Journal für Mobilität und Verkehr

ISSN 2628-4154

www.dvwg.de

Generieren, verarbeiten, veröffentlichen: Einheitlicher Umgang mit Mobilitätsdaten in interdisziplinären Forschungsclustern

Svetlana Zubareva*, Louis Stille-Hönig*, Markus Lienkamp

Siehe Autor*innenangaben

Abstract

Mit der zunehmenden Verfügbarkeit digitaler Mobilitätsdaten eröffnen sich neue Möglichkeiten für die empirische Verkehrsforschung und die evidenzbasierte Verkehrsplanung. Gleichzeitig wächst der Bedarf an transparent dokumentierten, qualitativ hochwertigen und datenschutzkonformen Dateninfrastrukturen, die heterogene Datenströme langfristig nutzbar machen. Dieser Beitrag zeigt am Beispiel des Münchner Forschungsclusters für die Zukunft der Mobilität (MCube) die Vorteile einer gemeinsamen Datenstruktur zwischen den beteiligten Projekten. Anhand von digitalen Mobilitätstagebüchern, in denen die zurückgelegten Wege erfasst werden, sowie Carsharing-Flottendaten aus unterschiedlichen Forschungsprojekten wird die zentrale Verarbeitung von Mobilitätsdaten beschrieben. Dabei wird ein mehrstufiger Bereinigungsprozess zur Harmonisierung von Formaten, zum Umgang mit fehlenden Werten sowie zur Validierung der Datensätze anhand bekannter Mobilitätsmuster dargelegt. Auch der Workflow zur Anonymisierung, der den Schutz der Teilnehmenden gewährleistet und gleichzeitig den analytischen Nutzen für Querschnitts- und Längsschnittanalysen bewahrt, wird erläutert. Der Beitrag liefert ein praxisnahes Vorgehensmodell für den Aufbau und den Betrieb einer Mobilitätsdatenbank und regt dazu an, die Verarbeitung von Mobilitätsdaten innerhalb von Forschungsclustern zu zentralisieren. Die Erkenntnisse und Methoden des Beitrags richten sich an Verkehrsforschende, Dateninfrastrukturbetreibende und Verkehrsplanende. Die dargelegten Prozesse erleichtern die Veröffentlichung von Daten sowohl für wissenschaftliche Publikationen als auch für die Bürger*innenbeteiligung und die Öffentlichkeitsarbeit.

Schlagwörter:

Datenmanagement, Mobilität, Interdisziplinarität, Forschungscluster, Datenschutz

1 Einleitung

Durch die zunehmende Verfügbarkeit digitaler Mobilitätsdaten ergeben sich stetig neue Möglichkeiten der empirischen Verkehrsforschung. Die Daten werden dabei durch unterschiedlichste Geräte wie Smartphones, Sensoren und Fahrzeuge erfasst und lassen eine detaillierte Untersuchung des Mobilitätsverhaltens zu. Allerdings ist dies nur möglich, wenn die unterschiedlichen Daten harmonisiert und kombiniert bereitgestellt werden. Denn so können Verkehrsangebote und deren Nutzung räumlich sowie zeitlich differenziert analysiert werden. Die gewonnenen Erkenntnisse können zudem als Grundlage für eine evidenzbasierte Verkehrsplanung dienen (Waldorf et al., 2026).

Die Zusammenführung dieser Daten ist allerdings mit verschiedenen Herausforderungen verbunden, da die Quellen sich hinsichtlich ihrer Formate, zeitlichen Auflösung, räumlichen Genauigkeit und inhaltlichen Struktur unterscheiden. Weiterhin wird die Datenverarbeitung häufig für jedes Forschungsprojekt neu aufgesetzt und durchgeführt. So entstehen voneinander getrennte, heterogene Datenbestände und projektspezifische Verarbeitungsschritte, welche eine spätere Zusammenführung

*korrespondierende Autor*innen

und Wiederverwendung erschweren. Dies ist insbesondere dann problematisch, wenn verschiedene Forschungsprojekte einem Cluster angehören und somit auf ein übergeordnetes Ziel hinarbeiten.

Darüber hinaus enthalten Mobilitätsdaten häufig einen direkten oder indirekten Personenbezug und müssen dementsprechend nach hohen Datenschutz-Standards verarbeitet werden. Neben der Anonymisierung ist eine nachvollziehbare Dokumentation der Verarbeitung erforderlich. Durch eine zentrale Dateninfrastruktur lassen sich einheitliche Qualitätsanforderungen umsetzen und Daten projektübergreifend verfügbar machen.

Der vorliegende Beitrag untersucht daher, wie heterogene Mobilitätsdaten innerhalb eines interdisziplinären Forschungsclusters zentral bereinigt, anonymisiert und bereitgestellt werden können und welche Vorteile sich daraus in der Praxis ergeben. Hierzu wird ein praxisnahes Vorgehensmodell für den Aufbau und Betrieb einer clusterinternen Mobilitätsdatenbank vorgestellt. Die Umsetzung wird anhand digitaler Mobilitätstagebücher und Carsharing-Flottendaten erläutert. Abschließend werden die Vorteile und Grenzen der zentralen Datenverarbeitung diskutiert. Zudem wird untersucht, inwiefern sich das Vorgehensmodell auf andere Forschungscluster und Organisationen übertragen lässt.

