
Förderung im Programm „Neue Fahrzeug- und Systemtechnologien (NFST)“

SenseTrAI

Verbundprojekt: Sensorintegrierte 3D-Druck- Leichtbauteile und KI-unterstützte Zustandsüberwachung für Bahntechnikkomponenten und -systeme

Schlussbericht

Förderkennzeichen	19I21033B
Laufzeit	01.08.2021 – 31.01.2025
Teilvorhaben	Entwicklung und Erprobung eines Systems für Management, Analyse und Verwertung von Systemdaten
Verfasser	Martin Schweikert, Michael Heinlein vedisys AG, Dieselstraße 12, 64347 Griesheim Sebastian Baumann, Christian Preusche DATAbility GmbH, Karlsbader Str. 10, 64295 Darmstadt (im F&E Unterauftrag)



Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Inhaltsverzeichnis

Inhaltsverzeichnis	II
I.1 Aufgabenstellung	2
I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde	3
I.3 Planung und Ablauf des Vorhabens	3
I.4 Wissenschaftlicher und technischer Anknüpfungsstand	5
I.5 Zusammenarbeit mit anderen Stellen	6
II.1 Verwendung der Zuwendung und des erzielten Ergebnisses	8
AP 2 Use Case Entwicklung	8
AP5 Entwicklung eines PHM-Systems	10
AP5.1 Anforderungen Datenbank	11
AP5.2 Implementierung und Umsetzung des Datenbanksystems	13
AP5.3 Erweiterung von Analytik-Services	15
AP5.4 Test und Evaluierung des Datenbanksystems	17
AP5.5 Inbetriebnahme des Datenbanksystems	18
AP5.6 Anforderungsanalyse für die Datenanalytik	19
AP5.7 Datendeskription	20
AP5.8 Algorithmenentwicklung PHM	21
AP5.9 Verifikation und Validierung der Datenanalytik	24
AP6 Demonstrator	25
AP7 Feldtest & Nutzenbewertung	27
II.2 Erläuterung der wichtigsten Positionen des zahlenmäßigen Nachweises	29
II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit	29
II.4 Voraussichtlicher Nutzen	29
II.4 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen	30
II.6 Erfolgte oder geplanten Veröffentlichungen der Ergebnisse	30

Teil I

Kurzdarstellung

I.1 Aufgabenstellung

Die Instandhaltung von Schienenfahrzeugen ist derzeit stark auf fristengesteuerte Revisionen oder die reaktive Reparatur von Ausfällen ausgelegt. Dies führt zu ineffizienten Stillstandzeiten, da Züge für Wartungsarbeiten aus dem Betrieb genommen werden müssen, auch wenn keine unmittelbare Notwendigkeit besteht, oder ungeplante Ausfälle zu hohen Kosten und Pönalen (Strafen) führen. Die Digitalisierung und neue Technologien wie Künstliche Intelligenz (KI) bieten hier Potenziale für eine prädiktive Analytik, bspw. für eine vorausschauende Instandhaltung (Predictive Maintenance), die sich jedoch im Bereich Bahntechnik noch in den Anfängen befindet.

Zu den aktuellen Herausforderungen von Eisenbahnverkehrsunternehmen (EVU) zählen:

- Hoher Aufwand für Nachrüstung und Zertifizierung von digitalen Systemen in bestehenden Zügen durch Hersteller, was die Kosten für EVUs erhöht und insbesondere kleinere EVUs unattraktiv macht.
- Heterogene Flotten bei vielen EVUs (verschiedene Hersteller und Baujahre) erschweren den flächendeckenden Einsatz herstellerunspezifischer Lösungen.
- Begrenzte Datenverfügbarkeit und -qualität: Hersteller bieten bis dato vermehrt Datenvisualisierungen und Reporte an, und die Nachrüstung zusätzlicher Sensorik ist komplex, kostenintensiv und birgt Herausforderungen bei der Integration sowie der Datenverarbeitung und -speicherung in den bestehenden Informationssystemen.
- Mangel an nachvollziehbaren Datenanalytiksystemen (Explainable AI), die bspw. eine transparente Diagnose und Prognose von Systemzuständen ermöglichen und das Vertrauen der Anwender in die Ergebnisse aufbauen.
- Fehlende gesamtheitliche, fahrzeugtyp- und herstellerübergreifende Ansätze, die ohne teure Nachrüstungen auskommen und die betriebliche Erfahrung aktiv einbeziehen.

Im Rahmen des Verbundvorhabens SenseTrAI_n werden Leichtbautechnologien durch die Kombination additiver Fertigungsverfahren mit bauteilintegrierter Sensorik sowie einer KI-gestützten Datenmanagement und -auswertung entwickelt.

Das Projekt SenseTrAI_n zielt darauf ab, innovative Lösungen für die prädiktive Analytik, im Speziellen für die zustandsbasierte, prädiktive Instandhaltung von Schienenfahrzeugkomponenten zu entwickeln und zu qualifizieren. Das Ziel des Verbundprojektes ist die Entwicklung einer neuen, gesamtheitlichen Lösung zur drahtlosen und kontinuierlichen Zustandsüberwachung sensorintegrierter, additiv gefertigter Bauteile. Neben entsprechenden LPBF-Prozessketten wurden sowohl geeignete Sensor-, Datenmanagement-, KI-Analyse- sowie Monitoringkonzepte entwickelt. Abschließend wurde ein Prototyp im Zugbetrieb integriert, verprobt und evaluiert. Das Vorhaben kombiniert die Vorteile der additiven Fertigung (3D-Druck) mit integrierter Sensorik und intelligenten KI-Algorithmen, um in diesem Teilvorhaben ein gesamtheitliches System für Management, Analyse und Verwertung von Systemdaten der Bahntechnik zu entwickeln und zu erproben.

Die zentralen Aufgabenstellungen umfassten folgende Aspekte:

1. Entwicklung und Qualifizierung von Verfahren zur Integration von Sensoren in metallische Bauteile mittels additiver Fertigungsverfahren (Laser Powder Bed Fusion, kurz LPBF). Dies ermöglicht die wirkstellennahe und hochpräzise Überwachung von Betriebszuständen, indem Sensoren direkt in sicherheitskritische Strukturbauteile eingebettet werden. Dadurch entfallen aufwendige externe Sensoranbringungen und Verkabelungen, und die Messdatenqualität wird signifikant verbessert.
2. Aufbau einer abgestimmten, nachrüstbaren Datenanalytik-Infrastruktur zur drahtlosen Übertragung, zentralen Speicherung und automatisierten Auswertung der gewonnenen Messdaten. Hierbei liegt der Fokus auf der Skalierbarkeit des Datenbanksystems und der Sicherheit der Kommunikationskanäle.
3. Entwicklung intelligenter KI-Algorithmen für die Diagnose und Prognose des Gesundheitszustandes von Bauteilen (Explainable AI). Diese Algorithmen sollen Systemexperten bei der Bewertung unterstützen und

transparente, nachvollziehbare Aussagen zur prädiktiven Instandhaltung sowie zur Optimierung lasttragender Strukturen ermöglichen.

4. Entwicklung und Etablierung von Standards für den Einsatz von 3D-Druck, Sensorik und KI in der Bahntechnik, um innovative Nachweiskonzepte zu ermöglichen, die fahrzeugtyp- und herstellerübergreifend einsetzbar sind und die Notwendigkeit teurer Nachrüstungen durch Zughersteller reduzieren.
5. Aktive Einbindung der betrieblichen Erfahrung von Servicedienstleistern und EVUs, um praxisnahe und effektive Lösungen zu gewährleisten.

