

Projektbericht RWI-GEO-Connect

Philipp Breidenbach, Inga Heinze,
Philip Raatz, Sandra Schaffner



Dezember 2023

RWI-GEO-Connect

Philipp Breidenbach¹ Inga Heinze¹, Philip Raatz¹, Sandra Schaffner¹

Dezember 2023

¹ RWI Leibniz-Institut für Wirtschaftsforschung

Abstract

Im vorliegenden Kurzbericht zum RWI-GEO-Connect Projekt, das im Rahmen des Konsortiums KonsortSWD der Nationalen Forschungsdaten Infrastruktur (NFDI) durchgeführt wurde, werden die zentralen Projektergebnisse präsentiert und ein Überblick über den Projektverlauf gegeben. Das primäre Ziel des Projekts bestand darin, Datensätze anderer Forschungsdatenzentren zu identifizieren, die sich potenziell mit den RWI Flagship Datensätzen RWI-GEO-GRID und RWI-GEO-RED verknüpfen lassen. Zudem wurden entsprechende Verknüpfungen angebahnt, während gleichzeitig Herausforderungen und Schwierigkeiten bei diesem Prozess erfasst wurden.

Keywords: gemeinsamer Datenzugang, Datenverknüpfung, ECONOMICS, RWI-GEO-GRID, RWI-GEO-RED

Inhaltsverzeichnis

1. Einführung.....	3
2. Projektbeschreibung und -verlauf.....	4
2.1. Exploration des Verknüpfungspotenzials kleinräumiger Datensätze	4
2.1.1. Erster Austausch zur Verknüpfung von Daten.....	4
2.1.2. Konkretisierung des weiteren Vorgehens zur Kooperation.....	5
2.2. Evaluierung der Gesprächsrunden	5
2.3. Skizzierung eines konkreten Fahrplans für alle Kooperationspartner	6
3. Ergebnisse	7

1. Einführung

Empirische Analysen mit kleinräumigen Daten gewinnen in der ökonomischen und sozialwissenschaftlichen Forschung zusehends an Bedeutung. Die zunehmende Verfügbarkeit von qualitativ hochwertigen Forschungsdaten auf kleiner geografischer Ebene bietet vielfältige Potentiale. Insbesondere in der Verknüpfung mit Individual- und Befragungsdaten bieten kleinräumige Kontextinformationen einen deutlichen Mehrwert für innovative Forschungsthemen. Allerdings stehen administrative Daten in Deutschland bisher vor allem auf Ebene der Gemeinden oder Kreise zur Verfügung. Kreise und Gemeinden können jedoch lokale Gegebenheiten innerhalb von Nachbarschaften nicht immer adäquat beschreiben. Zudem variieren sie stark in Bezug auf ihre Größe, und infolge vielfacher Gebietsreformen ebenfalls über die Zeit. Im Gegensatz dazu gewährleisten Daten auf Rasterebene eine vergleichbare Darstellung über Raum und Zeit hinweg, und ermöglichen präzise Analysen auf kleinstem Raum. So rückt die Verknüpfung von Datensätzen über unterschiedliche regionale Ebenen zunehmend in den Fokus, insbesondere die Verknüpfung auf kleinräumiger Ebene (z.B. 1x1 km²). Diese ermöglicht der Wissenschaft neuen innovativen Forschungsideen nachzugehen und bereits bekannte Datensätze aus einer neuen Perspektive kennenzulernen.

Forschende stehen dabei vor verschiedenen Herausforderungen. Erstens mangelt es an einer vollständigen Übersicht über existierende kleinräumige Datensätze sowie einer Darstellung der verfügbaren Geo-Ebenen. Zweitens unterliegen kleinräumige Rasterdaten zum Schutz sensibler Information häufig strengen Auflagen des Datenschutzes. Dadurch lassen sich vorhandene Datensätze häufig nur eingeschränkt kombinieren. Ziel des Projektes RWI-GEO-Connect war es, mittels der Verknüpfung von Daten auf kleinräumiger GEO-Ebene neue Perspektiven und Analysepotenziale für die Wissenschaft anzubahnen und aufzuzeigen sowie etwaige Schwierigkeiten und Herausforderungen zu identifizieren.

