

Aufbau eines Data Inventorys für KI Lösungen im Mobilitätsbereich

Cluster Mobility & Logistics

Aufbau & Problemstellung eines Mobility Data Inventorys

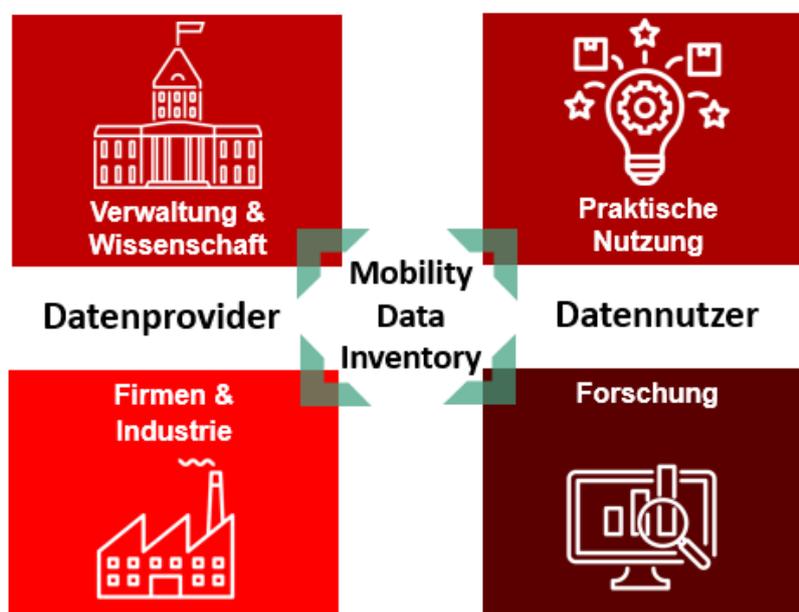
Stand 28.06.2022



Ausgangslage

Um die Verbreitung und Nutzbarkeit von KI-Anwendungen weiter zu erhöhen, sind große Mengen an Daten erforderlich, die als Grundlage für jegliche Art von KI-Systemen dienen. Gerade im Bereich der Mobilität werden dieses in einem großen Maße erhoben, jedoch entstehen für die Datennutzer dabei maßgebliche Probleme die eine zielgerichtete Verwendung erheblich erschweren. Zum einen werden die Daten von vielen verschiedenen Quellen und Organisationen erhoben, was dazu führt, dass es keine Sammlung bzw. Abstimmung der Datenerhebungen gibt (z.B. Lichtsignalanlagen, Wechselverkehrszeichen, Parkleitsysteme, Carsharing usw.) und zum anderen liegen die gesammelten Daten in sehr unterschiedlichen Datenformaten und -formen vor. Dies führt dazu, dass eine Vielzahl der eigentlich verfügbaren Daten nicht genutzt werden und deren Potential nicht ausgeschöpft wird. Um diese ungenutzte Ressourcen effizienter einzusetzen ist ein Bestreben von goAIR, eine zentrale Datenbank mit Mobilitätsdaten aus dem Raum Regensburg zu erschaffen.

Die Idee hinter der Realisierung eines solchen Data Inventorys / Data Hubs ist die systematische Zusammenführung von unterschiedlichen, bereits vorhandenen Mobilitätsdaten aus dem Raum Regensburg. Diese hat zum Ziel, eine digitale Datenplattform zu kreieren, die einen einfachen Zugang für potentielle Nutzer ermöglicht. Dadurch ergibt sich ein erhebliches Nutzungspotential gerade für SMEs, für die ein fehlender Datenzugang oftmals die entscheidende Hürde bei der Entwicklung und Umsetzungen von KI-Anwendungen darstellt. Dabei steht gerade die Verknüpfung der verschiedenen Datenprovider mit diversen Datennutzern im Vordergrund um die bekannte und aus der Wirtschaftsförderung vielfach erprobte Triple Helix Struktur aus Verwaltung, Wissenschaft und Wirtschaft auch im Gebiet der Datennutzung zu etablieren und so das Ökosystem Regensburg im Bereich KI zu stärken.



