

## Entwicklung eines Konzepts und Lastenhefts für eine "Szenariendatenbank" zur Bewertung der Sicherheitswirkung hochautomatisierter Fahrfunktionen

FA 82.719

Forschungsstelle: IKEM – Institut für Klimaschutz, Energie und Mobilität e. V., Berlin

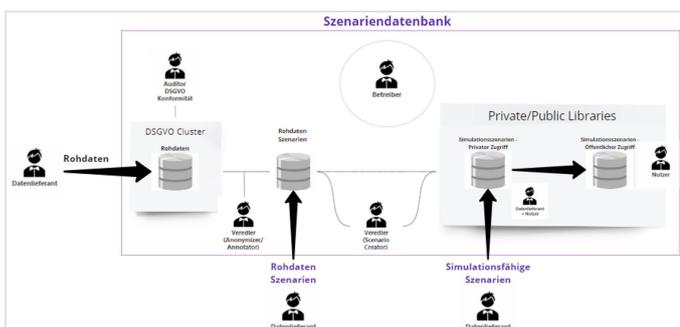
Bearbeiter: Siemon, B. / Wiegand, C. / Ehrich, H. / Klinge, A. / Krampitz, M. / Ehrich, H.

Auftraggeber: Bundesanstalt für Straßenwesen, Bergisch Gladbach

Abschluss: November 2021

### 1 Kurzbericht der Arbeitsinhalte

Der vorliegende Kurzbericht gibt einen Einblick in die verschiedenen Dimensionen der Konzeptionierung und Lastenheftentwicklung einer kooperativen Szenariendatenbank. Die Ausarbeitung gliedert sich in die Teilbereiche rechtswissenschaftlicher, wirtschaftswissenschaftlicher und technischer Rahmen- und Umsetzungsbedingungen einer nutzenmaximierenden Szenariendatenbank.



**Bild 1: Use-Case der zu entwickelnden Szenariendatenbank**

#### 1.1 Ökonomische Rahmenbedingungen

Zur Analyse der Marktarchitektur der Szenariendatenbank wurden die verschiedenen Stakeholder zuerst identifiziert und anschließend in die verschiedenen Kategorien öffentlich, privat sowie privat-öffentlich geclustert. Aus dieser Kategorisierung konnten die wichtigsten Akteure und deren Betriebs- sowie Nutzerinteressen identifiziert werden. Diese dienen als Grundlage für den Aufbau des Rollenmodells, des zugrundeliegenden ausdifferenzierten Anwendungsfalls und für die Zusammenstellung der rechtlichen, ökonomischen sowie technischen Anforderungen an die einzelnen Rollen. Für einen erfolgreichen Betrieb der Szenariendatenbank wurde ein Use-Case Szenario mit fünf Rollen (Datenlieferant, Veredler, Auditor und Betreiber) entwickelt. Der Datenlieferant trägt bei ausreichender Incentivierung zur Zulieferung der Daten, welche zur Erstellung der Szenarien benötigt werden, bei. Der Veredler labelt/anonymisiert daraufhin die zur Verfügung stehenden Datensätze, um eine Einspeisung als Simulationsszenarien zu