Grundlage für die Verknüpfung sind die Datensätze RWI-GEO-GRID¹ und RWI-GEO-RED² des Forschungsdatenzentrum Ruhr (FDZ Ruhr) am RWI – Leibniz-Institut für Wirtschaftsforschung. Die Datensätze charakterisieren sich durch eine kleinräumige Verfügbarkeit auf Grid-Ebene. Der RWI-GEO-GRID Datensatz beruht auf Daten der microm GmbH und bietet Informationen zu sozioökonomischen Merkmalen auf Grid-Ebene. Diese umfassen Angaben zur Anzahl sowie Zusammensetzung von Haushalten und Gebäuden, zur Bevölkerung und Altersstruktur nach Geschlecht, zum Migrationshintergrund, ökonomischen Faktoren wie der Arbeitslosenquote, Anteilen der PKW-Marken, -Segmenten und Antriebsarten, zum Kreditausfall und der Kaufkraft, sowie Bevölkerungsbewegungen und gesellschaftlichen Milieus. Beobachtungsgegenstand ist hierbei das einzelne 1X1 km Grid. Der RWI-GEO-RED Datensatz beruht auf Daten aus einer Kooperation mit ImmobilienScout24. Diese umfassen einen Großteil der Informationen zu den Immobilieninseraten, die auf der Webseite des Anbieters ImmobilienScout24 hinterlegt sind,

¹ RWI-GEO-GRID: Socio-Economic Data on grid level. (n.d.). RWI Essen. <https://www.rwi-essen.de/forschung-beratung/weitere/forschungsdatenzentrum-ruhr/datenangebot/rwi-geo-grid-socio-economic-data-on-grid-level>

² RWI-GEO-RED/X: Real estate data and price indices. (n.d.). RWI Essen. <https://www.rwi-essen.de/forschung-beratung/weitere/forschungsdatenzentrum-ruhr/datenangebot/rwi-geo-red-real-estate-data>

ergänzt um Informationen zur Adresse der inserierten Immobilie, zur Laufzeit der Anzeige, zur Anzahl der Klicks, sowie Angaben zur Art des Anbieters. Auf Grundlage dieser Datenbasis stellt das FDZ Ruhr Immobilienpreisindizes unter dem Namen RWI-GEO-REDX auf kleinräumiger Basis zur Verfügung. Im Rahmen des Projekts RWI-GEO-Connect wurde das Potential der Verknüpfung der Datensätze des FDZ Ruhr mit Datensätzen anderer akkreditierter Forschungsdatenzentren analysiert und eine Grundlage für die konkrete Umsetzung von Kooperationen zur Verknüpfung mit anderen kleinräumigen Daten geschaffen.

Im Weiteren wird auf das Projektvorgehen und die Projektziele eingegangen.

2. Projektbeschreibung und -verlauf

Das Vorgehen des Projektes verlief in mehreren Projektphasen. In einem ersten Schritt wurden alle beim Rat für Sozial- und Wirtschaftsdaten (RatSWD)³ akkreditierten Forschungsdatenzentren (FDZen) kontaktiert und zur Teilnahme an einer Onlinebefragung eingeladen. Ziel der Befragung war es, einen Überblick über die Existenz von kleinräumigen Daten in den FDZen zu schaffen und Anknüpfungspunkte für Kooperationen zu finden. Im zweiten Schritt führte das FDZ Ruhr Gespräche mit den FDZen, die über entsprechende Daten verfügen. Ergaben sich aus der ersten Gesprächsrunde Anknüpfungspunkte für Kooperationen, führte das FDZ Ruhr im dritten Schritt eine weitere Gesprächsrunde zur Vereinbarung eines Vorgehens zur konkreten Verknüpfung der verschiedenen Datensätze.

Zum Auftakt des Projektes stellte das Projektteam des FDZ Ruhr die grundlegenden Projektziele und die geplanten drei Arbeitsschritte im Rahmen der 28. Sitzung des Ausschusses für Forschungsdateninfrastrukturen (FDI-Ausschuss)⁴ vor.

2.1. Exploration des Verknüpfungspotenzials kleinräumiger Datensätze

In der Onlineumfrage wurden georeferenzierte Datensätze der FDZen abgefragt, die die Grundlage für die möglichen Kooperationspotenzialen darstellen. Die Onlineumfrage stand allen FDI-Ausschussmitgliedern in der Feldzeit (Juni 2023) zur Beantwortung offen. Mit insgesamt 30 teilnehmenden FDZen konnte ein sehr positiver Rücklauf von 71% erzielt werden. Aus den Antworten ergab sich für zwölf FDZen eine potenzielle Datenverknüpfbarkeit. Für die Auswahl war neben der Identifikation relevanter georeferenzierter Datensätze und dem Kooperationsinteresse vor allem die abgefragten inhaltlichen und technischen Möglichkeiten der Datenverknüpfung ausschlaggebend. Diese zwölf FDZen wurden im Anschluss an die Umfrage zu den bilateralen Gesprächen im zweiten Schritt eingeladen.