Konzeptidee

Um eine Verbesserung der Datensituation in Regensburg anzustreben und eine erste Anlaufstelle für Firmen in Bezug auf Mobilitätsdaten zu bieten, wurde im Rahmen von goAIR ein erster Data Inventory Ansatz erarbeitet. Um einen möglichst problemlosen Zugang mit einer einfachen Speichermöglichkeit und hoher Nutzbarkeit zu verbinden, wurden die Daten nicht direkt gesammelt, sondern Metainformationen der Daten in Tabellenform gespeichert. Um eine sinnvolle Metadatenstruktur zu erarbeiten wurde in Zusammenarbeit mit dem technologischen Institut für Künstliche Intelligenz TIKI eine maßgeschneiderte Tabellenstruktur entwickelt, die die wichtigsten Metainformationen übersichtlich darstellt und eine vereinfachte Übersicht über vorhandene Mobilitätsdaten im Raum Regensburg bietet. Die nachfolgende Grafik zeigt die gewählte Tabellenstruktur mit den 15 ausgewählten Attributen und ihren Beschreibungen.

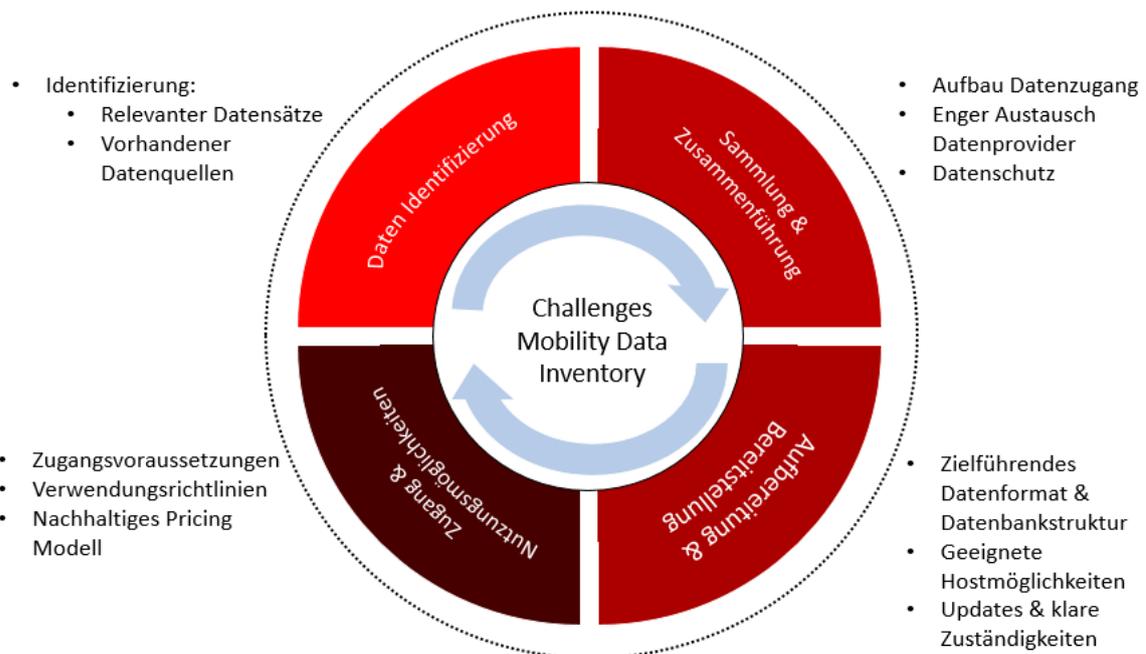
Beschreibung	Eindeutiger Bezeichner für den Datensatz	Sprechender Titel	Freitext mit den wesentlichen Eigenschaften	Primäre Verwendung des Datensatzes	Organisation	Kontaktdaten des Ansprechpartners bei einer beabsichtigten Nutzung	Zeitpunkt der Erstellung der aktuellen Version
Attribut	ID	Name	Beschreibung	Zweck	Datenproduzent	Datenmanager	Erzeugungszeitpunkt