2 Hintergrund und Stand der Forschung

Die systematische Erfassung, Aufbereitung und Bereitstellung von Mobilitätsdaten ist ein zentrales Thema der aktuellen Verkehrsforschung (Bellini et al., 2024).

Auf nationaler Ebene verfolgt die Nationale Forschungsdateninfrastruktur (NFDI) das Ziel, Forschungsdaten disziplinenübergreifend zu sichern und zugänglich zu machen (DFG, o. J.). Eine auf den Verkehrsbereich zugeschnittene Infrastruktur fehlt in diesem Rahmen bislang, wodurch Cluster bei der Datenverarbeitung kaum auf übergeordnete Standards zurückgreifen können. Ohne diesen Bezug entwickeln sich Formate, Metadaten und Aufbereitungsregeln in jedem Projekt eigenständig, was eine spätere Zusammenführung erschwert.

Der Mobility Data Space (MDS, ehemals mobiDS), entstanden im mFUND-Förderprogramm des Bundesverkehrsministeriums, verfolgt einen anderen Ansatz. Er baut auf der Referenzarchitektur der International Data Spaces auf und regelt den kontrollierten Austausch von Daten zwischen Kommunen, Unternehmen und Dienstleistenden unter Wahrung der Datensouveränität (Fraunhofer IVI, 2022). Im Zentrum steht damit das Teilen und Aufbereiten bereits vorhandener Daten zwischen verschiedenen Akteuren, nicht die interne Harmonisierung und mehrstufige Bereinigung heterogener Rohdaten innerhalb eines Forschungsclusters.

Ein Beispiel für eine langfristig etablierte, domänenspezifische Erhebung ist die German In-Depth Accident Study (GIDAS), die seit 1999 in Hannover und Dresden sowie seit 2023 auch in München detaillierte Unfalldaten erfasst und für die Wissenschaft aufbereitet (GIDAS, 2024). GIDAS zeigt, wie konsequente, über Jahre hinweg konsistente Datenerhebung in einer klar abgegrenzten Domäne methodisch gelingen kann. Der Ansatz ist jedoch auf Unfalldaten beschränkt und nicht darauf angelegt, unterschiedliche Mobilitätsdatenquellen zusammenzuführen. Andere Mobilitätsdaten, etwa Mobilitätstagebücher oder Flottendaten, folgen anderen Erhebungslogiken und lassen sich nicht ohne Weiteres an ein solches Schema anbinden.

Die genannten Ansätze verdeutlichen drei wiederkehrende Herausforderungen. Erstens sind die Quellen heterogen, was Formate, Auflösung und Semantik betrifft. Zweitens muss die Verarbeitung personenbezogener Mobilitätsdaten datenschutzkonform erfolgen. Drittens fehlt es an einheitlichen Standards für die Erhebung und Aufbereitung. Damit bleibt offen, wie eine Forschungsinfrastruktur konkret aussehen muss, die heterogene Mobilitätsdaten innerhalb eines interdisziplinären Forschungsclusters zentralisiert, nachnutzbar macht und dabei reproduzierbar dokumentiert. Gegenüber den skizzierten Initiativen verbindet das vorgestellte Modell übergeordnete Infrastrukturkonzepte mit der konkreten, reproduzierbaren Datenverarbeitung in der Verkehrs- und Mobilitätsforschung.

3 Daten im Kontext des Forschungsclusters MCube

Das Münchner Cluster für die Zukunft der Mobilität (MCube) hat das Ziel, Mobilitätsinnovationen zu entwickeln und in die Umsetzung zu bringen. Das Vorhaben basiert auf der Initiative „Clusters4Future", welche von dem Bundesministerium für Forschung, Technologie und Raumfahrt getragen wird.



Abbildung 1: Übersicht der MCube Projekte und Standorte in den Phasen 1 und 2 (MCube, o. J.)

In diesem Zusammenhang werden über neun Jahre bis zu 45 Millionen Euro Fördermittel in drei Umsetzungsphasen bereitgestellt (BMFTR, o. J.-a). Im Jahr 2021 ist MCube mit 14 Forschungsprojekten in die erste Umsetzungsphase gestartet, welche im November 2024 durch die zweite Umsetzungsphase mit neun Forschungsprojekten abgelöst wurde. Die Projekte haben das übergeordnete Ziel, die Qualität von Luft, Raum und Zeit zu verbessern, und bauen über die Projektphasen hinweg aufeinander auf (Clusterplattform Deutschland, o. J.; MCube, o. J.). Die Forschungsprojekte sowie deren Standorte werden in Abbildung 1 dargestellt.