2.1.1. Erster Austausch zur Verknüpfung von Daten

Im Rahmen dieser Auftaktgespräche im zweiten Schritt wurden die Datensätze detailliert vorgestellt sowie Verknüpfungspotenziale inhaltlicher Art sondiert. Im Fokus stand dabei

³RatSWD - KonsortSWD. (2023, September 26). KonsortSWD. <https://www.konsortswd.de/ueber-uns/ratswd/>

⁴ Ständiger Ausschuss Forschungsdateninfrastruktur (FDI Ausschuss) - KonsortSWD. (2023, Oktober 24). KonsortSWD. <https://www.konsortswd.de/ueber-uns/fdi-ausschuss/>

zusätzliche und detaillierter Informationen zu relevanten Datensätzen sowie Mehrwerte einer Datenverknüpfung zu sammeln und zu diskutieren. Zur strukturierten Erfassung der Informationen wurde von Seiten des FDZ Ruhr ein Gesprächsleitfaden erstellt, der die gezielte Abfrage wichtiger Verknüpfungscharakteristika der Datensätze, sowie die Identifikation offenkundiger Hürden und Herausforderungen ermöglichen sollte. Besonderes Augenmerk lag dabei auf der Perspektive der Forschenden. Unter diesem Aspekt wurden folgende Fragestellungen diskutiert:

- Ergibt eine Verknüpfung aus der Perspektive der Forschenden Sinn?
- Welche möglichen Forschungsfragen könnten mit einer Datenverknüpfung beantwortet werden?
- Welche Datensätze eignen sich inhaltlich und technisch für eine Verknüpfung?
- Wie sind der Datenzugang und die rechtlichen Möglichkeiten geregelt. Wie könnte ein Datenzugang ermöglicht werden?
- Welche Hindernisse und Hürden könnten sich ergeben?
- Welche weiteren Ansprechpersonen sollten bei einem Nachfolgegespräch dabei sein?

Aus den Gesprächen wurde jeweils ein Protokoll verfasst, das unter anderem das weitere Vorgehen sowie – wenn beiderseitig als sinnvoll erachtet – die Vereinbarung eines zweiten Gespräches beinhaltet.

2.1.2. Konkretisierung des weiteren Vorgehens zur Kooperation

Im Rahmen der Folgegespräche im dritten Schritt des Projektes galt es über die identifizierten Verknüpfungspotenziale hinaus, konkrete Anknüpfungspunkte zu definieren und konkrete Hürden und Herausforderungen zu diskutieren. Hierbei lag ein besonderer Fokus auf inhaltlichen, technischen und rechtlichen Herausforderungen, z.B. welche Datencharakteristika der RWI-GEO Datensätze für eine Verknüpfung bereitgestellt werden sollen, auf welcher GEO-Ebene die Verknüpfung erfolgt und welche vertraglichen Absprachen für die Kooperationen gegebenenfalls notwendig sind. Ergebnis der zweiten Gesprächsrunde war ein konkretes Vorgehen für die Anbahnung der Kooperationen zwischen den FDZen. Für die zweite Runde der Gespräche wurde jeweils eine Stunde Gesprächszeit angesetzt. Insgesamt wurde mit den folgenden FDZen ein zweites Gespräch geführt und ein konkretes Vorgehen vereinbart⁵: GESIS Data Secure Center (FDZ Wahlen und FDZ Allbus), BIBB-FDZ, FDZ KBA, FDZ-LifBi und FDZ-DZA.

2.2. Evaluierung der Gesprächsrunden

Auf Grundlage der geführten Gesprächsrunden erfolgt eine abschließende Konsolidierung der Gesprächsergebnisse zur Identifizierung von Synergien und Schnittmengen. Als Ergebnis konnte der Mehrheit der FDZen (im Folgenden Kooperationspartner) ein einheitliches Vorgehen übermittelt werden.

⁵ Mit dem FDZ SOEP wurden bereits im Vorfeld zum RWI-GEO-Connect Projekt Gespräche für eine Datenverknüpfung geführt, die im Rahmen dieses Projektes fortgeführt und intensiviert wurden.