Es ist zu erwähnen, dass der zuerst genannte Punkt dem Teilvorhaben der Partner ME-Meßsysteme und Fraunhofer-Institut für Lasertechnik obliegt, die nachfolgenden zwei Punkte im Speziellen für das in diesem Schlussbericht adressierte Teilvorhaben der vedisys AG aufgestellt sind sowie die letzten beiden Punkte Teilvorhaben-übergreifend behandelt wurden.

I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Vorhaben SenseTrAI_n ist ein Verbundvorhaben der Partner ME-Meßsysteme GmbH, welcher die Konsortialführung übernahm, Fraunhofer-Institut für Lasertechnik (ILT) und vedisys AG. Als assoziierte Partner waren die DATAbility GmbH (im F&E Unterauftrag der vedisys AG) sowie die DB Systemtechnik GmbH im Vorhaben tätig. Das Vorhaben wurde im Rahmen des Fachprogramms „Neue Fahrzeug- und Systemtechnologien“ der Programmsäule „Innovative Fahrzeuge“ vom Bundesministerium für Wirtschaft und Energie gefördert.

Die ursprüngliche Projektlaufzeit war vom 01.08.2021 bis zum 31.07.2024 geplant. Aufgrund von Verzögerungen bei der Aufzeichnung und Bereitstellung realer Messdaten aus Testfahrten wurde eine kostenneutrale Verlängerung bis zum 31.01.2025 genehmigt.

Während die anfängliche Bestandsaufnahme sowie die Anforderungsermittlung gemeinsam von den beteiligten Projektpartnern durchgeführt wurde, befasste sich die vedisys AG im Speziellen mit der Datenmanagementlösung. Aufgrund dieser datengetriebenen Entwicklungen war eine Bereitstellung von Daten aus dem Fahrzeugbetrieb notwendig. Öffentlich zur Verfügung stehende Ersatzdaten konnten wertvolle Erkenntnisse für die Entwicklung eines übergreifenden Datenmanagementsystems sowie hinsichtlich der Leistungsfähigkeit datenbasierter Modelle liefern. Jedoch waren diese für die Arbeit am ursprünglichen Untersuchungsfokus notwendig - ungünstiger Weise jedoch über große Teile der Projektlaufzeit jedoch nicht verfügbar. Die vedisys AG hatte daher Änderungen am Projektplan erarbeitet und Maßnahmen vorgeschlagen, mit denen auf die Datenlieferungen der Konsortialpartner verzichtet werden könnte, ohne die primären Projektziele zu gefährden.

Im Vergleich zur ursprünglichen Vorhabenbeschreibung wurden daher auch alternative Daten aus dem Betriebsbereich eines Eisenbahnverkehrsunternehmens herangezogen. Dieser Vorschlag wurde vom Projektträger sowie vom Konsortialführer positiv aufgenommen. Im Nachfolgenden werden daher die Ergebnisse der ursprünglich geplanten sowie geänderten Vorhabenbeschreibung als gemeinsames Projektergebnis vorgestellt.

I.3 Planung und Ablauf des Vorhabens

Das Vorhaben wurde in zwei Teilvorhaben untergliedert, die sich einerseits mit der Entwicklung von Hardware und Prozessen sowie der Entwicklung und Erprobung eines Systems für Management, Analyse und Verwertung von Systemdaten beschäftigen. Das gesamte Arbeitsprogramm gliederte sich in die folgenden sieben Arbeitspakete:

- AP 1: Koordination, Dokumentation und Öffentlichkeitsarbeit
- AP 2: Bauteilauswahl und Use-Case Entwicklung
- AP 3: Entwicklung geeigneter Sensorik
- AP 4: Entwicklung einer in-situ LPBF-Integrationsstrategie
- AP 5: Entwicklung und Erprobung eines Systems für Management, Analyse und Verwertung von Systemdaten

- AP 6: Demonstrator
- AP 7: Feldtest und Nutzenbewertung

Abbildung 1 stellt den geplanten zeitlichen Verlauf der Arbeiten in einem Balkendiagramm dar. Mit einer sechsmonatigen kostenneutralen Verlängerung wurde das Projekt zum 31.12.2019 abgeschlossen.

Projektmonate	3	6	9	12	15	18	21	24	27	30	33	36	39	42
Projektjahr	Jahr 1			Jahr 2						Jahr 2			Verlängerung	
AP1 Koordination, Dokumentation und Öffentlichkeitsarbeit														
AP 2 Use-Case Entwicklung														
Meilenstein 1: Spezifikation/Pflichtenhefte														
AP 3 Sensorik (Teilvorhaben 1)														
AP 3.1 Entwurf Multisensor														
AP 3.2 Entwurf Energieversorgung														
AP 3.3 Entwurf Übertragungstechnik														
AP 3.4 Implementierung Sensor														
Meilenstein 2: Sensorentwurf														
AP 4 LPBF-Integration (Teilvorhaben 1)														
AP 4.1 Konstruktion/Bauteiladaptation														
AP 4.2 Prozessbedingungen														
AP 4.3 Prozesskette Integration														
AP 4.4 Einfluss Prozessunterbrechung														
Meilenstein 3: Prozessentwicklung														
AP 5 Entwicklung eines PHM-Systems (Teilvorhaben 2)														
AP 5.1 Anforderungen Datenbank														
AP 5.2 Implementierung Datenbank														
AP 5.3 Erweiterung Analytik-Services														
AP 5.4 Test & Evaluierung Datenbank														
AP 5.5 Inbetriebnahme Datenbank														
Meilenstein 4: Implementierung & Standardisierung														
AP 5.6 Anforderungsanalyse Analytik														
AP 5.7 Datendeskription														
AP 5.8 Algorithmenentwicklung														
AP 5.9 Verifikation & Validierung														
AP 5.10 Entscheidungssystem														
Meilenstein 5: Funktionsfähigkeit & Entscheidungsregeln														
AP 6 Demonstrator														
AP 7 Feldtest & Nutzenbewertung														
AP 7.1 Feldtest & Prüfstandversuche														
AP 7.2 Nutzenbewertung & Handbuch														
Meilenstein 6: Nutzerhandbuch für additiv gefertigte Bauteile														

Abbildung 1: Zeitplan & Meilensteine

I.4 Wissenschaftlicher und technischer Anknüpfungsstand

Neben bekannten wissenschaftlichen Methoden konnte für das Vorhaben auf eine technische Basis für das Datenanalyse- und Datenmanagementsystem zurückgegriffen werden.

Das Ziel der Disziplin Prognostics and Health Management (kurz PHM) ist es, durch die erfassten Daten und Informationen und unter dem Einsatz (intelligenter) Algorithmen – bspw. durch Verfahren der künstlichen Intelligenz (z. B. für Bilderkennung sowie Text-/Semantikanalyse) und des maschinellen Lernens (z. B. für Regression und Klassifikation) – den aktuellen System- bzw. Bauteilzustand (Gesundheitszustand) abzuschätzen (Diagnose). In einem weiteren Schritt sollen diese Aussagen als Basis dienen, um die verbleibende Nutzungszeit (Restlebenszeit; engl. Remaining Useful Lifetime (kurz RUL)) bis zu einem Ausfall eines technischen Systems, wie eines Elektromotors, Getriebes oder Lagers sowie einer einzelnen Komponente wie Quer- und Längsträger in einer Struktur, vorherzusagen (Prognose). Basierend auf der Prognose können dann unerwartete Ausfälle (Stillstand, Nichtverfügbarkeit von Funktionen und/oder Defekte) frühzeitig durch geeignete Gegenmaßnahmen vermieden und der gesamte Instand- haltungs- und Wartungsprozess (z. B. inkl. der Ersatzteilbeschaffung und -logistik) optimiert werden. Neben dem enormen Potential Kosten durch die Vermeidung reaktiver Maßnahmen zur Instandsetzung (Ausfall bereits aufgetreten) einzusparen kann SHM/PHM ebenso zur Erhöhung der Systemsicherheit beitragen.

