HOCHSCHULE ESSLINGEN

Gefördert durch:

Bundesministerium für Forschung, Technologie und Raumfahrt

Finanziert von der Europäischen Union
NextGenerationEU

Anonymisierte Erfassung und Nutzung von Mobilitäts- und Bewegungsdaten

Prof. Dr. Dominik Schoop

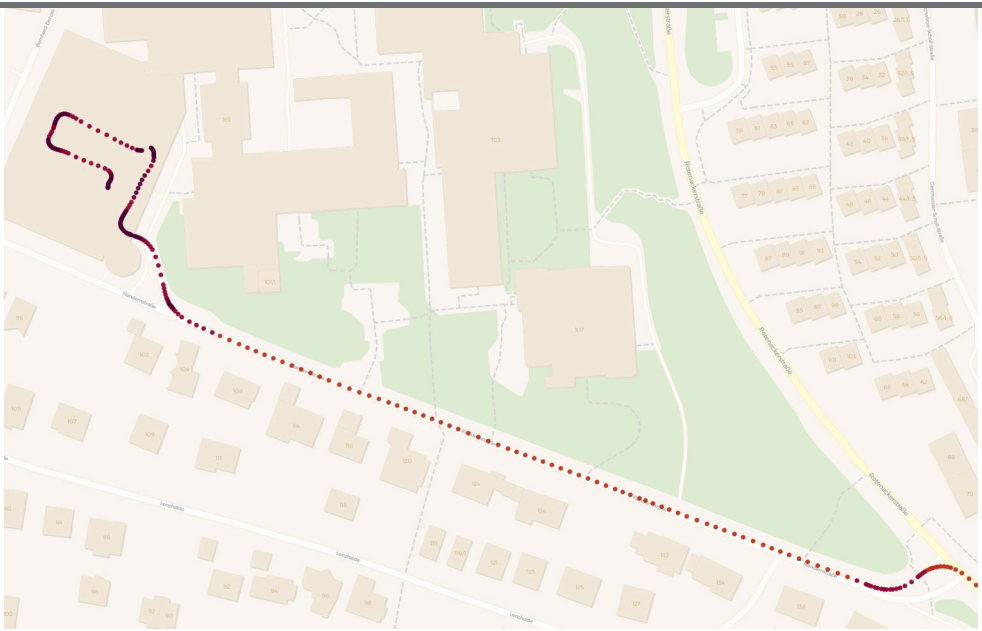
Hochschule Esslingen, Fakultät Informatik und Informationstechnik

3. Juni 2025

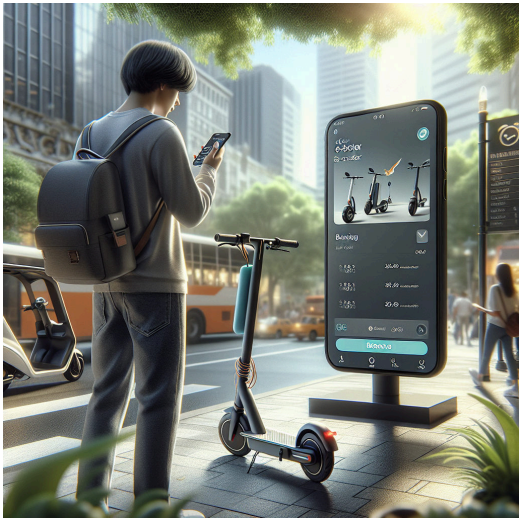
Agenda

(Kartenquelle: OpenStreetMap, `www.openstreetmap.org`)

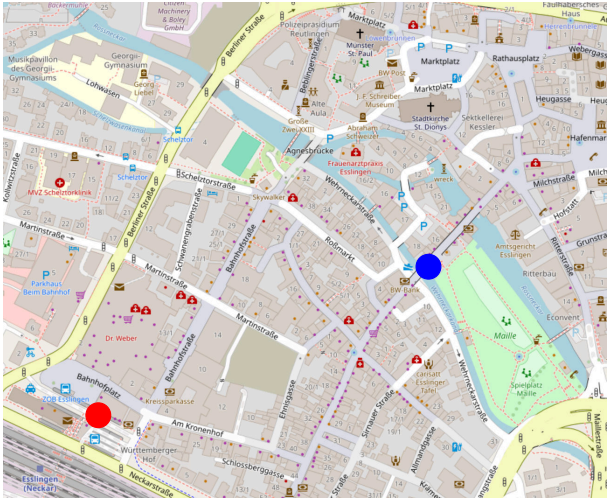
(Bildquelle: D. Schoop)



(Kartenquelle: OpenStreetMap, Trajektoriendarstellung mit Keppler.gl)

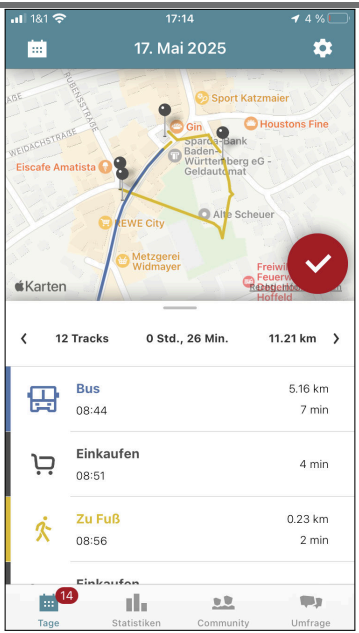


(Bildquelle: bwGPT)