In München werden in diesem Zusammenhang von Carsharing über Verkehrssimulationen bis hin zu Beteiligungs- und Partizipationsverfahren verschiedenste Facetten der Mobilität berücksichtigt. Das ist eine Chance für die interdisziplinäre Entwicklung von Innovationen, bringt aber auch Herausforderungen, wie das Management heterogener Daten, mit sich. MCube zeichnet sich dabei durch seinen Anwendungsbezug aus (MCube, o. J.). Forschungs- und Innovationsansätze wie die Incentivierung von klimafreundlicher Mobilität oder das Teilen privater Fahrzeuge werden in den Projektphasen umgesetzt und datenbasiert evaluiert (BMFTR, o. J.-b). Dieser Fokus auf lokale Umsetzungs- und Transformationsprozesse wird auch im wissenschaftlichen Diskurs als wichtiges Instrument für das Schaffen und Verstetigen von Innovationen beschrieben (Engels et al., 2019).

Die in der ersten Umsetzungsphase gewonnenen Daten wurden bereits im Rahmen von wissenschaftlichen Arbeiten untersucht und veröffentlicht (Loder et al., 2024; Nachtigall et al., 2025; Waldorf et al., 2026). Die Daten wurden in diesem Zusammenhang clusterübergreifend verarbeitet und für die Veröffentlichung aufbereitet. Dies wird auch in der zweiten Umsetzungsphase fortgeführt und soll in diesem Artikel am Beispiel der Projekte „Komfortables und effizientes Fahrzeug-Sharing auf Privatgrund (ComfficientShare 2.0)" und „Reallabor Future Mobility Campus Garching (ReFuMoLab_Garching)" näher dargestellt werden. In dem Forschungsprojekt ComfficientShare 2.0 werden Carsharing-Fahrten mithilfe von GPS-Loggern und Buchungsdaten aus einer Carsharing-App erfasst. Darüber hinaus werden mithilfe der TUM-Mobility-App digitale Mobilitätstagebücher der Nutzenden erstellt, in denen die zurückgelegten Wege fortlaufend mit Zeitpunkt, Länge und Verkehrsmittel erfasst werden. Die Kombination dieser Datensätze ermöglicht es, die Nutzung des Carsharings im Zusammenhang mit der Alltagsmobilität zu untersuchen. ReFuMoLab_Garching stellt den Projektteilnehmenden ebenfalls die TUM-Mobility-App zur Erstellung digitaler Mobilitätstagebücher zur Verfügung (MCube, o. J.).

4 Aufbau und Betrieb der Mobilitätsdatenbank

Bei der Mobilitätsdatenbank handelt es sich um einen selbst gehosteten PostgreSQL-Server mit der räumlichen Erweiterung PostGIS (PostgreSQL Global Development Group, 2024; PostGIS Project Steering Committee, 2024). Dieser ist modular über projektspezifische Schemata erweiterbar und wird durch einen zentralen Admin verwaltet.

4.1 Datenintegration und Harmonisierung

Die erhobenen Daten liegen in unterschiedlichen Formaten vor und werden über verschiedene Schnittstellen bereitgestellt (Tabelle 1). Ziel der Integration ist es, diese heterogenen Quellen in einer einheitlichen Datenstruktur abzubilden, in der jede Fahrt (Trip) als harmonisierter Datensatz mit Geometrie, Zeit und Länge vorliegt.

Tabelle 1: Daten, Schnittstellen und Formate

Daten	Schnitt-stelle	Inhalt
Carsharing Buchungen	Betreiber-plattform	Start / Ende / User-ID / Fahrzeug / Rolle
Fahrzeug-positionen	Betreiber-plattform	≈ 1 Position/Minute; Position / Uhrzeit / Fahrzeug
Fahrzeug-positionen	GPS-Logger	1 Hz; Position / Uhrzeit / Fahrzeug
Mobilitätstage-buch – Nutzende	TUM-Mobility-App-API	User-ID / E-Mail
Mobilitätstage-buch – Fahrten	TUM-Mobility-App-API	Startzeit / Endzeit / Länge / Verkehrsmittel
Online-Umfrage	Lime Survey	E-Mail / Soziodemographische Daten

Den Ausgangspunkt der Datenerhebung durch das Mobilitätstagebuch bildet eine Online-Umfrage über Lime Survey. Dort melden sich die Teilnehmenden an und erhalten anschließend die Zugangsdaten für die TUM-Mobility-App. Neben den Kontaktdaten werden dabei soziodemographische Merkmale erhoben. Über eine pseudonyme Kennung (UUID) können diese Angaben anschließend mit den erfassten Mobilitätsmustern verknüpft werden. Hierdurch lassen sich Unterschiede im Mobilitätsverhalten verschiedener Personengruppen untersuchen. Die Datenquellen sind direkt an die zentrale Datenbank angebunden. Die technische Umsetzung der Anonymisierung und Bereitstellung wird im Abschnitt „Datenbereitstellung und Nutzung“ erläutert.

Für ComfficientShare 2.0 umfasst das Schema drei Tabellen und eine Ansicht: eine Tabelle mit den Buchungsinformationen, eine Tabelle mit den Fahrzeugpositionen aus GPS-Loggern und Betreiberplattform sowie eine Tabelle, in der Buchungen und Positionen zu Fahrten verknüpft werden. Da die hochfrequenten Positionsdaten Rückschlüsse auf Personen sowie Wohn- und Arbeitsorte zulassen, werden für Analysen ausschließlich anonymisierte Ansichten bereitgestellt. Der Datenabruf und das Laden in die Datenbank erfolgen weitgehend automatisiert über einen weiteren Server im selben Netzwerk.