Führende System- und Anlagenhersteller bieten bereits unterschiedliche Monitoringsysteme mit einer dazugehörige Softwarelösung zur Datenauswertung an. Vorwiegend erfolgen aber die implementierten Diagnosen auf der Basis einfacher Schwellwertanalysen. Diese besitzen jedoch erhebliche Einbußen für eine Differenzierung von Degradationen sowie in der Genauigkeit und Aussagekraft bei der Betrachtung komplexer Systeme und Komponenten, bspw. für einen Kompressor oder ein lasttragendes Bauteil in dynamischer Einsatzumgebung auf einem fahrenden Zug. Dabei ist insbesondere zu beachten, dass gleiche Komponenten wie z. B. Elektromotoren, Lager oder Querträger auf unterschiedlichen Zügen unterschiedliche Ausprägungen (Individualität der Bauteile) aufweisen können. Darunter leidet derzeit noch der Aufbau von skalierbaren und allgemeingültigen Aussagen datenbasierter Modelle sowie die Interpretation der erzeugten numerischen Ergebnisse durch Menschen. Eine zentrale Fragestellung bleibt daher, inwieweit die Kenntnisse aktueller (Diagnose) sowie zukünftiger Systemzustände (Prognose) so aufbereitet werden können, dass sie als Basis für menschliche Entscheidungsträger*Innen dienen, um die Lebenszyklen von Systemen und Bauteilen sowie deren Wartungs- und Instandhaltungsmaßnahmen optimal auszunutzen.

Derartige Assistenzfunktionen fehlen derzeit noch in der Breite, da gewonnene Informationen über Zustände nicht weiter aggregiert und mit den Anforderungen an die aktuelle und zukünftige Betriebsführung abgeglichen werden. Das liegt einerseits daran, dass die PHM/SHM Analyseprozesse dabei vorwiegend datenbasierte Methoden verwenden, deren Anpassung und Optimierung für dedizierte Komponenten und Einsatzcharakteristiken derzeit noch einen erheblichen manuellen Aufwand darstellen. Im Fokus der Forschung und Entwicklung stehen aktuell bspw. Tätigkeiten zur Datenmanipulation ((Short-Time) Fourier Transformation, Wavelet Analyse sowie statistische Parameter z.B. Kurtosis, RMS...), zur geeigneten Merkmalauswahl (Cluster Analyse, Principal Component Analysis...), zur Modellierung einer eindeutigen Zustandsbestimmung anhand ausgewählter Merkmale sowie der Aufbau robuster Modelle für aussagekräftige Prognosehorizonte (Gaussian-Process-Regression, Neuronale Netzwerke...).

Dank einer offenen Struktur des Systems VEDIS®|ble der vedisys AG bot sich die Weiterentwicklung und Optimierung dieses Software-Pakets bezüglich einer Datenbankstruktur bzw. eines Datenbankschemas zur Verarbeitung von Zeitreihendaten aus den technischen Systemen (Zügen und Zugkomponenten) für EVUs und für die Verkehrsträger für dieses Projekt an. Das Ziel war der Machbarkeitsnachweis für die Eignung dieses Systems sowie Forschung und Entwicklung notwendiger Erweiterungen. Als Erweiterung waren für das Vorhabenziel Verknüpfungen des bestehenden Datenverwaltungssystems für Eventdaten mit kontinuierlichen Sensordaten sowie die Einbindung von Wartungsprognosen und Fehlerrückmeldungen (auch Texte) zusammen mit den Leistungsdaten der einzelnen Züge zu realisieren. Damit sollen Rückschlüsse auf mögliche Ursachen einer Überbeanspruchung von Fahrzeugteilen betrachtet werden können und somit nicht nur vorausschauend

Wartungsprozesse optimiert, sondern möglichst auch Probleme in der Streckeninfrastruktur identifiziert werden können (z.B. Streckenabschnitte in denen Bauteile immer überbeansprucht werden).

Das Software-Paket VEDIS® ist bereits modular aufgebaut und arbeitet in der Grundstruktur datenbankunabhängig, sodass die Erweiterung, Anpassung und Optimierung einer geeigneten Datenbank modular integrierbar waren. Das Software- Paket wurde im Projekt an die individuelle Betriebsanforderungen der unterschiedlichen Use Cases angepasst, um daraus Standardisierungsmöglichkeiten ableiten zu können.

Die Analysemodule von DATAbility (Anpassung und Weiterentwicklung im Rahmen dieses Vorhabens) boten aufbauend auf dem neu entwickelten Datenbanksystem einen logischen Baustein für genau die Analysen der Daten, um auch zukünftig bestehenden und weiteren Kunden erweiterte Auswerte-, Überwachungs- und Analysefunktionalitäten bieten zu können. Um auch die Konfigurierbarkeit des Datenbank- und Analytiksystems für die Anbindung an zukünftige EVUs und Verkehrsträger nach individuellen Anforderungen möglichst skalierbar zuzulassen, wurde ein Vorgehen für einen standardisierten Prozesse etabliert, um auf neue Use Cases geeignet reagieren zu können.

DATAbility hat in das Projekt ein eigenes Rapid Prototyping Framework für den agilen Aufbau und Test (hybrider) Prozessketten für komplexe Datenanalytikaufgaben eingebracht. Aus Vorarbeiten des Unternehmens resultierten bereits fundierte Expertise, ein tiefes Prozessverständnis und ein breiter Methodenpool zur Diagnose und für die Zustandsschätzung zum Einsatz für die Bauteilüberwachung sowie für die prädiktive Wartung und Instandhaltung bei Maschinen- und Anlagensystemen, bspw. für Elektromotoren, Getriebe, Kupplungen und Lager. Zudem bestanden Aufträge des Unternehmens für Entwicklungen im Bereich Mobilität und Logistik.

Im Rahmen des Projekts wurden für Recherche von Informationen und dem Abgleich wissenschaftlicher und technischer Ergebnisse rollierend Informationsdienste genutzt, bspw. SpringerLink und ScienceDirect sowie öffentliche Patentinformationssysteme. Zusätzlich wurden Marktstudien und Veröffentlichungen von Branchenverbänden wie Verband Deutscher Verkehrsunternehmen (VDV) und Branchenverband der deutschen Informations- und Telekommunikationsbranche (bitkom e. V.) herangezogen. Die interne Projektdokumentation erfolgte über ein zentrales Dokumentenmanagementsystem (MS Teams/SharePoint etc.), in dem alle Projektpartner Zugriff auf Protokolle, Arbeitsergebnisse und Zwischenberichte hatten. Die technische Dokumentation wurde laufend gepflegt. Zur Verbreitung der Projektergebnisse wurden ein Projektflyer sowie Beiträge auf der Projektwebsite und in sozialen Medien erstellt. Darüber hinaus wurden Ergebnisse auf Fachveranstaltungen präsentiert und eine Fachpublikation über einen Branchenverband veröffentlicht. Die Projektergebnisse werden langfristig bei den beteiligten Partnern archiviert und sollen in künftige Entwicklungs- oder Standardisierungsprozesse einfließen.