ermöglichen. Der Betreiber ist neben der Gewährleistung eines reibungslosen technischen Betriebs (Hosting, Nutzerverwaltung, Lizenzmanagement, Software-Releasepflege, Backup) verantwortlich für die finale Auswahl sowie die Festlegung der Szenarien. Neben den erforderlichen Kompetenzen zum Umgang mit annotierten Daten muss der Betreiber die Erstellung von Testumgebungen und die Analyse von Testdaten ermöglichen. Der Auditor prüft die Szenarien auf ihre DSGVO-Konformität beziehungsweise ihre Qualität und kann unabhängig vom Betreiber agieren. Die Nutzer der Datenbank können je nach Stakeholdercluster daraufhin die zur Verfügung stehenden Simulationsszenarien zur Sicherheitsprüfung autonomer Fahrfunktionen, Ausweitung der technischen Kompetenz oder zur Identifizierung von Infrastrukturbedarf nutzen. Es besteht die Möglichkeit, dass ein Stakeholder mehr als eine Rolle übernimmt. Vorstellbar ist, dass ein Stakeholder zum Beispiel zugleich die Rolle des Betreibers und Veredlers oder des Datenlieferanten und des Nutzers ausfüllt. Die Rollenbesetzung ist in jedem Fall abhängig von dem jeweiligen Partizipationsanreiz oder der jeweiligen Hürde. In einer umfangreichen Expert:innenbefragung mit 42 Teilnehmenden aus den verschiedenen Stakeholdergruppen wurden Anreize und Hürden der Datenbankpartizipation evaluiert und auf dieser Basis mögliche Rollenbesetzungen im hypothetischen Betrieb abgeleitet. In einem aufbauenden Workshop mit 24 Teilnehmenden konnten die Ergebnisse validiert und angepasst sowie die abgeleiteten Finanzierungs- und Betreibermodellvorschläge mit Branchenexpert:innen analysiert werden. Die Ergebnisse der Befragung und des aufbauenden Workshops lassen sich unter folgenden Kategorien zusammenfassen:

#### Wahrscheinlichste Rollenbesetzung im Kernbetrieb

- Veredler: Private beziehungsweise öffentlich-private Institutionen (KI-Unternehmen, Technologie Start-ups).
- Datenbankbetrieb: Privater Sektor (Toolhersteller, IT-Infrastruktur-Unternehmen) beziehungsweise beauftragt/zuschussfinanziert durch den öffentlichen Sektor.
- Auditor: Private beziehungsweise öffentlich-private Institutionen (Bundesbehörde, technischer Dienst), wobei insbesondere die Überprüfung der DSGVO-Konformität durch einen nach ISO zertifizierten Auditor erfolgt.
- Nutzer/Datenlieferant: alle Cluster abhängig von der jeweiligen Incentivierung und des Nutzerzugangs.

Finanzierungs- und Betreibermodellvorschläge für kosten-deckenden Betrieb

- Zuschussfinanzierung beim initialen Aufbau der Datenbank (CAPEX).
- Finanzierung der laufenden Kosten (OPEX) durch Kombination aus konstantem Mitgliedsbeitrag und Paketierung. Weitere Möglichkeiten wären auch Pay per Dataset Zahlungsmodalitäten.
- Option der Kommerzialisierung von Inhalten im späteren Betrieb durch beispielsweise Nutzung von Simulationen zur Verknüpfung von verschiedenen Szenarien sowie die Veredelung/Verknüpfung von Rohdaten.
- Schaffung von Anreizen zur Datenlieferung durch Vergütungssystem mittels Punktegutschriften, welche zum Erwerb neuer Daten beziehungsweise Szenarien genutzt werden können.
- Möglichkeit, ein modulares System zu integrieren, mit der Option weitere Dienstleistungen dazuzubuchen (beispielsweise integrierte Anonymisierung).
- Absicherung für Hersteller von neu entwickelten Fahrfunktionen.
- Zugang zu IT-Infrastruktur und vereinfachtes Datenmanagement.
- Flexible Paketierungsmöglichkeit des Use-Case ermöglicht Differenzierung der Nutzerzugänge und spricht dadurch unterschiedliche Stakeholdergruppen gleichermaßen an.
- Sicherstellung der DSGVO-Konformität.

Um den Vergleich zu bereits etablierten Szenariendatenbanken herzustellen, wurden mit den Betreibern der Datenbanken ENVITED Marketplace des eingetragenen Vereins Automotive Solution Center for Simulation und SafetyPool von Deepen AI und WMG University of Warwick Interviews geführt. Dabei wurden die Rollenverteilungen, Wertversprechen, Schlüsselressourcen und Finanzierungsmechanismen untersucht. Ähnlichkeiten bestehen in der Akzeptanz aller gängigen Dateiformate sowie in der Freigabe der Veröffentlichung simulationsfähiger Szenarien durch die jeweiligen Datenlieferanten und dem Ziel, eine Vielzahl an Simulationsszenarien zum Verkauf/Tausch anzubieten.