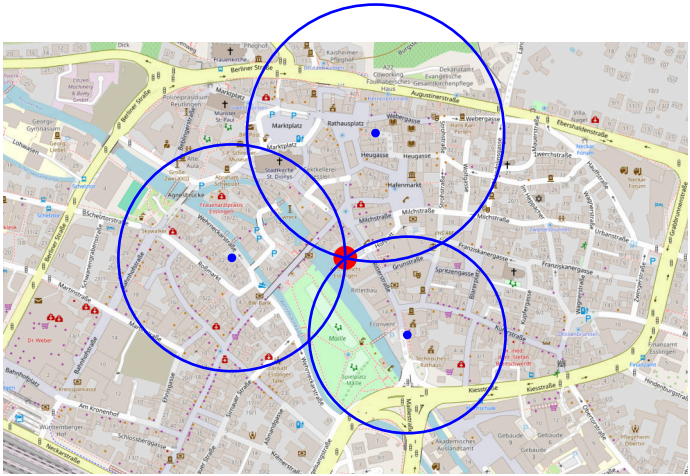
(Kartenquelle: OpenStreetMap)

(Bildquelle: bwegt Mobilität für Baden-Württemberg)

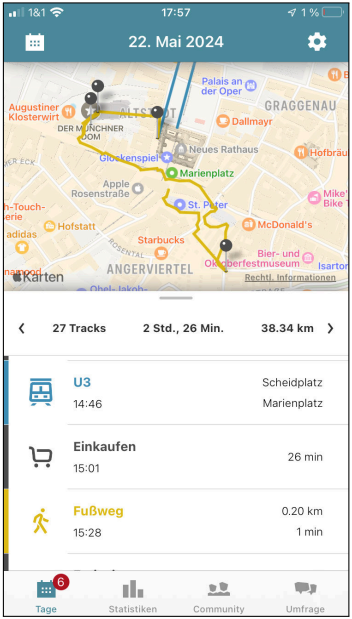


(Bildquelle: MOTIONTAG App, D. Schoop)

(Bildquelle: bwGPT)



(Kartenquelle: OpenStreetMap `openstreetmap.org`)



- Trajektorie = Bewegungspfad eines Objekts als Sequenz von Bewegungspunkten (Ort, Zeit, Richtung, Geschwindigkeit)
- Bewegungsdaten = Sequenzen von Trajektorien möglicherweise Personen und Fahrzeugen zugeordnet
- Mobilitätsdaten
  = Bewegungsdaten + Infrastrukturdaten + Verkehrsmittel + Sozialdaten + ...

(Bildquelle: MOTIONTAG App, D. Schoop)

Einzelperson

- Nutzung von Location Based Services
- Informationsangebote: zeitaktuelle Fahrpläne, Verkehrsmittelauslastung, ...
- ortsgebundenes Ticketing
- Selbst-Protokollierung
- Steuerung von indirekten Funktionen (Heizen beim Heimkommen, ...)

Mobilitätsdienstanbieter, Dritte

- Geschäftsmöglichkeiten
- bedarfsgerechte, ökonomische und ökologische Mobilitätsdienstleitungen
- Steuerung von Mobilitätsangeboten
- Planung von Mobiltätsangeboten
- Abrechnung von Mobilitätsangeboten
- Einnahmenaufteilung ÖPNV