Häufigkeit und Zeitpunkte der Erneuerung, Ergänzung, Erweiterung	Ablageort, bzw. Ort, an dem die Daten abgeholt werden können, z.B. URL	Einzeldateien Text, Bild, Zahlen/ Statistiken, Sensor oder Datenbank	Einzeldateien, z.B. csv, jpeg, json, xml; Datenbanken, z.B. SQL, Graph, Key-Value	Zeitraum, für den die Daten vorliegen	Abgedeckter räumlicher Bereich	Link zu Lizenzen	Letzte Aktualisierung der Informationen, d.h. der Aktualisierung des Datenkatalog-Eintrags selbst
Aktualisierung	Speicher	Typ	Format	Zeit	Geodaten	Nutzungsrechte / beschränkung	Stand

Nach der Erarbeitung dieser Struktur, wurden verschiedene relevante Organisationen kontaktiert um Zugang zu verschiedenen Mobilitätsdaten zu erhalten und darauf aufbauend die Tabelle zu befüllen. Dabei wurde ein verstärkter Fokus darauf gerichtet, in einer Pilotphase Daten von verschiedenen Datenprovidern aus den Bereichen Verwaltung, Forschung und von Firmen in die Struktur aufzunehmen um in einem engen Austausch mit den unterschiedlichen Akteuren etwaige Stärken und Schwächen der gewählten Struktur herauszuarbeiten und diese gegebenenfalls zu verbessern. Dazu wurden in 1:1 Gesprächen Daten von drei Firmen aus den Sektoren Automotive, Bau und IT in die Datenstruktur aufgenommen. Zusätzlich wurden Daten einer Hochschule sowie des Sachgebiets Straßenverkehrstechnik berücksichtigt. Im Rahmen dieses Austausches wurde allerdings eine Reihe von Schwachstellen der ausgewählten Struktur deutlich, die je nach Organisation stark voneinander abwichen und im Zuge eines größer angelegten Data Inventory's Beachtung finden müssten, was eine größere Umsetzung im geplanten Format verhinderte.

Problemstellung

Die Problemstellung der Umsetzung eines praxisrelevanten, nutzerfreundlichen Data Inventory's beinhaltet weitestgehend vier verschiedene Themenschwerpunkte. Die Identifizierung relevanter Daten, deren Sammlung, Zusammenführung, Aufbereitung und Bereitstellung sowie die Zugangs- und Nutzungsmöglichkeiten für die potentiellen Nutzergruppen.