Während die Datenbank von SafetyPool eine internationale Ausrichtung verfolgt und Anreize setzt durch einen kostenlosen Zugang und die Verwendung eines Punktesystems, beschränkt der ENVITED Marketplace den Zugang auf Stakeholder mit Expertise in AD/ADAS Systemen und zeichnet sich insbesondere durch seine Vereinsstruktur aus. Im Vergleich zu den Kernfunktionalitäten dieser bereits bestehenden Datenbanken, zeichnet sich der Use-Case aus Nutzersicht durch zusätzliche Kernfunktionalitäten aus. Dazu zählen unter anderem:

- Generierung einer Datengrundlage für Gesetzesänderungen, insbesondere die gegebenenfalls erforderliche regulatorische Nachbesserung und Feinsteuerung und das Ableiten von Verkehrssicherheitsmaßnahmen.
- Bereitstellung von Grundlagen zur Ausweitung der technischen Kompetenzen.
- Identifizierung des Bedarfs von Infrastruktur, technischen Diensten, KI-Systemen, Tools und weiteren Geschäftsmodellen (F&E).

Dadurch schafft der Use-Case einen Zusatznutzen und setzt deutliche Anreize für die unterschiedlichen Stakeholdergruppen. Diese Ergebnisse der ökonomischen Betrachtung wurden stetig in den Bezug der technischen Ausgestaltung gesetzt, um einen technisch umsetzbaren und ökonomisch effizienten Anwendungsfall abzuleiten (vergleiche Bild 1)

### 1.2 Technische Rahmenarchitektur

Die unterschiedlichen Anwendungsfälle, die durch diese Datenbank unterstützt werden sollen, sind der Ausgangspunkt der technische Rahmenarchitektur. Insbesondere werden Szenarien im R&D-Bereich in Software-in-the-Loop, Hardware-in-the-Loop, Data-Reprocessing-Anwendungsfällen aber auch Realfahrten auf einem Proving Ground für hochautomatisiertes Fahren benötigt und tragen dazu bei, solche automatisierte Fahrfunktionen freigeben zu können. Weiter ist zu berücksichtigen, dass schon diverse Datenbanken (zum Beispiel Safety pool, Envited Market Place, Music Database, Gidas-Datenbank etc.) wie auch verschiedene Datasets (zum Beispiel NuScenes, ApolloScapes etc.) im Markt verfügbar sind. Des Weiteren werden Dienstleistungen angeboten, um Szenarien für dedizierte Plattformen bereitzustellen. Keine dieser Datenbanken, Datasets oder Dienstleistungen erfüllt alle Anforderungen bezüglich der erwähnten Anwendungsfälle und ist allein für die Absicherung hochautomatisierter Fahrfunktionen ausreichend. Unter dem Gesichtspunkt, dass eine große Anzahl von Szenarien erforderlich ist, um eine belastbare Sicherheitsaussage über (automatisierte) Fahrfunktionen treffen zu können, und die Anwendungsfälle bezüglich dieser Fahrfunktion sehr detailliert und auch herstellerepezifisch sind, muss die Datenbank eine tiefgehende Suchfunktionalität beinhalten. Diese Funktionalität muss den entsprechenden Anwendungen gerecht werden. Neben rechtlichen Rahmenbedingungen, die vorgeben, bei Verwendung von personenbezogenen Rohdaten die Privatheit zu schützen, stellen auch im Markt etablierte oder sich etablierende Standards wichtige Anforderungen an die Architektur. Die wichtigsten ASAM Standards, die durch eine Datenbank insbesondere im Kontext Simulation zu unterstützen sind, sind folgend aufgelistet:

- ASAM OpenDrive
- ASAM OpenSzenario

- ASAM OpenODD (basiert auf BSI PAS 1883)
- ASAM OpenLabel
- ASAM OpenXOntology

Da es sich um eine kollaborative Datenbank handeln soll, ist es unabdingbar, dass die Datenbank in einer Cloud gehostet wird. Die technische Rahmenarchitektur lässt sich vereinfacht in die folgenden Teilaspekte unterteilen:

- Anbinden anderer Datenbanken und Datasets
  - Modulares Konzept zur Integration unterschiedlicher Datenbanken
  - Ontologische Datenbankstruktur
- Einspeisen und Überführen von Datensätzen in die kollaborative Datenbank
  - Modulares Konzept zur Einspeisung von Rohdaten
  - Einspeisen von Rohdaten (DSGVO-Server)
  - Einspeisen von anonymisierten und annotierten Szenarien
  - Einspeisen von Simulationsszenarien
  - Einspeisen von funktionalen und abstrakten Szenarien
- Nutzerzugang
  - Konfigurierbare Dashboards zum Auswerten der Metadaten
  - Definition einer ODD (mit vorgegebener Taxonomie)
  - Suchanfragen basierend auf einer ODD
  - Einfache Suchanfragen unabhängig von der ODD
  - Suchanfragen, die auch die Beschreibungen in den Metadaten durchsuchen
  - Einbinden von Szenarien, die ausschließlich für den Nutzer in einem privaten Bereich verwendet werden
  - Möglichkeit des Einspeisens/Publizierens in die kollaborative Datenbank (s.o.)
- Metadaten
  - Format, basierend auf mindestens einer gewählten ODD-Taxonomie

- Erweiterungsmöglichkeiten durch den Nutzer (zusätzliche Tags/Beschreibungen)
- Beschreibung, was das Szenario beinhaltet (zum Beispiel ODD), aber auch nicht betrachtet wird (zum Beispiel keine 3D-Assets, keine Fußgänger)
- Nachvollziehen, woher diese Daten kommen und wie sie erstellt wurde (zum Beispiel auf Basis von Expertenwissen oder basierend auf Messung einer Realfahrt oder basierend auf der PCM der Gidas-Datenbank)
- Bewertungsschema – Wie wichtig ist dieses Szenario im Kontext der Datenbank aber auch im Kontext der, durch den Nutzer definierten ODD

### – Rechtemanagement

- Datasets oder zusätzliche Datenbanken nicht zwingend für jeden zugänglich, da entsprechende "Lizenzen" verfügbar sein müssen
- Privater Bereich für eigene Daten und Analysen

### – Billing-Service

## 1.3 Rechtliche Rahmenbedingungen

Die Analyse des Rechtsrahmens setzt zunächst an der Erörterung des rechtlichen Nutzens der Datenbank an. Die Szenariendatenbank ist aus Sicht der Hersteller und Entwickler insbesondere für die Erstellung eines Sicherheitskonzepts zwecks Zulassung automatisierter Fahrfunktionen und Einhaltung von Complianceanforderungen (angesichts von Haftungsrisiken, Produktsicherheit etc.) von rechtlicher Relevanz. Weiterhin sind die in der Szenariendatenbank bereitgestellten Informationen für die gesetzliche Feinsteuerung im Bereich der Zulassung automatisierter Fahrfunktionen sowie das Vorantreiben von Formen der deregulierten Selbstregulierung wie beispielsweise die Entwicklung von bereichsspezifischen Standards bedeutsam. Anschließend werden die aus datenschutzrechtlicher Sicht bedeutsamen Rechtfertigungsgründe für Datenverarbeitungen personenbezogener Daten, Einwilligung, Interessenabwägung und die Erhebung von Halterdaten aufgrund der neuen gesetzlichen Rechtfertigungsgrundlage in § 1 g StVG sowie die Besonderheiten und Anforderungen mit Blick auf die Privilegierung von Forschungsdaten dargelegt. Schließlich werden mögliche Rechtsformen mit Blick auf den Datenbankbetrieb erörtert. In Betracht kommen aufgrund haftungsrechtlicher Erwägungen der eingetragene Verein und die GmbH (oder UG). Auch die unter Umständen aufgrund der steuerlichen Begünstigung und mit Blick auf Forschungsdaten vorteilhafte Inzidenzwirkung, sowie mit Imagevorteilen verbundene sinnvolle Einstufung als "gemeinnützig" wird erörtert. Schließlich wird auch die