1.5 Zusammenarbeit mit anderen Stellen

Im betrachteten Projektzeitraum fand neben dem rollierenden Austausch mit den Konsortialpartnern auch ein inhaltlicher Austausch mit weiteren Akteuren aus den Bereichen Eisenbahnverkehr und Bahntechnik statt. Über das Projekt hinaus konnten so potentielle Verwertungen nach Projektende sowie Optimierungsmöglichkeiten im Rahmen der Projektlaufzeit diskutiert werden.

Teil II

Eingehende Darstellung

II.1 Verwendung der Zuwendung und des erzielten Ergebnisses

Nachfolgend werden die durchgeführten Arbeiten und die erzielten Ergebnisse dargestellt. Die entsprechenden Abschnitte orientieren sich an den in Kapitel I.3 aufgeführten Hauptarbeitspaketen AP2, AP5, AP6 und AP7.

AP 2 Use Case Entwicklung

In Arbeitspaket AP2 wurde zusammen mit der DATAbility GmbH sowie dem Anwendungsgeber der DB Systemtechnik AG geeignete Anwendungsfälle im Rahmen von Workshops erörtert. Als Basis lieferte DB Systemtechnik entsprechende Übersichten zu möglichen Bauteilen und relevanten Subsystemen aus der derzeitigen Praxis. Auf Basis eines ersten Technologie-Scouting während der Antragsphase wurden bereits erste Use Cases exemplarisch ermittelt. Diese wurden nun gemeinsam zu Projektbeginn detaillierter, insbesondere in Bezug auf die digitale Eignung für Daten- und Analyserrelevanz, bewertet und hinsichtlich der Eignung für eine weitere Erarbeitung von Konzepten bewertet. Zudem besteht ein entscheidendes Kriterium in der Bewertung eines Mehrgewinns für EVUs, um eine zukünftige Marktfähigkeit schon während der Konzeptionsphase sicherstellen zu können.

Im Ergebnis konnte eine Auflistung möglicher Anwendungsfälle für eine weitere Bewertung erstellt werden. Ein Auszug daraus ist nachfolgend im Zwischenbericht dargestellt. Neben den exemplarischen Anwendungsfällen aus der Antragsphase, welche den Radsatzlagerdeckel für eine Überwachung von Achskomponenten, die Türmechanik an den Einstiegen der Fernverkehrszüge sowie die Primärfeder- / der Schlingerdämpfer beinhalteten, konnten 12 weitere zusammengetragen werden. Zusätzlich zum Funktions- und Anwendungsgebiet der Instandhaltung wurden mit dem Zugbetrieb, der Infrastrukturüberwachung sowie der Betriebsführung weitere, ähnlich gelagerte Kategorien für den Anwendungsgeber identifiziert.

Tabelle 1: Auflistung und Kurzzusammenfassung identifizierter Anwendungsfälle zusammen mit dem Anwendungsgeber

ID	Kategorie	Name	Beschreibung & Ziele
1	Instandhaltung	Türen	Türstörung vorhersehen, um ungeplante Ausfälle und Pönalen zu reduzieren; Mechaniküberwachung durch Messsensorik für Kraft und Spannungen sowie weitere Regelparameter des Systems etablieren, bspw. durch Stromdaten des Türaufzuges über Information-BUS des Zuges
2	Instandhaltung	Toiletten	Ausfälle führen zu erheblichen Kundenunmut; neben mechanischen oder elektrischen Ursachen können aber auch menschliche Ursachen wegen extra herbeigeführter Verstopfung verursacht werden; Möglichkeit zur Komponentenüberwachung kann der Blower zur Vakkumerzeugung geben
3	Instandhaltung	Gap-Closer / Gap-Filler (Türen)	Zusätzliche Einrichtung an den Zügen, deren Inbetriebnahme recht jung ist; sofern es Störungen am Gap-Closer gibt, wirkt sich dies ebenfalls auf eine Türsperrung aus - ggfs. kann ein Zug damit nicht weiterfahren; Überwachung des elektromechanischen Aktuators erscheint neben der Messbox an Getriebe-Bauteilen sehr sinnvoll
4	Instandhaltung	Dämpferverschleiß	Mechanischer Verschleiß am Dämpfer führt zu Komfortreduktion im Fahrgastraum und kann zusätzlich Zugstabilisierungsprobleme mit sich bringen; Messungen von Beschleunigungen und Rollraten vorhanden. Daraus sollen Lastkollektive ermittelt und Aussage über Anomalien gegeben werden; Erfassung / Ermittlung der Anzahl von Schwingspielen könnte dahingehend ein aussichtsreichen KPI darstellen, die Ableitung daraus gibt Aussage über Dehnungsgeschwindigkeit geben
5	Instandhaltung	Gummibuchsen	Bauteil wird derzeit nicht aktiv überwacht und bietet damit hohes Digitalisierungspotential; visuelle Überprüfung erfordert erhebliche Montageaufwände; die Applikation von Sensorik über die Messbox bzw. die Integration im Gestänge am Einbauort des Bauteils soll Lastkollektive ermitteln lassen; belastbare Aussagen über den Gesundheitszustand des Materials und dessen Feder-Dämpfer-Eigenschaft sollen generiert werden
6	Instandhaltung	Klimaanlage / Elektrische Komponenten	Parallelen zu Toiletten-Komponenten → Elektromotorenantrieb zur Phasenstromüberwachung (minimalinvasiv); Bspw. Überwachung von Verdampfer / Kompressor erscheint sinnvoll; jedoch ist die Komponente nicht gut zugänglich und auf dem zugseitigen informations-BUS liegen aufgrund des Zulieferteils nicht viele Messdaten vor
7	Instandhaltung	Pantograf	Die Prüfung und Nachweisführung an diesem Bauteil ist aufgrund einer notwendigen Zertifizierung aufwändig und sensibel; maßgeblich für

			Bauteilstörungen und Ausfälle ist der Kohleblockverschleiß sowie zu ertragende Erschütterungen im Zusammenspiel der Infrastruktur mit Gleis und Oberleitung – werden Bewegungen nicht ausreichend gedämpft, kann Oberleitung Schaden nehmen, ist Anpressdruck an Oberleitung nicht ausreichend, erfolgt keine ausreichende Energiezuführung zum Zug; die Dämpfungsüberwachung mit einer Messbox scheint sinnvoll
8	Infrastrukturüberwachung	Streckennetzbewertung	Durch eine Messdatenaufzeichnung an Strukturbauteilen wie Achsen, Radsätzen, Drehgestell oder Wagenkasten sollen nicht nur Lastkollektive auf Zugkomponenten ermittelt, sondern auch der Zustand der Infrastruktur wie Trassen bewertet werden. Durch die intelligente Auswertung der Messsensorik soll ein kontinuierliches Streckenmonitoring ermöglicht werden, welches Anomalien und Drifts frühzeitig erkennt. Diese sollen Indikatoren für Fehlstellen an Rädern in Bezug auf Flachstellen, Spurbabweichungen an Gleisen oder Hohllagen im Gleisbett geben.
9	Betrieb	Nachweis von Pönalen	Vermeidung von Pönalen durch Nachweis von Komfortthematiken: Türstörung (wann, wie lange, Behebung, Nachweis Funktionalität), Ableiten von Kennzahlen und Qualitätskriterien (z.B. objektive Pünktlichkeit); Auswertung und Bereitstellung von Qualitätsdatenanalysen für die Bewertung von Bonus-Malus-Zuschreibungen
10	Betrieb	Verspätungsanalyse	Verspätungsverlaufsanalysen und Prognosen zur frühzeitigen Erkennung von Betriebsabweichung (Anomalien)
11	Betrieb	Verspätungsursachen	Erkennung von Trends, Korrelationen und Ursachen von Verspätungen und deren Auswirkungen auf den Betriebsablauf; Einbeziehung technischer Störungen in die Bewertung zur Charakterisierung möglicher Auswirkungen; Verwertung dieser Informationen für die Planung
12	Betrieb	Streckenauslastung	Nutzung von Fahrgastzählung zur Bewertung der Ausnutzung von Streckenabschnitten sowie der Belastung technischer Systeme (bspw. Türen, Toiletten, Sitze, Klimaanlage etc.)
13	Betrieb	Szenario-Planer	Entwicklung eines Tools zur Bewertung von Auswirkungen von Veränderungen im Fahrplan auf Anschluss- und Instandsetzungsmöglichkeiten
14	Betriebsführung	Multimodales Routing	Tarifübergreifendes Routing mit weiteren Verkehrsträgern Zug, Bus, Bahn, zu Fuß, Scooter, Car-Sharing und On-Demand-Verkehre zur Abwicklung betriebsbedingter Störungen