1. Eine umfassende Identifizierung von relevanten Daten und den dazugehörigen Providern bedarf einer proaktiven Kontaktaufnahme mit relevanten Akteuren, die mobilitätsbezogenen Daten besitzen und gewillt sind, diese für eine gemeinsame Nutzung bereit zu stellen. In der Praxis ergeben sich daraus bereits zwei große Schwierigkeiten. Zum einen ist die Identifizierung der Datenprovider sehr schwierig, da diese häufig selbst organisationsübergreifend nicht wissen über welche Daten sie genau verfügen und zum anderen ist ein Überzeugungsprozess an vielen Stellen notwendig, um die Provider von den Vorteilen einer Datenbereitstellung zu überzeugen.
2. Bei der Sammlung und Zusammenführung der Daten spielen die zuvor genannten Herausforderungen gerade für den Zugang zu den konkreten Daten eine große Rolle. Darüber hinaus muss ein konkreter Rahmen geschaffen werden, der die Erfüllung aller datenschutzrechtlichen Vorgaben garantiert um einen Datenaustausch auch rechtlich zu ermöglichen.
3. Die Aufbereitung und Bereitstellung der Daten stellt die wohl größte Schwierigkeit bei dem Aufbau eines Data Inventorys dar und war maßgeblich dafür verantwortlich, die zugrundeliegende Idee eines einfach angelegten Data Inventorys mit Metadaten weiter zu verfolgen. Zunächst muss dabei ein ziel führendes Datenformat sowie eine passende Datenbankstruktur ausgewählt werden, was den Charakter des Inventorys maßgeblich beeinflusst. Die im Pilotprojekt gewählte Metadatenstruktur hat sich dabei als sinnvoll erwiesen, jedoch bereitet diese in der dargelegten Form unterschiedlichen Akteuren Probleme. Die Datenprovider können ihre Daten selbst nicht immer mit den Metaattributen darstellen, da sie diese selbst nicht kennen, bzw. nicht anwenden können. Dies betrifft vor allem die Verwaltung in der eine Vielzahl an unterschiedlichsten Daten in sehr großen Mengen erhoben werden. Dazu müsste von Seiten des Inventory Providers diese Kategorisierung teilweise selbst erfolgen beziehungsweise eine Hilfestellung für die Datenprovider erfolgen. Gleichzeitig haben allerdings die Datennutzer das Problem, dass eine Metadatenbank mit 15 Attributen fast schon zu sehr vereinfacht ist, um Ihnen einen konkreten Mehrwert zu bieten. Zudem muss das Konzept für den reibungslosen Zugang zu den eigentlichen Daten mit umgesetzt werden. Dabei muss vom Betreiber des Data Inventory's eine geeignete Hostmöglichkeit ausgewählt werden und die Daten müssen regelmäßig gepflegt und aktualisiert werden.
4. Der vierte Schwerpunkt der Problemstellung bezieht sich auf die Erarbeitung eines Zugangs- und Nutzungskonzeptes. Um die in den Punkten eins bis drei dargelegten Herausforderungen zu meistern, ist die Erstellung eines Data Inventorys mit einem erheblichen Ressourcen- und Personalaufwand verbunden. Damit sich dieser Aufwand rechtfertigt, ist es notwendig, ein Pricing Modell zu entwickeln, das zumindest bei erfolgreicher Verwendung der bereitgestellten Daten, einen Mehrwert für den Betreiber in Form einer Bezahlung oder anderen Verbindlichkeiten bietet. Zudem muss die Nutzung des Inventorys und den daraus entstehenden Anwendungen, klar geregelt sein um etwaige rechtliche Probleme und Interessenskonflikte bereits im Vorfeld auszuschließen.

Relevante Vergleichsprojekte

Als Orientierungshilfe für die Umsetzung eines Data Hubs für Mobilitätsdaten im Raum Regensburg können bereits erfolgreiche Vorhaben anderer Regionen dienen, die sich dem Thema Datennutzung verschrieben haben. Ein prominentes Beispiel davor ist das Berlin Data Mobility Hub. Ziel dieser Initiative ist die Erstellung einer digitalen Daten- und Wissensplattform zur systematischen Zusammenführung berlinbezogener Datenbestände des Personen- und Güterverkehrs unter besonderer Berücksichtigung der Erlangung von Erkenntnisgewinnen über die Auswirkungen der Covid-19-Pandemie auf das Berliner Verkehrswesen.

Die Plattform basiert auf einer umfangreichen Identifikation relevanter Datenquellen und der Analyse bestehender datenbezogener Bedarfs- und Problemlagen der Berliner Verkehrslandschaft. In diesem Rahmen wird darauf abgezielt, entsprechende Akteure zur Bereitstellung verkehrsbezogener Datenbestände zu gewinnen. Das Hub umfasst neben quantitativer Livedaten öffentlicher und privater Verkehrsanbieter sowie Mobilfunk- oder Navigationsbetreiber auch qualitative Daten sowie historische Datensätze. Ergänzend zu diesen Daten werden im Rahmen des Hubs integrierte Analysefunktionen angeboten, die den Nutzern einen direkten Mehrwert bieten. Gefördert wird das Vorhaben im Rahmen der Exzellenzstrategie von Bund und Ländern und geleitet von der TU Berlin und der FU Berlin.