2.3. Skizzierung eines konkreten Fahrplans für alle Kooperationspartner

Nach Abschluss der zweiten Runde der Gespräche hat das FDZ Ruhr folgendes Konzept erstellt, das den Kooperationspartnern vorgestellt wird:

- Bereitstellung eines „mixed“ Immobilien Preis-Indizes, basierend auf den RWI-GEO-RED Daten.
- Die Berechnung erfolgt auf Basis der Preise für Hauskäufe, Wohnungskäufe sowie Wohnungsmieten (gewichtet nach den Beobachtungen je GEO-Ebene) und wird mit der neuen Welle des Datensatzes RWI-GEO-RED im Januar 2024 durchgeführt.
- Der Preisindex wird auf 1km² Grid-Ebene berechnet und mittels einer GEO-Bridge auf „höhere“ Geo-Ebenen hochgerechnet, sodass das Datenprodukt auf Grid-, Gemeinde- und Kreisebene verfügbar ist.
- Für die Berechnung auf höhere GEO-Ebenen wird der Gebietsstand von 2019 verwendet
- Für das neue Datenprodukt wird eine Datensatzbeschreibung angefertigt, die auch das Berechnungsverfahren des Preisindex beinhaltet.
- Für die Nutzung dieses neuen Datenproduktes bedarf es keines Datennutzungsvertrags. Die Erfassung der Datennutzenden erfolgt über das kooperierende FDZ.
- Für die Bewerbung des neuen Datenproduktes und der Verschneidungsmöglichkeit wird ein Mustertext für die jeweilige Homepage entwickelt.

Ein abweichendes Vorgehen ergibt sich für die FDZen FDZ SOEP und FDZ KBA. Hier wurde folgender Fahrplan abgestimmt:

- Das FDZ KBA interessiert sich an einer Verschneidungsmöglichkeit mit dem RWI-GEO-GRID Daten, um das Zentrale Fahrzeugregister (ZFZR) mit sozioökonomischen Aspekten zu ergänzen
 - o Daten des RWI-GEO-GRID unterliegen Nutzungsverträgen zwischen RWI und Forschenden. Die Datennutzende, die das verknüpfte Datenprodukt nutzen wollen, müssen getrennte Nutzungsverträge mit dem FDZ KBA und dem FDZ Ruhr am RWI abschließen.
- Mit dem FDZ SOEP gab es bereits im Vorfeld des RWI-GEO-Connect Projektes einen Austausch zur Verschneidung der jeweiligen Datenbestände. Die Diskussionen und Abstimmung sind zwischen den beiden Einrichtungen bereits fortgeschrittener. Eine entsprechende Verschneidung wird hier über einen Remote Access Zugang auf das SOEP angestrebt. Hierzu erfolgten technische und rechtliche Abstimmung während der Projektlaufzeit

Die nachfolgende Matrix stellte die Verknüpfung der Datenbestände der FDZen dar:

Datensätze	Zentrale Fahrzeugregister (ZFZR)	Der Deutsche Alterssurvey (DEAS)	BIBB/BAuA - Erwerbstätigenbefragung 2018	GESIS Data Secure Center	Sozio-oekonomischen Panels	Nationales Bildungspanel (NEPS)
RWI-GEO-GRID	X		X		X (Einrichtung Remote Access)	
RWI-GEO-RED/REDX		X	X	X	X (Einrichtung Remote Access)	X

Als weiteres interessantes Datenprodukt soll perspektivisch auch der vom FDZ Ruhr entwickelte Amenity/Disamenity Index für eine Verschneidung angeboten werden. Es gehen aktuell Informationen zu kleinräumiger Belastung durch Feinstaubverschmutzung, Lärm, Hitzetage, Ozon und fehlende Grünflächen in den Index ein. Die Daten liegen auf der 1x1 km Ebene (INSPIRE Grids) vor und können über eine Aggregation auf höhere Gebietseinheiten angepasst werden.

Die hier dargestellten Datenverknüpfungen bzw. Kooperationen sollen in einer Kooperationsvereinbarung zwischen den beteiligten FDZen festgehalten werden. Diese umfasst im Wesentlichen die administrativen Abstimmungen zwischen den beteiligten FDZen bezüglich der Erfassung der Datennutzenden, Zitierung der Datensätze und Kommunikationswege zwischen den FDZen. Hinsichtlich der Kommunikation mit den Datennutzenden werden die Verschneidungsmöglichkeiten auf den jeweiligen Webseiten der FDZen beworben. Zudem ist eine Darstellung der Verschneidungsmöglichkeit über das Forum4Mica⁶ ebenfalls angedacht.