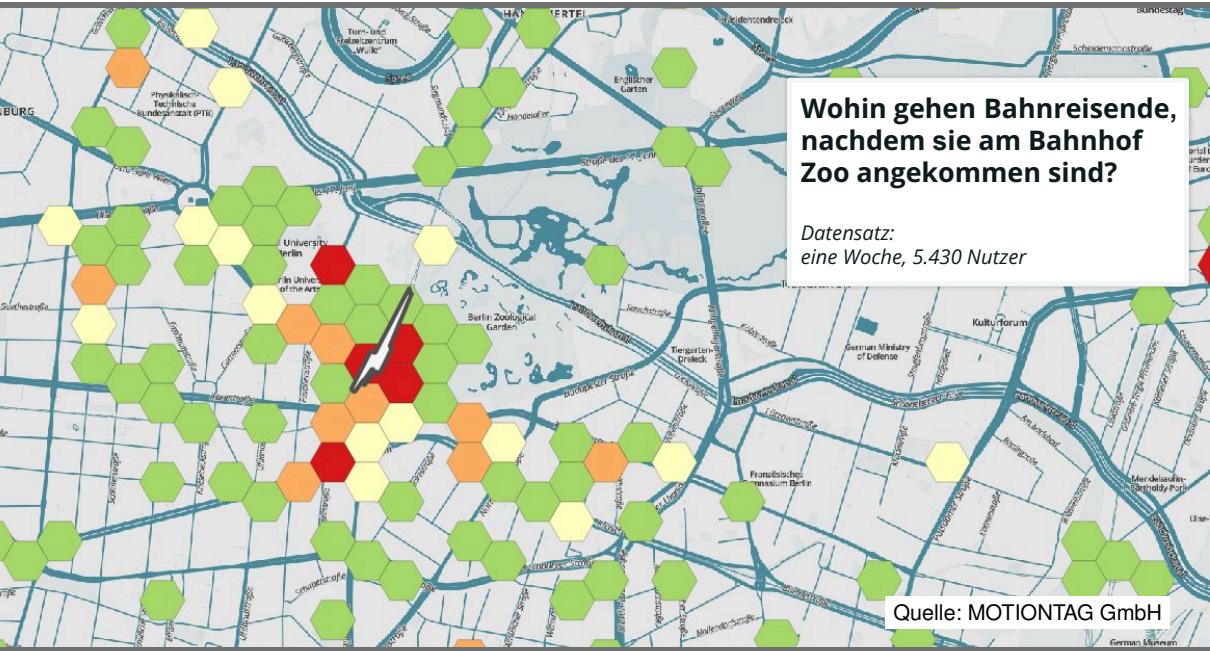
Kommunen

- Verkehrsflusssteuerung
- Gestaltung von Mobilitätsangeboten
- Verkehrs- und Stadtplanung
- Umweltmanagement

Gesellschaft

- bedarfsgerechte, ökonomische und ökologische Mobilität
- Beteiligung am Gesellschaftsleben
- Chancengleichheit
- Transparenz und Demokratie

Plattform für mobilitätsrelevante Daten des Landes (`https://www.mobidata-bw.de/dataset`):



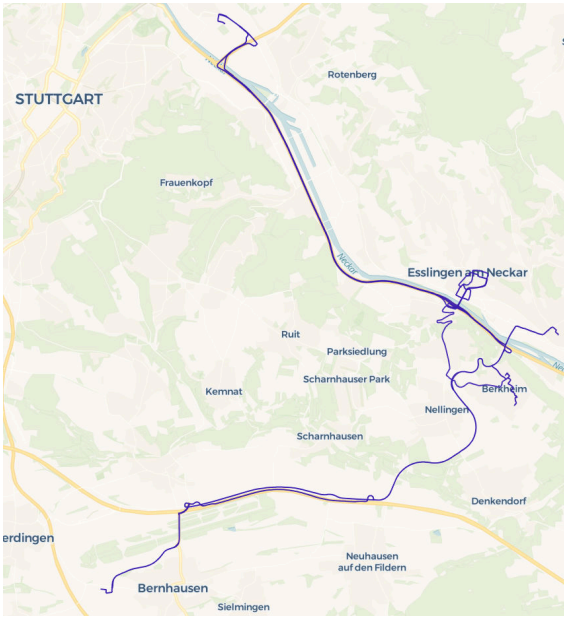
Thesen

- Eine bedarfsgerechte, ökonomische und ökologische Gestaltung des Lebensraumes einschließlich Verkehr benötigt Erkenntnisse über das Verhalten und den Bedarf von Personen.
- Manche Erkenntnisse können nur aus Mobilitäts- und Bewegungsdaten mit hohem Informationsgehalt von einer großen Anzahl von Personen gewonnen werden.

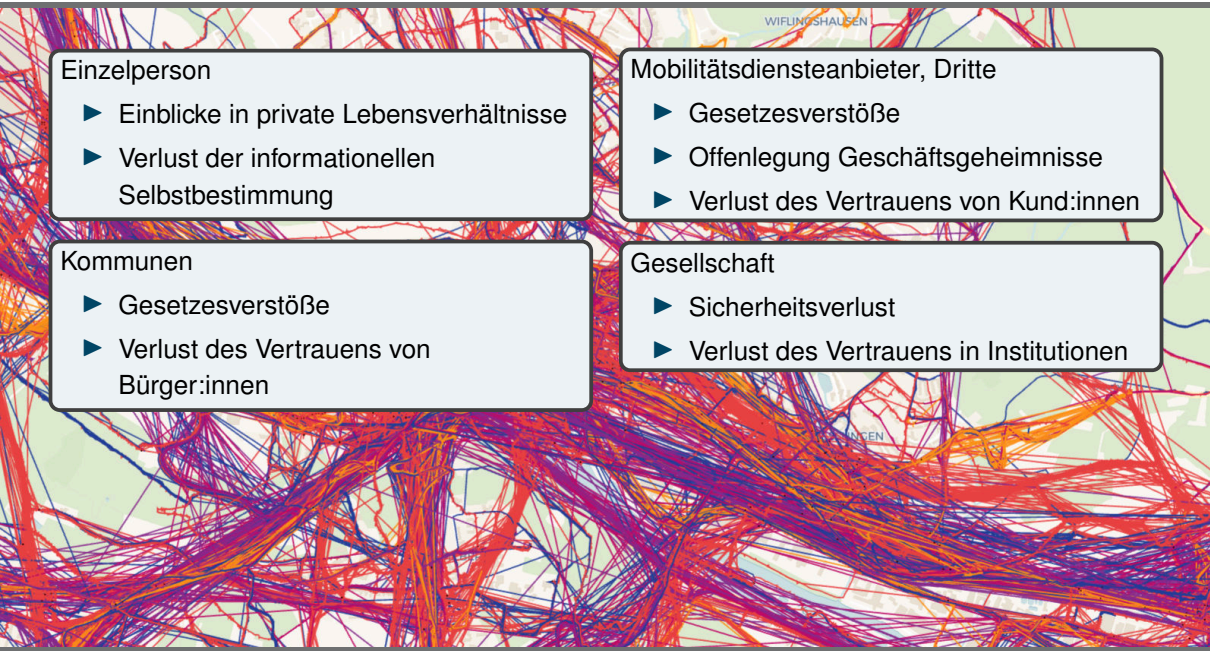
Mögliche Erkenntnisse aus den Trajektorien einer simulierten Person:

- wohnt neben der Kirche in Esslingen-Berkheim
- arbeitet als Kellner:in im Restaurant Einhorn in Esslingen
- geht gerne schwimmen
- ist Fan vom VfB Stuttgart
- fährt ein Auto der Marke Lexus
- ist evangelischen Glaubens
- raucht
- besucht häufig eine Spielothek

Bewegungsdaten können intimste personenbezogene Daten beinhalten einschließlich Daten nach Art. 9 EU-DSGVO.