Analog dazu werden die Daten der digitalen Mobilitätstagebücher täglich von einem externen Dienstleister bereitgestellt und in einem eigenen Schema gespeichert. Dieses enthält eine Tabelle mit den registrierten Nutzenden und eine Tabelle mit den Fahrten. Auch hier werden die Fahrten nach denselben Maßstäben anonymisiert und über eine Ansicht bereitgestellt.

4.2 Mehrstufiger Bereinigungsprozess

Nach der erfolgreichen Bereitstellung der Daten muss deren Qualität geprüft und sichergestellt werden. Dies gilt insbesondere für die Carsharing-Daten, da die digitalen Mobilitätstagebücher bereits in aufbereiteter Form bereitgestellt werden. Die Methodik zur Qualitätssicherung der Carsharing-Daten erfolgt in drei Schritten:

Schritt 1:

Zunächst werden unvollständige Buchungen und fehlende Positionswerte behandelt. Als Fahrt gilt nur eine abgeschlossene Buchung mit vorhandener Start- und Endzeit, wobei die Endzeit nach der Startzeit liegen muss. Zusammenhängende Positionsfolgen werden anhand von Standzeitlücken von mehr als 15 Minuten in einzelne Fahrsegmente getrennt (Schuessler & Axhausen, 2009).

Schritt 2:

Fahrten mit leerer oder ungültiger Geometrie, einer Länge unter 100 m oder einer Dauer unter zwei Minuten werden als Fehler verworfen. Die Schwellenwerte wurden anhand des vorliegenden Datensatzes sowie der Daten aus ComfficientShare 1.0 entwickelt. Da nicht alle Fahrzeuge mit einem GPS-

Logger ausgestattet sind, liegen für einzelne Fahrzeuge nur die geringer aufgelösten Positionsdaten der Betreiberplattform vor. Die Positionsdaten werden zunächst technisch bereinigt. Hierzu werden Koordinaten entfernt, deren Breitengrad außerhalb des Intervalls von −90° bis 90° oder deren Längengrad außerhalb des Intervalls von −180° bis 180° liegt (Zheng, 2015).

Schritt 3:

Die Punkte beider Quellen werden mit demselben Map-Matching-Verfahren verarbeitet: Ein auf einem Server im lokalen Netzwerk betriebener GraphHopper-Dienst ordnet die GPS-Punkte dem dort vorgehaltenen OpenStreetMap-Straßengraphen zu und rekonstruiert den Fahrtverlauf (GraphHopper GmbH, o. J.; Newson & Krumm, 2009; OpenStreetMap contributors, 2024). Fehlen vollständig verwertbare Positionen, bleibt die Buchung ohne Geometrie erhalten.

Die Ergebnisse des Bereinigungsprozesses werden in Abbildung 2 dargestellt. In der oberen Karte werden die Rohdaten dargestellt, während die untere Karte die bereinigten Trajektorien zeigt. Die räumliche Verteilung lässt erkennen, dass sich die Fahrten entlang der überregionalen Hauptverkehrsachsen konzentrieren.

Die Plausibilität der Carsharing-Fahrten wird anhand der erfassten Daten von Nachtigall et al. (2025) hinsichtlich der charakteristischen Fahrtlängen und -dauern geprüft. Weichen die beobachteten Muster deutlich von diesen Referenzen ab, so deutet dies auf verbleibende Fehler im Bereinigungsprozess hin. Dies wird gleichermaßen für die Mobilitätstagebücher mithilfe der Daten aus dem Vorgängerprojekt Mobilität.Leben durchgeführt und in Abbildung 3 exemplarisch anhand der Wegehäufigkeit im Tagesverlauf dargestellt. Es zeigt sich, dass die beobachteten Fahrtlängen und -dauern den aus großräumigen Erhebungen wie Mobilität in Deutschland bekannten Mustern entsprechen (Nobis & Kuhnimhof, 2018). Bei der Wegehäufigkeit über den Tag werden zwar Unterschiede sichtbar, doch dies lässt sich durch den Fokus auf die Pendlermobilität in der aktuellen MCube-Studie erklären.