Mit diesem zusammengetragenen Anwendungs-Portfolio lassen sich eine Vielzahl von übergeordneten Anforderungen an die additive Fertigung und Integration der Messsensorik, aber auch an die zur Verfügung stehenden Informationen und zu messenden Daten sowie die Verarbeitungslogik mittels Algorithmen diskutieren.

Im Projektverlauf konnten ursprünglich geplante reale Datensätze aus Testfahrten für die geplanten Analysen jedoch nicht ausreichend und entsprechend des Zeitplans zur Verfügung gestellt werden. Der Analyseumfang im Projekt wurde daher insoweit erweitert, indem weitere Analytik-Services identifiziert und das Portfolio entsprechend ausgebaut wurde. Die nachfolgende Abbildung 2 zeigt den im Teilvorhaben konkret mit Daten analysierten Untersuchungsumfang bis Projektende. Dieses Vorgehen wurde mit dem Projektträger abgestimmt.

Bei den dargestellten Anwendungsfällen 1 bis 4 handelt es sich um das Management und die Analyse originärer Fahrzeugdaten in der Mobilität mit dem Fokus auf Zustandsvorhersage und Gesundheitsmanagement technischer Systeme (Prognostics and Health Management, PHM). Mit Anwendungsfall 5 wurden die Daten fachlich für die Verwertung von Text durch sogenannte Large Language Models erweitert. Dieses Datenformat fällt beispielsweise in der Wartung sowie im Beziehungsmanagement mit Kunden (bspw. bei Beschwerden und Kundenanfragen) an.

Anwendungsfall 6 beschäftigte sich daher mit der Analyse weiterer Umfelddaten in Bezug auf Netz- und Fahrpläne sowie Wetter für Routenempfehlungen. Anwendungsfall 7 beschäftigt sich mit Bestelldaten am Anfang der Wertschöpfungskette für eine Mobilitätsleistung. Die Bearbeitung der Anwendungsfälle 5 bis 7 brachte in das Projekt eine reisendenzentriertere Sicht für die Entwicklung und Erprobung eines Systems für Management, Analyse und Verwertung von Systemdaten aus der Mobilität ein.

Die Arbeiten wurden mit Erreichung des Meilensteins MS1 mit der Erstellung der Spezifikationen zu den einzelnen Use Cases in Bezug auf das Datenmanagement und die Auswertungsfunktionalitäten abgeschlossen. Im Nachfolgenden werden einzelne Aspekte der betrachteten Use Cases für die thematische Aufbereitung der Arbeitspakete aufgegriffen.



Abbildung 2: Identifiziertes Analytik-Service-Portfolio

AP5 Entwicklung eines PHM-Systems

Heutige Verkehrssysteme sind von einer enormen Heterogenität geprägt, wenig miteinander vernetzt sowie überwiegend regional und dezentral in Datenbanken verwaltet. Eine übergeordnete Informationszusammenführung oder Anreicherung erfolgt bisher, wenn überhaupt, nur punktuell, sodass die notwendige datentechnische Vernetzung zwischen relevanten Stakeholdern innerhalb eines EVUs sowie weiterhin Verkehrsverbänden, Mobilitätsanbieter, Service-, Wartungs- und Instandhaltungsdienstleister sowie Mobilitätsnutzende bis dato faktisch ausbleibt. Aufgrund der Vielzahl proprietärer Systeme sowie dem vorherrschenden Digitalisierungsbedarf der Verkehrsträger ist eine weitergehende Datennutzung, beispielsweise in Form von Services, ohne ein Paradigmenwechsel für die Anbindung und Vernetzung bestehender Daten- und Informationslandschaften nicht möglich.

Ein Data Mesh wurde für die Zusammenführung und Verknüpfung von Daten und Informationen im Vorhaben SenseTrAIIn gewählt. Das Data Mesh setzt die Merkmale unterschiedlicher Daten- und Informationsquellen in Beziehung. Übergeordnet dienen Mediatoren für die effiziente Abwicklung von Anfragen sowie zur Konsistenzhaltung der Daten. Die Anbindung zusätzlicher Datenanbieter, bspw. durch Nachrüstung bestehender Fahrzeuge mit Telematik (IoT) oder die Einbeziehung von Umweltdaten (bspw. Wetter, Buchungsdaten oder Passagierzählungen) ist ebenfalls möglich. Nicht zuletzt lassen sich unternehmensübergreifend Services aufbauen, synchron halten und weiterentwickeln.

Ein Data Mesh schafft moderne Möglichkeiten für Services und Datenprodukte. Insbesondere werden für Betreiber von Verkehrsträgern ein umfassenderes und dynamischeres Asset Management & Monitoring mit Abweichungsanalysen, Diagnosen und Prognosen ermöglicht. Datenkonsumenten können bedarfsgerecht mit angeforderten Merkmalen versorgt, über Änderungen oder Neuerungen im Datenraum sowie bei Services informiert und in die Lage der selbstständigen Nutzung des Datenraumes versetzt werden. Damit lassen sich Verwaltungsredundanzen insbesondere für Stammdaten sowie ineffiziente Medien- und Informationsbrüche vermeiden. Im Gegensatz zu konventionellen Architekturen, wie beispielsweise ein Data Lake oder Data Sump, können effizientere Zugriffs- und Bereitstellungszeiten erreicht werden.

Trotz aller Vorteile eines Data Mesh für die Anwendung in SenseTrAIIn ist der Aufwand für Aufbau und Pflege nicht außer Acht zu lassen. Insbesondere das Mapping, die Synchronisierung und Aktualisierung des Data Mesh erfordern eine fortlaufend intensive Zusammenarbeit aller Akteure wie Datenserviceanbieter mit sogenannten Product

Owner (DATAbility). Datendomänen von Datenbereitstellern (Deutsche Bahn und andere) sowie Datenplattformbetreiber (vedisys AG). Alle Beziehungen der Daten und Informationen sind bedarfsgerecht aufzubereiten und mit weiteren Meta-Informationen anzureichern, um die Potentiale der engen Verzahnung von Quellen und Produkten/Services nutzen zu können. Da im Vorhaben SenseTrAIIn dieses Paradigma von Grund auf aufgebaut und für bestehende Datensysteme nachrüstbar gestaltet wird, überwiegen jedoch aus Sicht der vedisys AG und der DATAbility GmbH die Vorteile. In Abbildung 3 ist ein exemplarischer Aufbau eines Data Mesh mit den dahinterliegenden vier Prinzipien dargestellt.