(Quelle: Karte: OpenStreetMap, Daten: simuliert, Darstellung: Kepler.gl)



Einzelperson

- Einblicke in private Lebensverhältnisse
- Verlust der informationellen Selbstbestimmung

Mobilitätsdiensteanbieter, Dritte

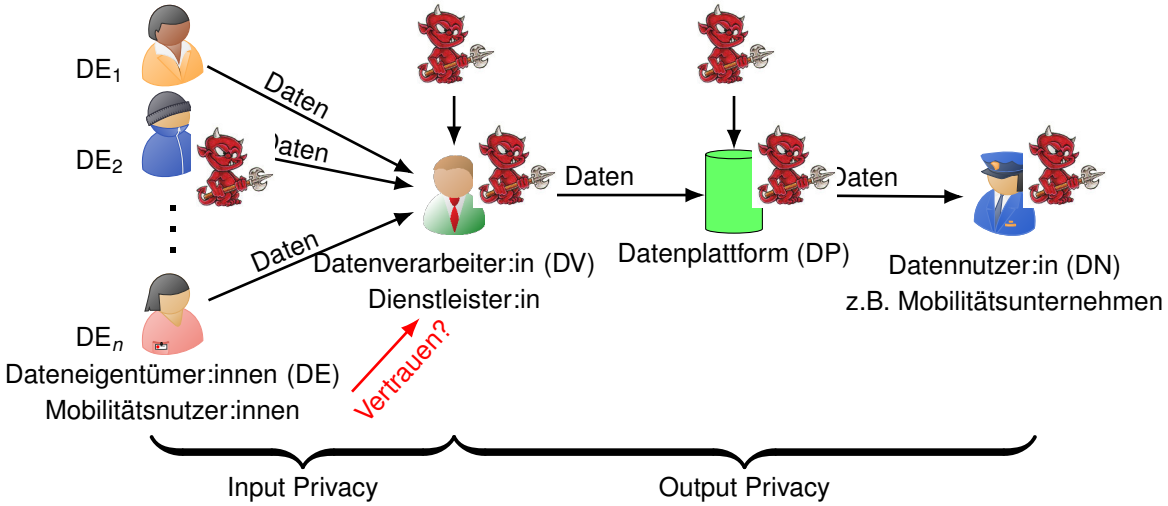
- Gesetzesverstöße
- Offenlegung Geschäftsgeheimnisse
- Verlust des Vertrauens von Kund:innen

Kommunen

- Gesetzesverstöße
- Verlust des Vertrauens von Bürger:innen

Gesellschaft

- Sicherheitsverlust
- Verlust des Vertrauens in Institutionen


DE1
DE2
DEn
Daten
Daten
Daten
Datenverarbeiter:in (DV)
Dienstleister:in
Daten
Datenplattform (DP)
Daten
Datennutzer:in (DN)
z.B. Mobilitätsunternehmen
Dateneigentümer:innen (DE)
Mobilitätsnutzer:innen
Vertrauen?
Input Privacy
Output Privacy

Herausforderung „Nützlichkeit anonymisierter Daten“

Ein Anonymisierungsverfahren

- transformiert Daten
- löscht Daten
- fügt Daten hinzu
- fasst Daten zusammen

$\Rightarrow$

Der Informationsgehalt anonymisierter Daten ist meist geringer als der der Originaldaten.

- Informationsgehalt ausreichend?
- Verbliebene Information die benötigte?

Herausforderung „Definition und Bestimmung der Anonymität“

Wann sind Daten überhaupt anonym?

Definition „anonyme Informationen“ nach Sätzen 5 und 6 EWG 26 EU-DSGVO

Informationen (Daten) sind anonym, wenn sie sich nicht auf natürliche Personen beziehen oder wenn eine betroffene Person nicht mehr identifiziert werden kann.