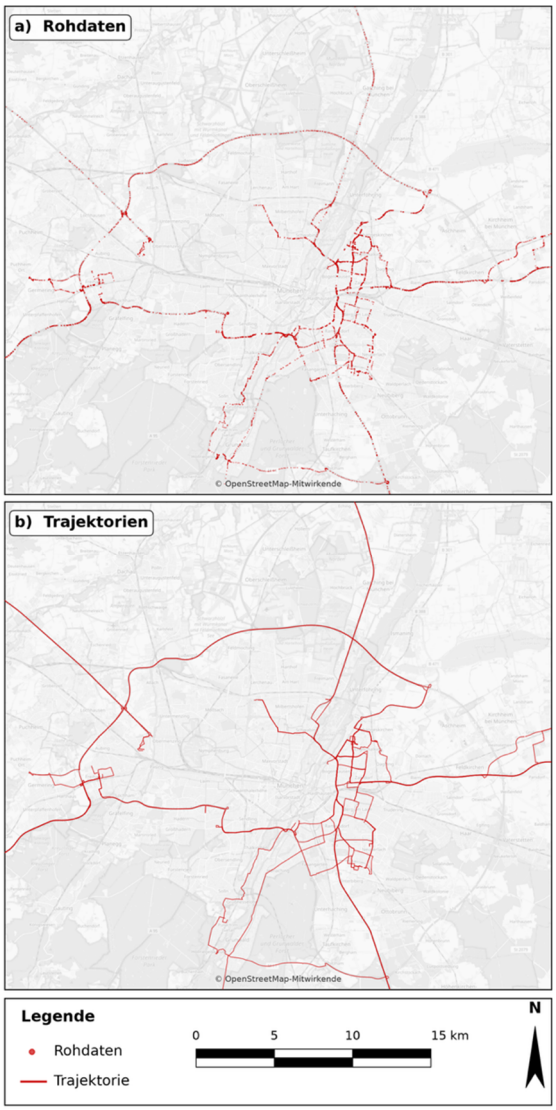


Abbildung 2: Aufbereitung der ComfficientShare 2.0 Fahrten in München

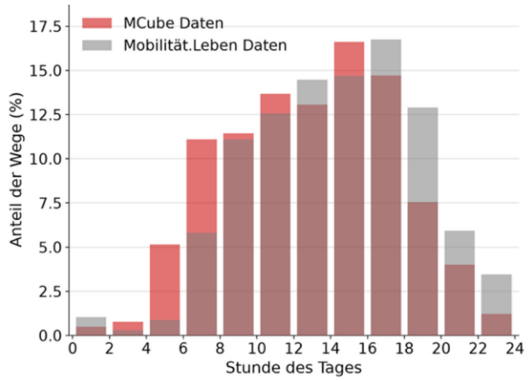


Abbildung 3: Verteilung der Wege im Tagesverlauf nach den zugrunde liegenden Datensätzen

4.3 Anonymisierung und Datenschutz

Im Rahmen des Projekts werden hochsensible Mobilitätsdaten erhoben. Daher ist eine Anonymisierung erforderlich, bevor diese für Forschungszwecke genutzt werden können. Die konkrete Ausgestaltung der Anonymisierung ist projektzielabhängig, da jedes Projekt unterschiedliche Forschungsfragen verfolgt. So erfordert beispielsweise ein Projekt zum Parkverhalten eine andere Datenaggregation als ein Projekt zur Mitarbeitendenmobilität. Um dennoch eine vergleichbare Datenbasis über alle Projekte hinweg zu gewährleisten, wurden die projektübergreifenden Anonymisierungsstufen noch vor der Datenerhebung definiert. Dadurch steht jedem Projekt dieselbe Datenmenge und -granularität zur Verfügung.

Die MCube-Datenbank setzt die Anonymisierung nicht als Einzelschritt um, sondern abgestuft, beginnend bei der Erhebung. Der gesamte Workflow der Datenaufbereitung ist in Abbildung 4 dargestellt. Grundlage ist die Pseudonymisierung. Alle Analysebestände sind über UUIDs referenziert, während direkte personenbezogene Merkmale wie E-Mail-Adressen ausschließlich in der separaten Nutzenden-Tabelle liegen. Sie ist von den analytischen Datensätzen getrennt, sodass Auswertungen ohne direkten Personenbezug erfolgen.

Das übergeordnete Ziel der Anonymisierung besteht darin, die Nachvollziehbarkeit einzelner Teilnehmender auszuschließen. Es soll verhindert werden, dass sich die Bewegungen einzelner Personen rekonstruieren lassen. Daher wurde jedem Teilnehmenden vor der Datenerhebung eine eindeutige UUID zugeordnet. Sämtliche weiteren Datenbestände lassen sich ausschließlich über diese UUID miteinander verknüpfen, ohne dass ein Rückschluss auf die jeweilige Person möglich ist.

Auf Basis dieser Pseudonymisierung werden drei Datenformate bereitgestellt. Zum einen die Umfragedaten, bei denen personenbezogene Angaben (E-Mail-Adresse sowie Vor- und Nachname) entfernt und durch die UUID ersetzt wurden. Zum anderen die GPS-Daten, die in Trajektorien und das Mikromobilitätsverhalten unterteilt werden. Trajektorien sind Datensätze mit der Geometrie einzelner Fahrten. Alle Trajektorien werden anonymisiert, indem ihre Start- und Endpunkte als Geohash der Genauigkeitsstufe 7 codiert werden (Balkić et al., 2012). Zusätzlich werden Dauer und Länge der Fahrt sowie Start- und Endzeitpunkt, jeweils auf 15 Minuten gerundet, bereitgestellt. Durch die räumliche und zeitliche Aggregation werden Rückschlüsse auf einzelne Personen verhindert.