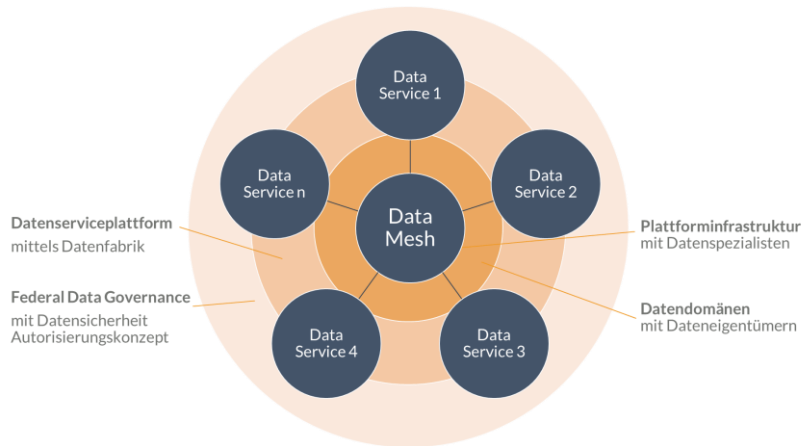


Abbildung 3: Schematische Aufbau eines Data Mesh Konzeptes für das Vorhaben

AP5.1 Anforderungen Datenbank

Für eine zentrale Verwaltung von Daten und Informationen für die entwickelten Anwendungsfälle und als Bindeglied der Teilsysteme einer Demonstrationslösung für das PHM System zur Zustandsüberwachung wurde von der vedisys AG ein geeignetes Datenbanksystem spezifiziert. Damit war es möglich, spezifische Anforderungen je Anwendungsfall in skalierbare, generalisierte Anforderungen bezogen auf ein adaptives Datenbanksystem zu überführen. Den grundlegenden Aufbau beschreibt die nachfolgende Abbildung 4 mit Input-Layer für unterschiedliche Datenquellen, dem Data Warehouse zur Verarbeitung und Speicherung sowie dem Output-Layer für die Zuführung in Services wie Reporting über Dashboards sowie Analysen und Modelle für Diagnosen und Prognosen.

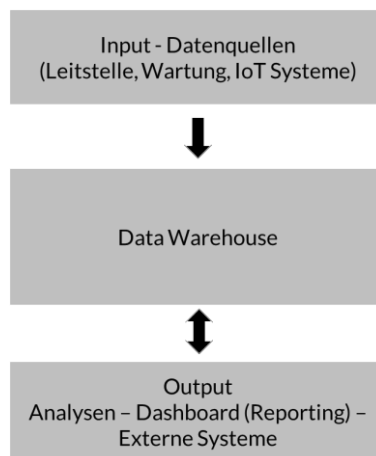


Abbildung 4: Struktur des Datenbanksystemkonzeptes

In Abbildung 5 ist die erstellte Zielarchitektur des Gesamtsystems und die Verknüpfung mit den Tätigkeiten im Projekt (Arbeitspaketen) schematisch dargestellt. Für die zentrale Verwaltung und Prozessierung der Daten wurde ein generalisiertes Datenbanksystem von der vedisys AG entwickelt. Das Analytikmodul der DATAbility GmbH

erlaubt die umfassende Untersuchung der gespeicherten Daten und erstellt anwendungsspezifische Dashboards und Reports für den Anwender. Da das Datenbanksystem generalisierte Schnittstellen unterstützt, ist es möglich, Daten aus unterschiedlichen Quellen einzubinden.

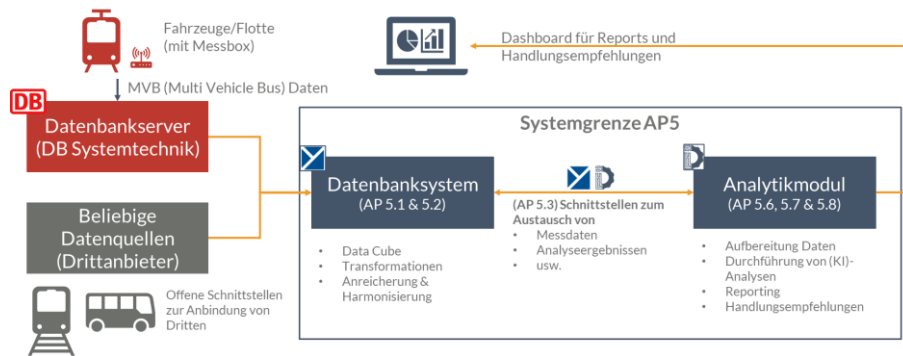


Abbildung 5: Schematische Zielarchitektur des Gesamtsystems

Seitens der vedisys AG wurde ein sogenannter Application Server aufgebaut. Dieser stellt die zentrale Komponente im Gesamtsystem dar und übernimmt bei allen Anfragen unter anderem die Rolle eines Mediators. Alle clientseitigen Anfragen werden dadurch bei Bedarf vollständig oder teilweise an weitere Systemkomponenten weitergeleitet. Diesbezüglich stellen die Services und Analysen der DATAbility GmbH eine weitere Systemkomponente dar. Die Definition der Services/Schnittstellen basiert dabei auf dem OpenAPI Standard 3.0.

Für eine fehlerfreie und verzögerungsarme Übertragung von bspw. Fahrzeugdaten an das Datenbanksystem, wurde im Speziellen auf ein bereits entwickeltes System der DB Systemtechnik zurückgegriffen. Dieses System ermöglicht durch eine Messbox die Aufnahme von Sensor- und Fahrzeugdaten sowie deren Übertragung an „Land“ während des Zugbetriebes (Telematikeinheit). Die Daten können an Land von einem Datenaustauschserver abgefragt werden. Da im Rahmen des Projektes keine abrupt auftretenden, systemkritischen Fehler in den Use Cases im Vordergrund stehen, wurde bei der Übertragung der Daten keine harte Echtzeit gefordert.

Darüber hinaus wird das System auf die Anbindung weiterer Drittanbieter von Daten und Informationen ausgelegt. Folgende Anforderungen wurden definiert:

- Generalisierbarkeit
- Performance wie Verarbeitungszeit für Anfragen (Lesen und Schreiben von Daten)
- Datensicherheit
- Redundanzen & Resistenz
- Mehrbenutzerbetrieb & Koordination
- Datenintegration & Datenstrukturierung
- Datenkonsistenz & Persistenz (auch transaktionell)
- Echtzeitfähigkeit bzgl. Speicherung, Bereitstellung, Analyse und Filterung

Dabei wurden insbesondere folgende Vorteile für bestehende Herausforderungen für das Vorhaben diskutiert und spezifiziert sowie für die Entwicklung und Implementierung berücksichtigt:

- **Datenkonsolidierung:** Die generische Datenmanagementlösung ermöglicht die Konsolidierung von Daten aus verschiedenen Quellen und Formaten an einem zentralen Ort. Dadurch wird es einfacher, Daten zu sammeln und auf sie zuzugreifen, unabhängig von ihrer ursprünglichen Struktur oder ihrem Format. Dies erleichtert die Zusammenführung bahntechnischer Event- und Zeitreihen-Systemdaten (Aggregation) für die Analyse.
- **Datenintegration:** Die generische Datenmanagementlösung bietet nun Funktionen zur Datenintegration, um unterschiedliche Datenströme und -formate zu harmonisieren. Damit bietet sie Standardisierungsmöglichkeiten für die Zusammenführung von Telematik und Betriebsinformationssystemen. Dahingehend können automatische Konvertierungsfunktionen oder

Mapping-Tools bereitgestellt werden, um Daten in ein einheitliches Format zu überführen und Kompatibilitätsproblem zu lösen.