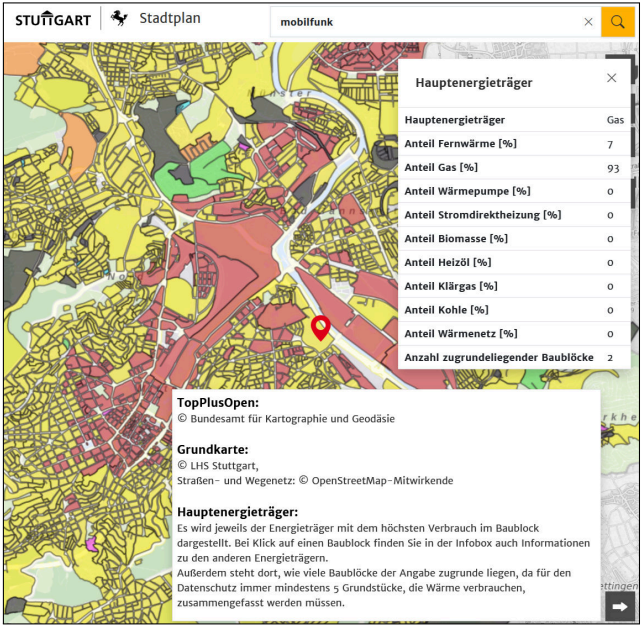
⇓

Anonymität ist ein qualitatives Maß.
Daten sind anonym oder personenbeziehbar.
Anwendungsbereich der EU-DSGVO ist eröffnet oder nicht.

Definitionen von „Anonymität“:

k-Anonymität Daten sind *k*-anonym, wenn jede Datenabfrage die Daten von mindestens *k* Personen zurückliefert.
1-anonym = Person identifiziert
je größer *k* desto größer die Anonymität

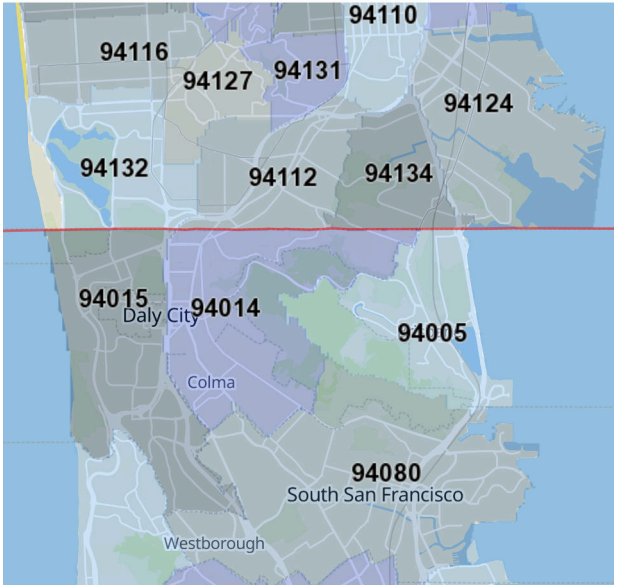
Anonymität ist ein quantitatives Maß.
Ab welchem Wert sind Daten anonym?



Ortsbezogene Information zu Hauptenergieträgern in Stuttgart auf `https://maps.stuttgart.de/stadtplan`

„Außerdem steht [im Datensatz], wie viele Baublöcke der Angabe zugrunde liegen, da für den Datenschutz immer mindestens 5 Grundstücke, die Wärme verbrauchen, zusammengefasst werden müssen.“

- Identifizierbarkeit hängt von Gruppengrößen ab, die durch Kombination von Attributen bestimmt werden
- Experiment: Bevölkerung in USA bzgl. Geburtstag, Geschlecht und ZIP-Code/County



ZIP codes in the south of San Francisco

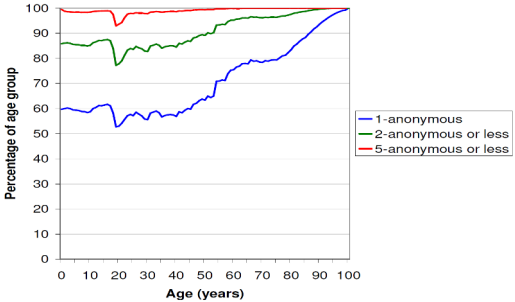


Figure 1: Anonymity of the U.S. population, by age, given {Gender, ZIP code, Full date of birth}.

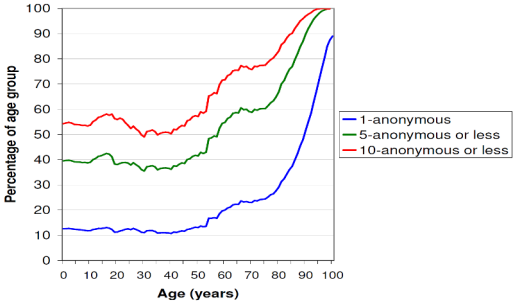


Figure 2: Anonymity of the U.S. population, by age, given {Gender, County, Full date of birth}.

(Quelle: [Gol06])

- 1-anonym = Person ist eindeutig identifizierbar
- *k*-anonym = jede Person erscheint mit mindestens *k* Personen identisch zu sein
- US Counties haben eine Population zwischen 18 und 705.000
- Verknüpfung mit anderen Datenquellen (z.B. Wählerlisten) kann zur Re-Identifizierung mit Namen führen

Definitionen von „Anonymität“:

k-Anonymität Daten sind k-anonym, wenn jede Datenabfrage die Daten von mindestens k Personen zurückliefert.
1-anonym = Person identifiziert
je größer k desto größer die Anonymität

Differential Privacy Ein Analyseverfahren gewährleistet „Differential Privacy“, wenn aus der Ausgabe des Verfahrens nicht bestimmt werden kann, ob eine beliebig bestimmte Person in den Originaldaten vorkam oder nicht.