Das Mikromobilitätsverhalten umfasst im Gegensatz zu den Trajektorien einzelne GPS-Punkte innerhalb eines definierten Radius. Diese liegen somit als Punktdaten und nicht als trajektorienbasierte Verläufe vor. Diese Daten sind ausschließlich für Teilnehmende innerhalb des jeweils betrachteten Projekts relevant. Die Nutzung ist somit projektspezifisch und auf den entsprechenden Kreis der Teilnehmenden beschränkt.

4.4 Datenbereitstellung und Nutzung

Abschließend werden die aufbereiteten Daten zentral vorgehalten und projektbezogen für die Nut-

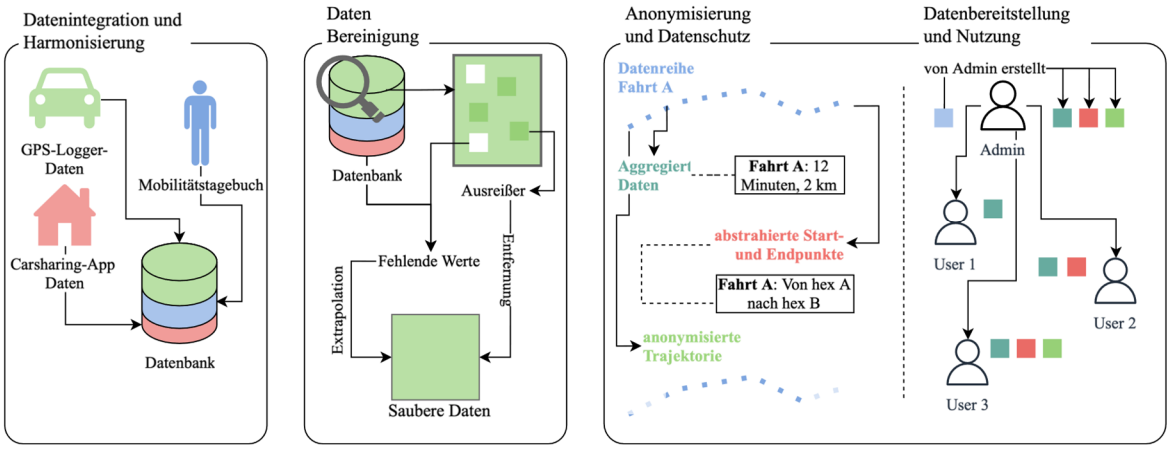


Abbildung 4: Workflow zur Datenaufbereitung

zung bereitgestellt. Die für die Weitergabe vorgesehenen, anonymisierten Datenbestände (vgl. Abschnitt Anonymisierung) werden dabei als Datenbankansichten (Views) anstatt physischen Tabellen angelegt. Dieser Ansatz hat den Vorteil, dass die Anonymisierungs- und Aggregationslogik zentral in den Views gepflegt wird. Sobald neue Rohdaten in die Datenbank importiert werden, stehen sie automatisch in der korrekt aggregierten, anonymisierten Form bereit, wodurch pro neuem Datensatz kein manueller Aufbereitungsschritt erforderlich ist. Die zuvor eingeführten Datenaggregationsstufen werden projektbezogen als entsprechende Ansichten abgebildet, sodass jedes Projekt auf die für seine Forschungsfrage vorgesehenen Datenformate zugreift. Die Zugriffskontrolle erfolgt als rollenbasierte Zugriffskontrolle (Role-Based Access Control, RBAC) (Sandhu et al., 1996): Ein Administrator weist den Datenbanknutzenden Rollen zu und regelt auf dieser Grundlage, welche Ansichten von wem eingesehen werden dürfen, wobei sich die Berechtigung an der jeweiligen Projektzugehörigkeit ausrichtet.

Die über die Views bereitgestellten Daten dienen nicht nur der projektbezogenen Forschungsauswertung, sondern finden darüber hinaus in unterschiedlichen Nutzungsszenarien Verwendung. So fließen sie in Publikationen ein, in denen Ergebnisse aus der Mobilitätsforschung der wissenschaftlichen Gemeinschaft sowie der Fachöffentlichkeit zugänglich gemacht werden. Zudem bilden sie eine Grundlage für die Bürger*innenbeteiligung. Auf Basis der anonymisierten und aggregierten Bestände lassen sich verkehrliche Sachverhalte evidenzbasiert und verständlich aufbereiten, um Bürgerinnen und Bürger in Planungs- und Entscheidungsprozesse einzubinden. Schließlich kommen die Daten auch in der Öffentlichkeitsarbeit zum Einsatz, etwa um Forschungsergebnisse und daraus abgeleitete Erkenntnisse einer breiteren Öffentlichkeit zu vermitteln. In allen Fällen greifen die Nutzenden ausschließlich auf die anonymisierten Views zu, sodass die Trennung zwischen Rohdatenhaltung und Weiterverwendung stets gewahrt bleibt.