- Skalierbarkeit: Die generische Datenmanagementlösung ist in der Lage, sowohl mit aktuellen als auch mit zukünftigen Datenströmen und -formaten umzugehen. Durch die Bereitstellung flexibler Datenstrukturen und Erweiterungsmöglichkeiten kann die Lösung an neue Datenanforderungen angepasst werden. Dadurch können kontinuierlich neue Datenquellen integriert und eine gesamtheitliche Systemlösung (Datenbankbereitstellung & Reporting) aufgebaut werden, ohne dass es zu größeren Hindernissen kommt.
- Datenqualität: Eine generische Datenmanagementlösung stellt Funktionen zur Überwachung und Verbesserung der Datenqualität bereit. Diese umfassen beispielsweise die Identifizierung und Bereinigung von fehlerhaften oder inkonsistenten Daten, die Erkennung von Duplikaten oder die Validierung von Daten gemäß vordefinierten Regeln. Durch die Gewährleistung einer hohen Datenqualität wird die Genauigkeit und Zuverlässigkeit der Analyse verbessert.
- Zugänglichkeit und Kollaborationsopportunität: Eine generische Datenmanagementlösung bietet Reporting-Möglichkeiten, über die Datenanalysten und andere Benutzer informiert werden. Durch die Unterstützung von Kollaborationsfunktionen können zudem Analyse erarbeiten, Erkenntnisse geteilt und gemeinsam Entscheidungen getroffen werden.

AP5.2 Implementierung und Umsetzung des Datenbanksystems

Das Datenbanksystem sollte im Aufbau und der zukünftigen Einsatzorientierung her in hohem Maße generisch aufgebaut sein. Im Projekt wurde im Entwurf und der Umsetzung daher darauf geachtet, dass das Datenbanksystem mit Hilfe eines modularen Aufbaus für die vorgegebene Anwendungsumgebung einfach konfigurierbar gehalten werden kann, um zusätzliche Datenquellen einfach anbinden bzw. auf Änderungen in den Originär-Datenquellen und -Schnittstellen schnell reagieren zu können. Die Vorteile dieser Flexibilität zeigten sich im Projekt bereits durch den Aufbau der alternativen Use Cases. Es wurde ein relationales Datenmodell aufgebaut, mit dem die innerhalb des Vorhabens generierten Daten von Fahrzeugen, deren Einsatzumläufen sowie Umgebungsdaten abgelegt und verarbeitet werden können. Hierzu wurden diverse Datenstrukturen einer Vielzahl bisher verfügbarer und zukünftig erwartbarer Parameter optimiert und generalisiert, damit einerseits eine Datenbankunabhängigkeit sowie die Konfigurierbarkeit für anfallende Messwerte und aufkommenden Event-Logs, deren Dimensionen und sonstiger notwendiger Stammdaten der Fahrzeuge ermöglicht werden kann.

Das aufgebaute generische Datenmanagementsystem besteht aus verschiedenen Komponenten und Elementen, die zusammenarbeiten, um Datenbanken zu verwalten, Backups durchzuführen und den Zugriff auf die Datenbanken zu ermöglichen.

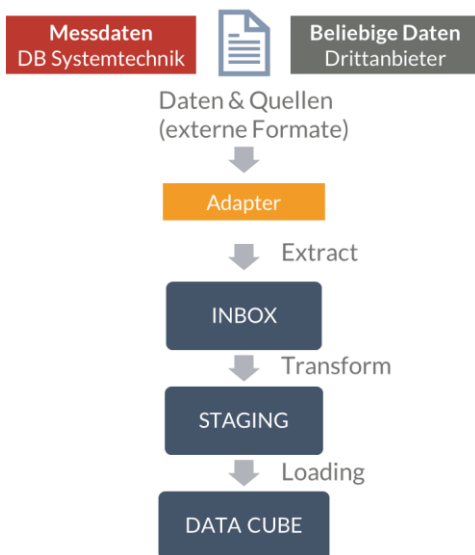
Der Application Server ist ein integraler Bestandteil von Web- und Unternehmensanwendungen. Er stellt ein Rahmengerüst bzw. eine Plattform dar, welches Anwendungen ausführen und konfigurieren kann sowie verschiedene Dienste bereitstellt, um den Betrieb des Datenbanksystems in einem Netzwerkumfeld zu unterstützen. Im Wesentlichen handelt es sich um eine Middleware oder auch Web-Service, die zwischen einer Benutzeroberfläche (z. B. einem Webbrowser) und den Datenbanken und/oder anderen Ressourcen, die von der Anwendung benötigt werden, vermittelt. Mit Hilfe des Application Servers können fachliche Dienste im Baukastensystem bereitgestellt werden. Das Rahmengerüst besitzt die Fähigkeit, zukünftig für die vedisys AG kundenseitige Kernfunktionalitäten des Unternehmens für Bereiche wie Bestellwesen, Produktion, Vertragsverwaltung und Abrechnung abzubilden und die Interdependenzen der einzelnen Bausteine auch im Sinne der Wartbarkeit zu standardisieren. Daneben ist die Plattform für das Vorhaben derartig zweckdienlich entwickelt worden, damit ebenso Endpunkte in einem hohen Abstraktionsgrad zur Datenintegration und Datenabfrage (auch pivotiert) bereitgestellt werden können. Für das aufgebaute Datenbanksystem ist es damit unerheblich, welche Art von Daten und Formate abgefragt, gespeichert und verarbeitet werden sollen oder für welche Use Cases diese einbezogen werden. Dies dient als Basis zum Betrieb der eigentlichen Analysefunktionen über Algorithmen und entwickelte Funktionalitäten (siehe AP 5.8 Algorithmenentwicklung PHM) – beispielsweise zur Muster- und Anomalieerkennung sowie für Diagnosen und Prognosen. Die Kommunikation und Schnittstellen erfolgt zumeist über REST-Services, wobei eine Vielzahl von Adaptern für beliebige Schnittstellen- und Datenformate zur

Validierung, Vervollständigung und Harmonisierung von Daten eingebunden werden können. Das System unterstützt ein rollengestütztes Berechtigungskonzept zur Eigenverwaltung, das den Zugriff auf die einzelnen Programmfunktionen regelt. Die Rollen werden auch als Benutzergruppen bezeichnet und dienen zur Zusammenfassung von mehreren Usern gleicher Rechte. Es können beliebige Benutzer und Benutzergruppen frei definiert werden, wobei ein Benutzer beliebig vielen Gruppen zugeordnet werden kann.