Ein weiteres bekanntes und relevantes Vergleichsprojekt bietet die MDM-Plattform, ein nationaler Zugangspunkt für multimodale Mobilitätsdaten. Im Rahmen dieses Projekts bietet das Bundesministerium für Verkehr und digitale Infrastruktur Anbietern und Nutzern von Mobilitätsdaten eine neutrale B2B-Plattform mit definierten Standards für den Datenaustausch. Dabei ist vor allem eine reichhaltige Datenbasis über aktuelle Fahrpläne, Staus, Baustellen, Mobilitätsoptionen und Parkmöglichkeiten enthalten.

Durch die Plattform wird ein sicherer Daten- und Kommunikationsstandard sowie transparente Konditionen und eine hohe Datensicherheit gewährleistet. Die Plattform ist dabei als eine Art Dienstleistung angedacht und bietet den Kunden technische sowie organisatorische Entlastungen. Das Vorhaben ist dabei durch die EU im Rahmen der Entwicklung intelligenter Verkehrssysteme (IVS) gefördert. Zukünftig ist da Ziel der Initiative, möglichst viele Mobilitätsdaten zu erschließen, über verschiedene Verkehrsmittel, Netzelemente und Akteure hinweg. Dazu gehören auch Mobilitätsdaten, die parallel auf anderen Datenportalen in Deutschland angeboten werden. Somit entsteht ein Synergiepotential mit anderen Plattformen, die sich vor allem für Datennutzer als nützlich erweisen.

Ausblick

Die beiden dargestellten Beispiele lassen bereits die mögliche Richtung eines Mobility Data Hubs in Regensburg erahnen. Um einen wirklichen Mehrwert für die Nutzer, aber auch gute Rahmenbedingungen für die Datenprovider zu ermöglichen, ist es notwendig eine Lösung zu entwickeln die das Thema ganzheitlich abdeckt. Das komplette Spektrum der dargestellten Problemstellungen muss dabei genauso berücksichtigt werden, wie die Bedürfnisse und Wünsche der Datennutzer- und Anbieter.

Damit wäre es möglich eine möglichst breite Gruppe beider Parteien zu anzusprechen und für das Projekt zu gewinnen. Können diese Rahmenbedingungen erfüllt werden, zeichnet sich ein großes Nutzungspotential für die Mobilitätsdaten ab, was die Umsetzung von KI-Projekten im Raum Regensburg für Firmen, allen voran SMEs, deutlich vereinfacht. Solch ein Vorhaben kann in dieser Größenordnung allerdings nicht von den in goAIR beteiligten Organisationen gestemmt werden, sondern es bedarf eigenen, nachhaltigen Strukturen, die ein Fortbestehen und eine Weiterentwicklung des Datenangebots gewährleisten. Damit könnte der Raum Regensburg seine Rolle als Vorreiter im Bereich KI allerdings weiter ausbauen und ein Angebot schaffen, das es in dieser Form nur in wenigen Ballungszentren, nicht aber in vergleichbaren Regionen gibt. Zudem kann dadurch die Zusammenarbeit der relevanten Stakeholder weiter gestärkt werden und es kann eine Basis für konkrete projektbasierte Umsetzungen von KI-Anwendungen geschaffen werden, die ein erhebliches wirtschaftliches sowie wissenschaftliches Potential beherbergen das nicht zuletzt für die ganze Region Regensburg einen Mehrwert bietet.

Vergleichsprojekte:

**MDM-Plattform – der Nationale Zugangspunkt für
multimodale Mobilitätsdaten**

<https://www.mdm-portal.de>



Berlin Mobility Data Hub

[ARTE: Berlin Mobility Data Hub \(tu-berlin.de\)](https://tu-berlin.de)