5 Diskussion

Das vorgestellte Vorgehensmodell zeigt, dass sich heterogene Mobilitätsdaten innerhalb eines interdisziplinären Forschungsclusters zentral integrieren, bereinigen und nachnutzbar aufbereiten lassen. Am Beispiel der Projekte ComfficientShare 2.0 und ReFuMoLab_Garching wird deutlich, wie Carsharing-Flottendaten und appgestützte Mobilitätstagebücher in einem gemeinsamen PostgreSQL/PostGIS-basierten Datenmodell zusammengeführt werden. Der Mehrwert einer solchen Zentralisierung liegt weniger in der Datenhaltung als in der einheitlichen Bereinigungs- und Anonymisierungskette: Sie reduziert Doppelarbeiten zwischen Cluster-Projekten, vereinfacht die Einhaltung der clusterweiten Datenschutzvereinbarung und macht die Prozessschritte für Publikationen und Nachnutzende transparent nachvollziehbar. Es zeigt sich, dass die Datenverarbeitung über die Projekte hinweg einheitlich gelingt, wodurch die Funktionsfähigkeit des Prozesses bereits im laufenden Betrieb deutlich wird.

Dem stehen Grenzen gegenüber. Der Integrationsaufwand wächst mit jeder neuen Datenquelle. E-Scooter-Telematik, Maut- oder mobilfunkbasierte Verkehrsdaten erfordern jeweils eine eigene Integration, Schema-Mappings und Bereinigungsregeln. Die Validierung an Referenzdaten ist nur dort möglich, wo geeignete Erhebungen wie MiD, SrV oder Mobilität.Leben vorliegen. Darüber hinaus können rechtliche Fortschreibungen etablierte Anonymisierungs-Workflows nachträglich infrage stellen. Zudem setzt das Modell eine funktionierende Cluster-Governance voraus: Nur wenn alle Beteiligten ihre Daten tatsächlich zentral einspeisen, entfaltet sich der skizzierte Mehrwert.

Die Übertragbarkeit auf vergleichbare Strukturen, etwa andere BMFTR-Zukunftscluster oder kommunale Datenräume, ist grundsätzlich gegeben, da die Prozesse generisch gehalten sind. Bei stark kommerziell getriebenen Datenräumen mit Datensouveränität zwischen eigenständigen Akteuren stößt das Modell jedoch an paradigmatische Grenzen (vgl. Mobility Data Space). Als zentrale Erkenntnis lässt sich festhalten: Die frühzeitige Festlegung auf ein gemeinsames Datenmodell, Datenschutzvereinbarungen und Anonymisierungs-Workflow ist entscheidend für einen effizienten Forschungsbetrieb.

6 Fazit und Ausblick

Der Beitrag hat ein Vorgehensmodell für den Aufbau und Betrieb einer clusterinternen Mobilitätsdatenbank am Beispiel von MCube vorgestellt. Es verbindet die Integration heterogener Quellen in ei-

nem einheitlichen Datenmodell mit einer mehrstufigen Bereinigung und einem Anonymisierungs-Workflow, der den Schutz der Teilnehmenden mit dem analytischen Nutzen für Querschnitts- und Längsschnittanalysen in Einklang bringt. Die Validierung anhand bekannter Mobilitätsmuster bestätigt, dass die zentral aufbereiteten Daten den erwarteten Mustern entsprechen. Zugleich werden Grenzen sichtbar, die aus der automatisierten Erkennung und der Heterogenität der Quellen erwachsen.

Daraus ergibt sich die Empfehlung, Mobilitätsdaten innerhalb von Organisationen und Forschungsclustern zentral zu verarbeiten und Datenbereinigung sowie Anonymisierung als gleichberechtigte Bestandteile des Forschungsdatenzyklus zu verankern, nicht als nachgelagerte Schritte.

Weitere Studien sollten die Erweiterung um neue Datenquellen (ÖPNV-Soll/Ist-Daten, weitere Sharing-Angebote, mobilfunkbasierte Daten), die Automatisierung von Schema-Mapping sowie die Ausreißererkennung etwa durch ML-gestützte Verfahren adressieren. Auch die Verankerung der FAIR-Prinzipien über Metadatenstandards und die Anbindung an übergeordnete Initiativen wie NFDI und Datenräume sind relevante Forschungsfelder.

Danksagung

Die vorliegende Arbeit wurde durch die Förderung des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) ermöglicht. Die Förderung ist Teil der Initiative Clusters4Future und ihres Mobilitätsclusters MCube (Förderkennzeichen: 03ZU2105CA).

Literaturverzeichnis

Balkić, Z., Šoštarić, D., & Horvat, G. (2012). GeoHash and UUID identifier for multi-agent systems. In G. Jezic, M. Kusek, N.-T. Nguyen, R. J. Howlett & L. C. Jain (Hrsg.), Agent and Multi-Agent Systems: Technologies and Applications (LNCS 7327, S. 290–298). Springer. https://doi.org/10.1007/978-3-642-30947-2_33

Bellini, P., Bilotta, S., Collini, E., Fanfani, M., & Nesi, P. (2024). Data sources and models for integrated mobility and transport solutions. Sensors, 24(2), 441. https://doi.org/10.3390/s24020441

Bundesministerium für Forschung, Technologie und Raumfahrt [BMFTR]. (o. J.-a). Eckpunkte der Förderung. Clusters4Future – Zukunftscluster-Initiative. https://www.clusters4future.de/foerderinitiative/eckpunkte-der-foerderung