Bedeutsamkeit IT-sicherheitsrechtlicher Vorgaben erläutert. Aufbauend auf die Analyse der rechtlichen Rahmenbedingungen werden praktische Handlungsvorgaben und Regulierungspotenzial für die Umsetzung der Datenbank abgeleitet.

Handlungsempfehlungen aus rechtlicher Sicht:

- IT-Sicherheit ist von Anfang an mitzudenken.
- Die Datenbank sollte, soweit es dem relevanten Informationsgehalt nicht abträglich ist, nicht mit personenbezogenen Daten gefüttert werden.
- Hilfsweise (falls dennoch personenbezogene Daten in Datenbank einfließen) sollten die Datenverarbeitungsprozesse DSGVO-konform ausgestaltet werden.
- Sämtliche Verarbeitungs-/ Informations-/ und Abwägungsprozesse sind zudem umfassend (mit Blick auf Haftungsrisiken über die allgemeine Rechenschaftspflicht hinaus) zu dokumentieren.
- Das Konzept des Forschungsvorhabens (insbesondere Fragestellung, Verantwortlichkeiten, herangezogene Datenarten, gegebenenfalls Abwägungsgründe, Methodik, Gemeinschaftsnutzen und die Veröffentlichung der wesentlichen Ergebnisse) und das Ergreifen geeigneter Garantien gemäß Art. 89 Abs. 1 und Abs. 2 DSGVO sollte transparent dargestellt werden, um in den Genuss der datenschutzrechtlichen Privilegierungen von Forschungsdaten zu kommen.
- Wenn die Datenbank privat betrieben wird, ist in dem Zusammenhang darauf zu achten, dass die rein kommerzielle Nutzung und der Bereich der Forschung und Entwicklung voneinander getrennt sind, sodass sich das Aufziehen einer ausgelagerten Forschungsdatenbank empfiehlt. Bei privater Finanzierung oder privaten Eigeninteressen ist ein Konzept zu entwickeln, welches eine direkte Einflussnahme auf den Erkenntnisprozess (zum Beispiel durch Weisungen) ausschließt. Daneben ist darzulegen, dass private (zum Beispiel wirtschaftliche) Interessen das Forschungsinteresse nicht dominieren.
- Insgesamt ist es lohnenswert, im kooperativen Austausch mit der verantwortlichen Datenschutzbehörde zu stehen.
- In Betracht kommt aus Haftungsgründen aus diesem Grund entweder die Ausgestaltung als GmbH oder als eingetragener Verein, wobei der GmbH wohl langfristig der Vorzug einzuräumen ist. Die Anerkennung als gemeinnützige GmbH hat zahlreiche Vorteile, ist aber nur dann denkbar, wenn das Gewinnausschüttungsverbot durch ein passendes Finanzierungsmodell aufgefangen wird. Vor der Gründung eines Vereins oder einer GmbH, welche die Privilegien des § 51 AO auf-

grund ihrer Gemeinnützigkeit erhalten soll, ist es ratsam, bereits vorab mit den zuständigen Finanzbehörden Kontakt aufzunehmen, um so sicherzustellen, dass die Anforderungen des Gemeinnützigkeitsrechts erfüllt sind.