ϵ-Differential Privacy Ein Analyseverfahren A gewährleistet „ϵ-Differential Privacy“, wenn die Wahrscheinlichkeit einer bestimmten Ausgabe O sich maximal um den Faktor e^{ϵ} unterscheidet, wenn eine Person in den Daten vorkommt oder nicht.
$Pr[A(D_1) \in O] \leq e^{\epsilon} \cdot Pr[A(D_2) \in O]$

$\Downarrow$

Anonymität ist ein quantitatives Maß.
Ab welchem Wert sind Daten anonym?

Thesen

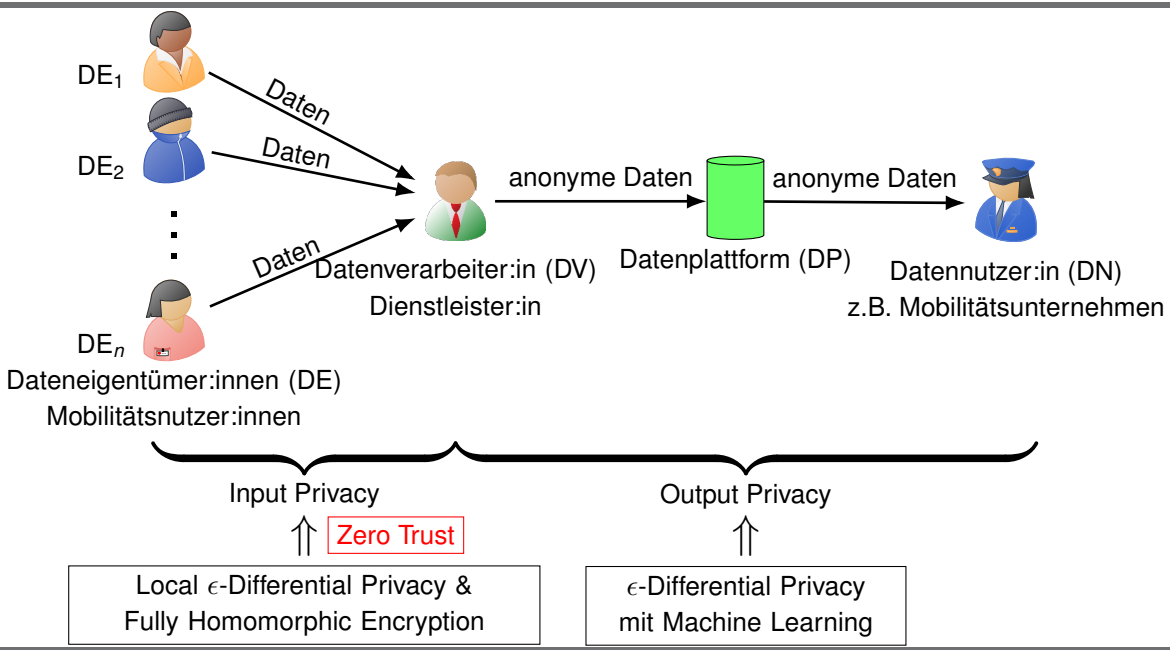
- Bewegungs- und Mobilitätsdaten entstehen dauernd, werden aber kaum genutzt
- Personen und Institutionen zögern aus Datenschutzgründen persönliche Bewegungsdaten herzugeben bzw. zu verarbeiten
- Detailreiche Bewegungsdaten sind für viele Anwendungsfälle nützlich

bessere Lösungsansätze erforderlich

Forschungsprojekt
„Anonymisierte Erfassung und Nutzung von Mobilitäts- und Bewegungsdaten (AnoMoB)“

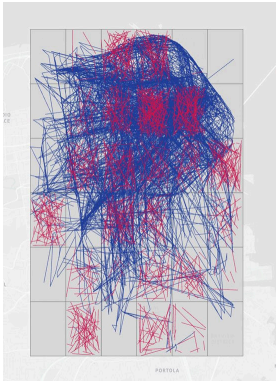
Agenda

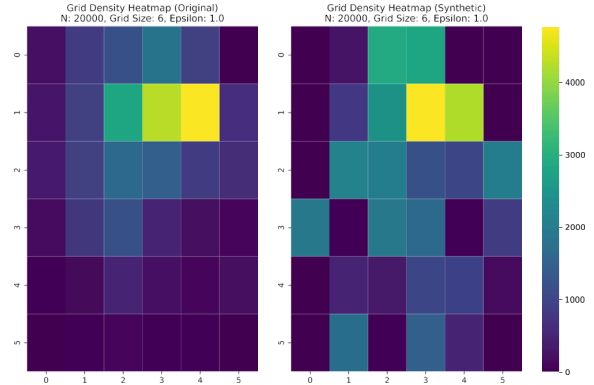
Projekttitel	Anonymisierte Erfassung und Nutzung von Mobilitäts- und Bewegungsdaten (AnoMoB)
Förderer	BMFTR  
Bekanntmachung	„Forschungsnetzwerk Anonymisierung für eine sichere Datennutzung“, 22 geförderte Projekte aufgeteilt auf 5 Cluster
Förderdauer	15.12.2022 – 14.12.2025
Cluster	Intelligenter Intermodaler Pendlerverkehr (IIP)
Projektpartner	Hochschule Esslingen
	Fraunhofer IAO, Stuttgart
	cantamen GmbH, Hannover
	MOTIONTAG GmbH, Potsdam
Projektsteckbrief	`https://www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/anomob`