3. Ergebnisse

Das RWI-GEO-Connect Projekt konnte erfolgreich die Verknüpfung der Datensätze RWI-GEO-GRID und RWI-GEO-RED/X mit georeferenzierten Datensätzen anderer FDZen forcieren und konkrete Kooperationen anbahnen. Das Projekt legt den Grundstein für den zeitnahen Zugang zu verknüpften Datensätzen zwischen Daten anderer FDZen und ausgewählten Scientific Use Files des FDZ Ruhr. Die Nutzung wird Forschenden über getrennte Nutzungsverträge zumeist in den Datensicherheitsräumen der Kooperationspartner ermöglicht.⁷ Dies ermöglicht die reibungslose und nutzerorientierte Verknüpfung der verschiedenen Datensätze und eröffnet neue spannende Analysepotenziale. Über den ablaufenden Projektzeitraum hinaus besteht das Ziel, die angebahnten Kooperationen im Laufe des ersten Halbjahres 2024 umzusetzen.

Im Laufe des Projekts ließen sich einige Herausforderungen bei der Verknüpfung von Datensätzen identifizieren. Folgendes Vorgehen hat sich bewährt und bietet Orientierung für verwandte Vorhaben. In einem ersten Schritt sollte skizziert werden, welchen Mehrwert die Verschneidung konkreter Datenprodukte mit sich bringt. Entsprechende Verknüpfungen sind vor allem sinnvoll, wenn sie bereits bestehende Datenbestände um neue Angebote erweitern oder bereits existierende Daten in einer höheren Qualität abbilden. Für die Verknüpfung auf der Grundlage räumlicher Variablen ist zu beachten, dass diese auf gleicher Ebene vorliegen oder auf dieser bereitgestellt werden können.

Im Falle von sensiblen Raster- oder Adressdaten ist neben dem zugrunde liegenden System zur Georeferenzierung außerdem zu klären, ob die Geodaten beiden Datenzentren vorliegen

⁶ Forum4MICA. (n.d.). Forum4MICA. <https://forum.lifbi.de/>

⁷ Die vom FDZ Ruhr bereitgestellten SUFs sind nicht auf die Nutzung in Datensicherheitsräumen beschränkt, sodass sie ohne datenschutzrechtliche Hürden gesichert und zugangsbeschränkt in Datensicherheitsräumen der Kooperationspartner genutzt werden können.

dürfen und unter welchen Auflagen diese zur Weitergabe an Kooperationspartner berechtigt sind. In Bezug auf administrative Grenzen in Deutschland sind zudem Änderungen im Zuge von Gebietsreformen zu berücksichtigen. Absprachen zu Formalia und dem konkreten Ablauf der Datennutzung sind zwischen den Forschungsdatenzentren in Kooperationsvereinbarungen oder in gesonderten Vertragswerken festzuhalten. Eine Vereinheitlichung des Vertragswerkes zur Datennutzung erscheint bei bilateralen Verknüpfungen nicht sinnvoll. Der administrative Aufwand bei beiden kooperierenden FDZen erweist sich als zu hoch. Als letzter Schritt vor der gemeinsamen Bereitstellung der Daten ist eine Testphase empfehlenswert. Diese soll sicherstellen, dass der Zugang zu den Daten reibungslos funktioniert. Die Kooperationspartner sollten darüber hinaus festlegen, wie sie das neue Datenangebot kommunizieren, damit aus der Kooperation auch tatsächlich ein Mehrwert für Forschende entsteht.

Impressum

Kontakt:

RWI Leibniz-Institut für Wirtschaftsforschung

Philip Raatz

Forschungsdatenmanager

Webseite: <https://www.rwi-essen.de/forschung-beratung/weitere/forschungsdatenzentrum-ruhr>

E-Mail: philip.raatz@rwi-essen.de

Essen, Dezember 2023

KonsortSWD:

KonsortSWD baut als Teil der Nationalen Forschungsdateninfrastruktur Angebote zur Unterstützung von Forschung mit Daten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften aus. Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll nutzungsorientiert ausgestaltet sein und die Bedürfnisse der Forschungscommunities berücksichtigen. Wichtiger Grundstein ist dabei das seit über zwei Jahrzehnten durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) aufgebaute Netzwerk von Forschungsdatenzentren.

KonsortSWD wird im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert – Projektnummer: 442494171.



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:
<https://creativecommons.org/licenses/by/4.0/>

DOI: 10.5281/zenodo.10473630

Zitationsvorschlag:

Breidenbach, P., Heinze, I., Raatz, P. & Schaffner, S. (2023). RWI-GEO-Connect. Kurzprojekt. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD). <https://doi.org/10.5281/zenodo.10473630>