Bundesministerium für Forschung, Technologie und Raumfahrt [BMFTR]. (o. J.-b). Zukunftscluster MCube. Clusters4Future – Zukunftscluster-Initiative. https://www.clusters4future.de/die-zukunftscluster/die-zukunftscluster-der-ersten-wettbewerbsrunde/mcube

Clusterplattform Deutschland. (o. J.). Münchner Cluster für die Zukunft der Mobilität in Metropolregionen (MCube). https://www.clusterplattform.de/CLUSTER/Redaktion/DE/Cluster/MCube_Zukunftscluster.html

Deutsche Forschungsgemeinschaft [DFG]. (o. J.). Nationale Forschungsdateninfrastruktur (NFDI). https://www.dfg.de/de/foerderung/foerderinitiativen/nfdi

Engels, F., Wentland, A., & Pfotenhauer, S. M. (2019). Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. Research Policy, 48(9), 103826. https://doi.org/10.1016/j.respol.2019.103826

Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme IVI. (2022). Mobility Data Space: Ein sicherer Datenraum für die souveräne und plattformübergreifende Bewirtschaftung von Mobilitätsdaten (Whitepaper, 2. Aufl.). https://www.mobility-dataspace.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_DE_20220603_web.pdf

GIDAS – German In-Depth Accident Study. (2024). About GIDAS. https://www.gidas.org/about-en.html

GraphHopper GmbH. (o. J.). GraphHopper Routing Engine [Software]. https://github.com/graphhopper

Loder, A., Dahmen, V., Waldorf, I., Álvarez-Ossorio Martinez, S., Cantner, F., & Bogenberger, K. (2024). The Mobilität.Leben study: A 20-month mobility-tracking panel (Arbeitsberichte Mobility Policy, Bd. 2). Technische Universität München. https://media-tum.ub.tum.de/1751562

MCube. (o. J.). MCube – Münchner Cluster für die Zukunft der Mobilität in Metropolregionen. Technische Universität München. https://mcube-cluster.de/

Nachtigall, N., Dietermann, B., & Lienkamp, M. (2025). Design and evaluation of a closed-group, electric residential car-sharing system – Findings from a case study in Munich, Germany. Cleaner and Responsible Consumption, 19, 100340. https://doi.org/10.1016/j.clrc.2025.100340

Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (S. 336–343). ACM. https://doi.org/10.1145/1653771.1653818

Nobis, C., & Kuhnimhof, T. (2018). Mobilität in Deutschland – MiD: Ergebnisbericht. Studie von infas, DLR, IVT und infas 360 im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur. https://www.mobilitaet-in-deutschland.de/archive/pdf/MiD2017_Ergebnisbericht.pdf

OpenStreetMap contributors. (2024). Planet dump [Datensatz]. https://planet.openstreetmap.org

PostGIS Project Steering Committee. (2024). PostGIS documentation. https://postgis.net/documentation/

PostgreSQL Global Development Group. (2024). PostgreSQL 16 documentation. https://www.postgresql.org/docs/16/

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. Computer, 29(2), 38–47. https://doi.org/10.1109/2.485845

Schuessler, N., & Axhausen, K. W. (2009). Processing raw data from Global Positioning Systems without additional information. Transportation Research Record, 2105(1), 28–36. https://doi.org/10.3141/2105-04

Waldorf, I., Loder, A., Müller, J., Dahmen, V., & Bogenberger, K. (2026). Potential for commuting demand management: Analysis of day-to-day modal variability based on a 1-year GPS-tracking dataset. Transportation Research Record. https://doi.org/10.1177/03611981251356498

Zheng, Y. (2015). Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology, 6(3), Artikel 29. https://doi.org/10.1145/2743025

Autor*innenangaben

Svetlana Zubareva
Technische Universität München
TUM School of Engineering and Design
Lehrstuhl für Fahrzeugtechnik
Boltzmannstraße 15
85748 Garching b. München

E-Mail: s.zubareva@tum.de

Louis Stille-Hönig
Technische Universität München
TUM School of Engineering and Design
Lehrstuhl für Fahrzeugtechnik
Boltzmannstraße 15
85748 Garching b. München

E-Mail: louis.stille-hoenig@tum.de

Markus Lienkamp
Technische Universität München
TUM School of Engineering and Design
Lehrstuhl für Fahrzeugtechnik
Boltzmannstraße 15
85748 Garching b. München

E-Mail: lienkamp@tum.de

Autor*innenbeitrag

Svetlana Zubareva und Louis Stille-Hönig haben zu gleichen Teilen zu dieser Arbeit beigetragen und teilen sich die Erstautorenschaft. Sie waren verantwortlich für Konzeptualisierung, Methodik, Software, formale Analyse, Datenkuratierung und Visualisierung, verfassten den Originalentwurf und wirkten an der Überprüfung und Bearbeitung des Manuskripts mit. Markus Lienkamp hat zur Konzeption und wissenschaftlichen Einordnung der Arbeit beigetragen, die Untersuchung im Rahmen des Forschungsprojekts wissenschaftlich betreut und das Manuskript kritisch überarbeitet. Alle Autor*innen haben die eingereichte Fassung gelesen und ihr zugestimmt.