DE1
DE2
DEn
Daten
Daten
Daten
Datenverarbeiter:in (DV)
Dienstleister:in
anonyme Daten
anonyme Daten
Datenplattform (DP)
Datennutzer:in (DN)
z.B. Mobilitätsunternehmen
Dateneigentümer:innen (DE)
Mobilitätsnutzer:innen
Input Privacy
Output Privacy
Zero Trust
Local ϵ-Differential Privacy &
Fully Homomorphic Encryption
ϵ-Differential Privacy
mit Machine Learning

- Ansatz der Local ϵ-Differential Privacy [Du+23]
- Dateneigentümer:innen verrauschen Start- und Endpunkt einer Trajektorie
- statistische Verteilung der Längen und Übergangswahrscheinlichkeiten von Trajektorien in einem Grid
- Generierung von synthetischen Trajektorien

Beispiel: GPS-Koordinaten von ca. 500 Taxis in San Francisco, USA



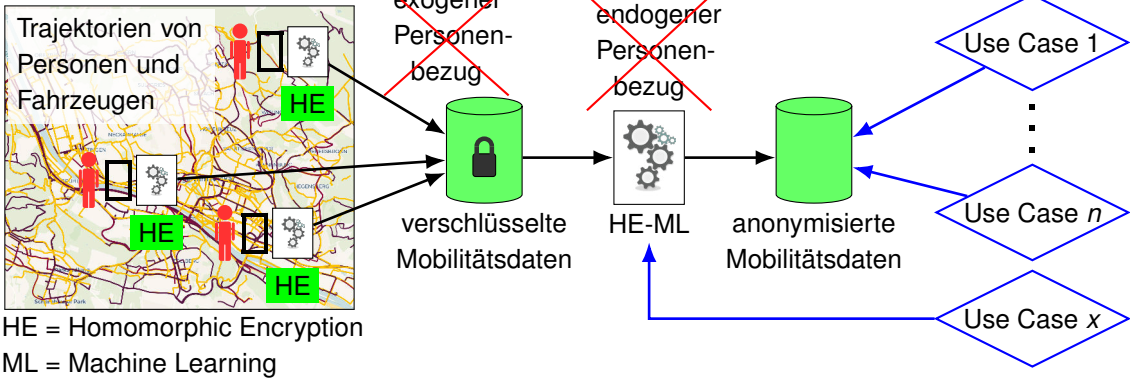


(Bildquelle: G. Gühring, D. Hu, Hochschule Esslingen)

- Synthetisierte Trajektorie wird mit einem Map-Matching-Algorithmus wieder auf die Straße gelegt.
- Die Punkte der synthetisieren Trajektorie werden auf die nächste Straße geschoben, Punkte werden mit Routing miteinander verbunden.
- Wichtige Points of Interest (POIs) der Mobilitätsnutzer:innen sind weiterhin identifizierbar.

(Bildquelle: G. Gühring, D. Hu, Hochschule Esslingen)

- Daten werden bei Eigentümer:in homomorph verschlüsselt.
- Datenverarbeiter:in kann Daten nicht einsehen aber anonymisieren bzw. anonymisiert auswerten.
- Unverschlüsselte, anonymisierte Mobilitätsdaten stehen für die Anwendungsfälle der Nutzer:innen zur Verfügung.



HE = Homomorphic Encryption
ML = Machine Learning

Homomorphe Verschlüsselung beruht auf gitterbasierter Kryptographie und dem Problem des Lernen mit Fehlern (learning with errors, LWE)

- Rauschen wird zum Klartext hinzugefügt.
- Das Rauschen vergrößert sich beträchtlich, wenn Schlüsseltexte multipliziert werden.
- Ein zu großes Rauschen verhindert die erfolgreiche Entschlüsselung von Schlüsseltexten (leeres Rausch-Budget)

$\Rightarrow$ Rauschen muss gesteuert werden

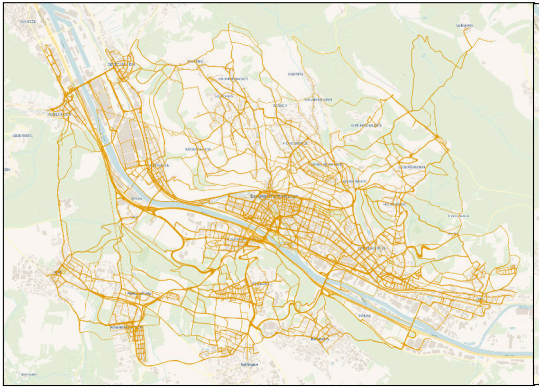
Option 1: Bootstrapping: Rauschen wird immer wieder reduziert

Option 2: tiefenbeschränkte Verschlüsselung: Anzahl der Multiplikationen wird durch das Rausch-Budget bestimmt

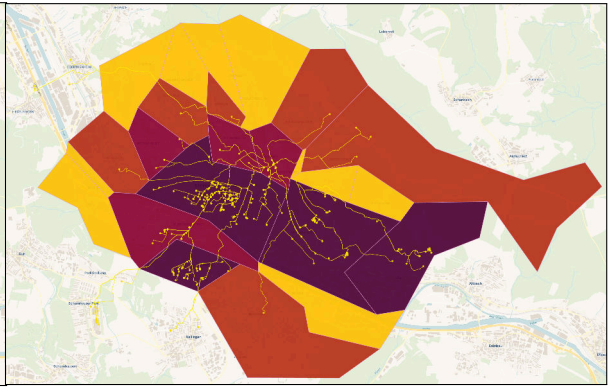
Herausforderung bzgl.

