

Verknüpfung von administrativen und Befragungsdaten: eine Herausforderung

Die Kombination von Umfrage- und Verwaltungsdaten nimmt zu, doch der Datenzugriff ist nach wie vor stark eingeschränkt

Schlagworte: Datensatzverknüpfung, Verwaltungsdatensätze, Umfragedaten

RELEVANZ DES THEMAS

Mit der Kombination sich gegenseitig ergänzender Verwaltungs- und Befragungsdaten erschließt sich der akademischen und politiknahen Forschung ein enormes Potenzial, das sich durch die verbesserte Verfügbarkeit von Datenquellen und technologische Fortschritte beim Datenabgleich zusehends vergrößert. Diese Entwicklungen spiegeln sich auch in der wachsenden Zahl wissenschaftlicher Arbeiten in der Arbeitsökonomie wider, die verknüpfte Erhebungs- und Verwaltungsdaten verwenden. Bedenken hinsichtlich der Datensicherheit und des Datenschutzes, schränken allerdings den Datenzugriff und Erkenntnisgewinn der Forschung ein.

WICHTIGE RESULTATE

Pro

- ⊕ Die Datenverknüpfung überwindet einige der Unzulänglichkeiten beider Datenquellen.
- ⊕ Verknüpfungen eröffnen neue Forschungsmöglichkeiten, indem sehr zuverlässige Verwaltungsdaten mit detaillierten Umfragedaten kombiniert werden.
- ⊕ Routinemäßig erfasste Verwaltungsdaten ergänzen preiswert und zuverlässig Umfragedaten.
- ⊕ Umfragekosten können potenziell sinken, wenn weniger Fragen gestellt zu werden brauchen.
- ⊕ Können sensible Daten – etwa zu den Arbeitseinkommen – aus Verwaltungsdatensätzen gezogen werden, verringert dies das Risiko von Befragungsabbrüchen und hohen Nichtantwortraten.

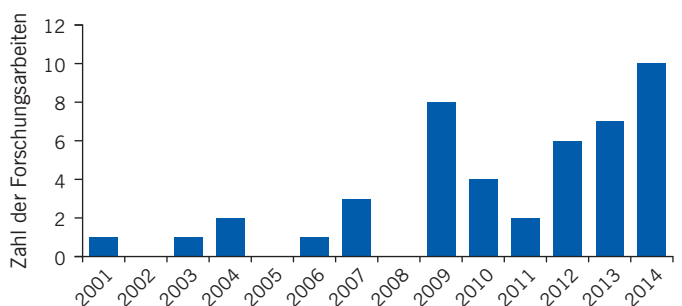
Contra

- ⊖ Die Verknüpfung von Daten kann wegen Verhandlungen mit Datenanbietern kostspielig und zeitaufwendig sein.
- ⊖ Datenschutzvorschriften sowie die Notwendigkeit der Anonymisierung von Daten schränken Datenzugriff und -inhalte ein.
- ⊖ Die Notwendigkeit, Einwilligungen zur Nutzung verknüpfter Daten einzuholen, kann Zustimmungsverzerrungen erzeugen und auch durch verringerte Antwortraten Selektionsprobleme hervorrufen.
- ⊖ Zuverlässige Verknüpfungen erfordern eine jeweils eindeutige Kennung; ohne solche wird die Verknüpfung mühsam und die Matching-Qualität sinkt.

KERNBOTSCHAFT DES AUTORS

Die Datenverknüpfung eröffnet neue Forschungsmöglichkeiten, indem zuverlässig Verwaltungsdatensätze mit detaillierten Umfragedaten kombiniert werden. Sicherergestellt sein muss, dass beide Datenquellen eindeutige persönliche Kennungen enthalten und die Umfrage eine ordnungsgemäß formulierte Einwilligungserklärung für die Auskunftspersonen enthält. Bedenken zu Datensicherheit auf Seiten des Datenanbieters (in der Regel einer staatlichen Institution) müssen im Vorfeld entkräftet werden, um zu verhindern, dass sie zu Zugangsbeschränkungen oder Forderungen nach strikter Anonymisierung der Daten führen und so das Forschungspotenzial verringern.

Die Verwendung von verknüpften Verwaltungs- und Erhebungsdaten nimmt in der Arbeitsmarktforschung deutlich zu



Quelle: Berechnungen des Autors basierend auf einer Auswertung der IZA Discussion Paper Series.