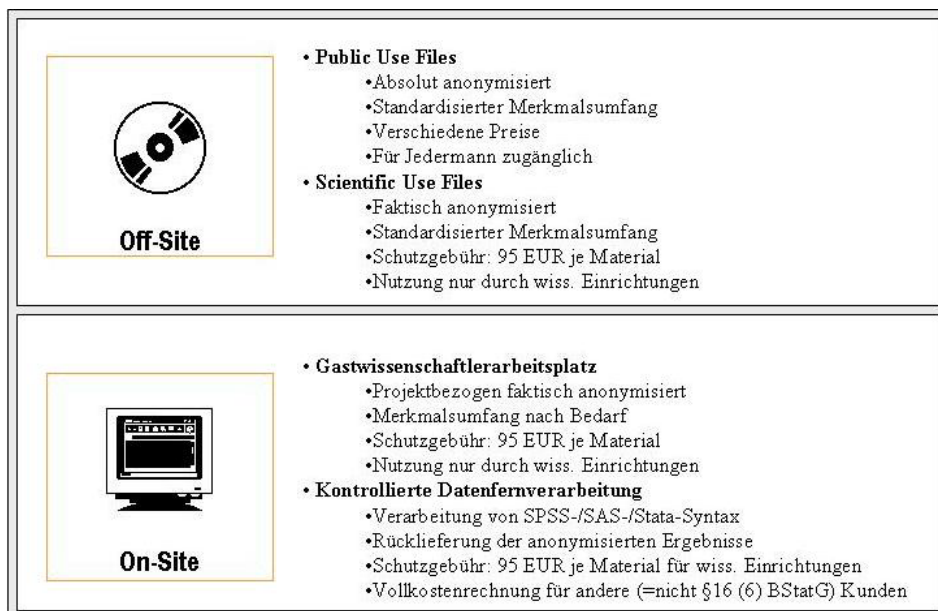


Abbildung 2: Datenzugangswege im Forschungsdatenzentrum der Statistischen Ämter der Länder

4 Datenzugangswege und Anonymisierungsmaßnahmen

Durch die Einrichtung verschiedener Datenzugangswege eröffnete sich für die Wissenschaft eine Form des Datenzugangs, die zuvor nicht möglich war. Ansatzpunkt dieser verschiedenen Nutzungswege ist es, die Reidentifikation einer Auskunftseinheit nicht allein durch Veränderung des Datenmaterials, sondern auch durch die Regulierung des Datenzugriffs sicherzustellen. Unterscheidet man nach dem Ort an dem die Daten sich bei der Analyse befinden (off-site = nicht in einem statistischen Landesamt; on-site = in einem statistischen Landesamt) lassen sich die verschiedenen Zugangswege wie in Abbildung 2 darstellen. Die Datensätze unterscheiden sich nach Grad der Anonymität, Merkmalsumfang, Preisen und Zugangsmöglichkeiten.

5 Datenangebot

Von den statistischen Ämtern der Länder werden rund 250 Statistiken regelmäßig durchgeführt, davon liegen im Forschungsdatenzentrum derzeit rund 60 Datenbestände zu breit gestreuten Themengebieten vor, die sich grob in die folgenden fünf Bereiche gliedern lassen:

- Sozialstatistiken
- Wirtschaftsstatistiken
- Finanz- und Steuerstatistiken
- Rechtspflegestatistiken
- Agrar- und Umweltstatistiken

Für die Agrarforschung sind insbesondere die Mikrodaten der Agrarstrukturerhebung von besonderem Interesse. Diese wird regelmäßig alle zwei Jahre im Mai durchgeführt. Das gleich bleibende Erhebungsprogramm erfasst Produktionsstrukturen und -kapazitäten der landwirtschaftlichen Betriebe sowie Angaben über die wirtschaftlichen und sozialen Verhältnisse der Betriebsinhaber und/oder -leiter. Darüber hinaus umfasst das Grundprogramm der Agrarstrukturerhebung auch Merkmale zur Bodennutzung und zu Viehbeständen sowie Angaben zu Betrieben mit ökologischem Landbau. Einen detaillierten Überblick über das gesamte Datenangebot des Forschungsdatenzentrums der statistischen Ämter der Länder befindet sich unter www.forschungsdatenzentrum.de.

Literaturverzeichnis

- [MUE91] Müller, W., Blien, U., Knoche, P., Wirth, H. u.a.: Die faktische Anonymität von Mikrodaten, Forum des Bundesstatistik, Band 19, 1991.
- [ZUE03] Zühlke, S., Zwick, M., Scharnhorst, S., Wende, T.: Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, Wirtschaft und Statistik, Ausgabe 10/2003.