- Rechenzeit
- Speicherbedarf

Aus welchen Stadtteilen von Esslingen pendeln wie viele Personen an den Hochschulstandort „Flandernstraße“? (berechnet aus 16.384 verschlüsselten Trajektorien)



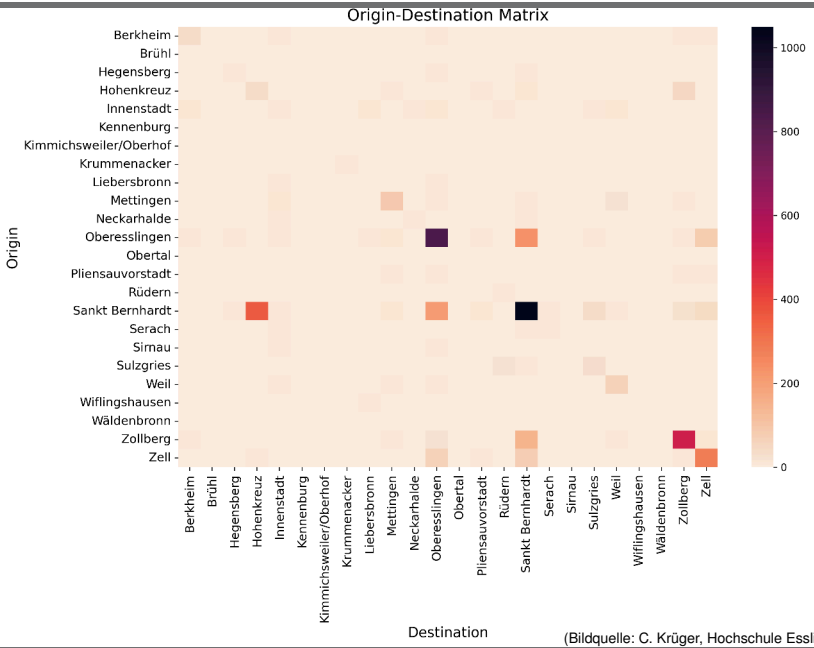
Originaltrajektorien
(simuliert)



Heat Map der Häufigkeiten
(mit relevanten Trajektorien)

(Bildquelle: OpenStreetMap, Kepler.gl, C. Krüger, Hochschule Esslingen)

Von welchem Teilort zu welchem Teilort bewegen sich Personen in Esslingen? (berechnet aus 16.384 verschlüsselten Trajektorien)



(Bildquelle: C. Krüger, Hochschule Esslingen)

- Personenbezogene Bewegungsdaten fallen in vielen digitalen Applikationen an, werden wegen des Datenschutzes aber nur begrenzt genutzt.
- Detaillierte Bewegungsdaten von Personen wären für viele Anwendungsszenarien nützlich, wobei meist nur die aggregierten Daten relevant sind.
- Bewegungsdaten einzelner Personen haben einen hohen Informationsgehalt und erlauben Einblick in das persönliche Leben.
- Dateneigentümer und Institutionen zögern daher Bewegungsdaten bereit zu stellen bzw. zu erheben.
- Die Gewährleistung von Input Privacy soll helfen, das Vertrauen von Bürger:innen und Kund:innen in Institutionen und Technologie zu wahren und damit die Bereitschaft fördern, Daten bereitzustellen.
- Methoden der Local ϵ-Differential Privacy und der homomorphen Verschlüsselung sind Ansätze für Input und Output Privacy.

[Arm+15] Frederik Armknecht u. a. *A Guide to Fully Homomorphic Encryption*. Cryptology ePrint Archive, Paper 2015/1192. https://eprint.iacr.org/2015/1192. 2015. URL: https://eprint.iacr.org/2015/1192.

[Bad+22] Ahmad Al Badawi u. a. *OpenFHE: Open-Source Fully Homomorphic Encryption Library*. Cryptology ePrint Archive, Paper 2022/915. https://eprint.iacr.org/2022/915. 2022. URL: https://eprint.iacr.org/2022/915.

[Du+23] Yuntao Du u. a. "LDPTrace: Locally Differentially Private Trajectory Synthesis". In: *Proc. VLDB Endow.* 16.8 (2023), S. 1897–1909. DOI: 10.14778/3594512.3594520. URL: https://www.vldb.org/pvldb/vol16/p1897-gao.pdf.

[Gol06] Philippe Golle. "Revisiting the uniqueness of simple demographics in the US population". In: *Proceedings of the 5th ACM workshop on Privacy in electronic society - WPES '06*. ACM Press, 2006. DOI: 10.1145/1179601.1179615.

[LB23] Janis Latus und Maxim Bickel. *Cryptanonyme Analyse von Mobilitätsdaten mittels homomorpher Verschlüsselung*. Techn. Ber. Hochschule Esslingen, 2023.

[PSG22] Michal Piorkowski, Natasa Sarafijanovic-Djukic und Matthias Grossglauser. *CRAWDAD epfl/mobility*. 2022. DOI: 10.15783/C7J010. URL: https://dx.doi.org/10.15783/C7J010.