Das System umfasst unterschiedliche Datenbanken, wobei jede Datenbank relevante Daten enthält und jeweils durch ein Set von Datenbank-Containern repräsentiert wird. Ein HAProxy (Englisch für "High Availability Proxy") ist darunter ein Open-Source-Software-Tool. Es wird eingesetzt, um den Datenverkehr auf mehrere Server zu verteilen (Load-Balancer) sowie die Skalierbarkeit, Verfügbarkeit und Zuverlässigkeit von Anwendungen zu verbessern. Der HAProxy ermöglicht den gesteuerten Zugriff auf unterschiedliche Datenbank- und Backup-Container und leitet bei Bedarf Datenverkehre auf andere verfügbare Server um. Zudem wird damit der Verbindungszustand einer Sitzung überwacht, um konsistente Sitzungen zu ermöglichen und Anwendungszustände einzuhalten. Jeder HAProxy kann über das Routing Network Anfragen zum Zugriff auf die Datenbank- oder etwaige Backup-Container erhalten. Er leitet die Anfragen dann an den jeweiligen Primary-Datenbank-Container weiter. Dadurch wird ein effizienter und verteilter Zugriff auf die Daten ermöglicht.

Hierzu baute die vedisys AG neben TCP-, und HTTP-Checks auch benutzerdefinierte Skripte zur Prüfung auf. Ein HAProxy unterstützt eine Vielzahl von Protokollen, darunter HTTP, HTTPS, TCP und UDP. Dadurch kann er in verschiedenen Anwendungsfällen eingesetzt werden, einschließlich Webanwendungen und API-Gateways, die im Vorhaben für eine moderne Anwendungsstruktur forciert werden. Zudem können damit umfangreiche Konfigurationsoptionen geboten werden, welche es Administratoren ermöglicht, das Verhalten und die Parameter nach spezifischen Anforderungen einer Applikation und Datenserviceanfrage anzupassen. Vorwiegend fallen darunter Aspekte wie Lastverteilungsalgorithmen, Protokoll- und Portweiterleitungen sowie Sitzungsbindungsregeln.

Zur automatisierten Konvertierung unterschiedlicher Datenformate ist ein internes (generisches) Format zur Ablage in einer Datenbank als Data Cube notwendig. Dieser Prozess wird mit ETL, aus dem Englischen für Extract, Transform und Loading bekannt, bezeichnet und ist in Abbildung 6 dargestellt.



Das aufgebaute Vorgehen beinhaltet dabei mehrere Schritte:

- Zuerst müssen die Daten aus den unterschiedlichen Datenquellen mittels geeigneter Adapter extrahiert werden (Extract).
- Danach werden die Daten in einer sogenannten Staging Area zwischengespeichert und vorverarbeitet, beispielsweise um Daten bei Bedarf bereinigen zu können (Transform).
- Zuletzt werden die aufbereiteten Daten in die generische Zieldatenbank eingeladen (Loading).

Im Rahmen des sogenannten Stagings werden die Daten validiert, mit Asset-Informationen angereichert für die weitere Verwendung mit en weiteren Daten und Informationen anderer Quellen harmonisiert (Mapping).

Abbildung 6: Schematische Anbindung

Als Schnittstelle eigneten sich generische REST APIs nach dem OpenAPI Standard. Diese API ermöglicht das kontrollierte Abfragen der benötigten Entitäten, bspw. die Abfrage benötigter Messereignissen (Events). Die empfangenen Nachrichten werden dann vom Datenbanksystem auf ihre Struktur hin validiert, um sicherzustellen, dass sie den vorher definierten Schemata für die weitere Datenverarbeitung entsprechen. Anschließend werden die Nachrichten aufbereitet und in die gewünschte Zielstruktur überführt. Dabei wurde bereits auf

Transformationsergebnissen zur Weiterverarbeitung und Analyse aus dem Staging-Prozess zurückgegriffen. Zudem wurden die Ebenen für das

- Datensystem zur Übersetzung und Optimierung von Anfragen an die Datenbank,
- Zugriffssystem zur Verwaltung physischer Sätze und Zugriffspfade auf die Datenbank und
- Speichersystem als Puffer und die Speicherverwaltung (auch extern) sowie
- die Datenbank zur physischen Speicherung von Daten

aufgebaut. Weiterhin wurde eine Metadatenverwaltung für weitere modellspezifische Daten wie Messreihen einzelner Fahrzeuge und eine Transaktionsverwaltung zur Synchronisation und Datensicherheit implementiert.

Insgesamt trägt eine generische Datenmanagementlösung dazu bei, die Komplexität der Datenanalyse zu reduzieren, die Effizienz zu steigern und die Genauigkeit der Ergebnisse zu verbessern. Eine entsprechende Struktur ist in Abbildung 7 dargestellt.

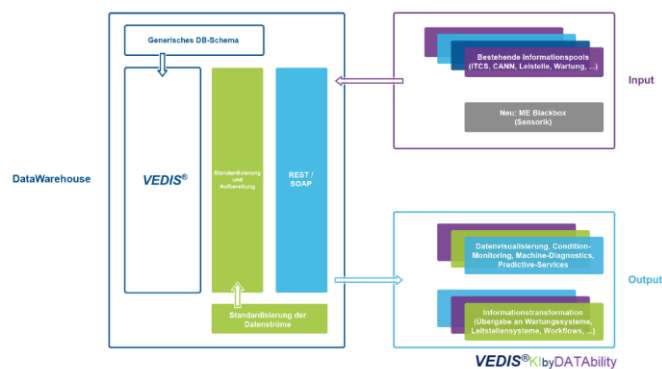


Abbildung 7: Schematische Interaktion der generischen Datenmanagementlösung mit Quellsystemen und Datenservices

Das damit aufgebaute generische Datenmanagementsystem der vedisys AG ermöglicht eine robuste Verwaltung von Datenbanken, Backups und Datenreplikation. Es gewährleistet eine hohe Verfügbarkeit, Datenkonsistenz und eine effiziente Nutzung der Ressourcen durch die Aufteilung von Schreibvorgängen und Backups auf die Primary-Datenbank. In Abbildung 8 ist die Zusammenführung der unterschiedlichen Rechnerknoten und Datenbanken und das Zusammenspiel aus Datenquellen und Anfragen dargestellt.

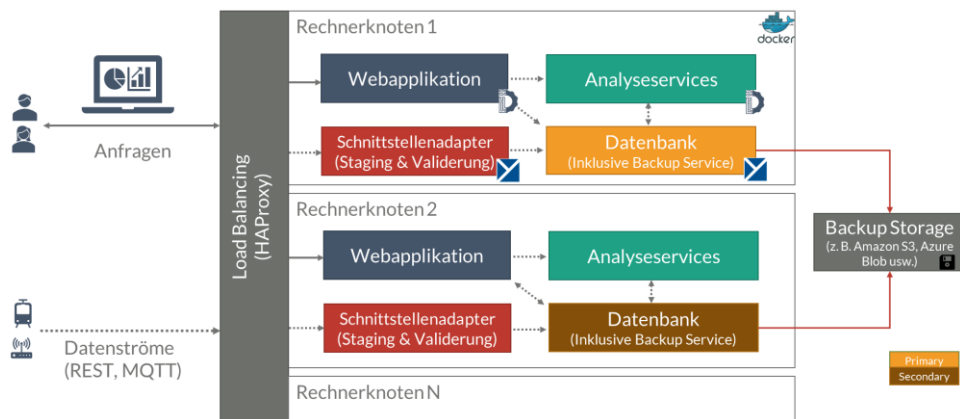


Abbildung 8: Schematischer Daten- und Informationsfluss des Datenmanagementsystems mit Rechnerknoten und Containern

AP5.3 Erweiterung von Analytik-Services

Im Rahmen der Tätigkeiten am Datenmanagementsystem wurden in diesem Arbeitspaket von der vedisys AG in enger Zusammenarbeit mit DATAbility Strukturen und Schnittstellen spezifiziert, um Datenbankfunktionalitäten

um Analytik-Services zu erweitern und zu integrieren. Zu Beginn fand eine gemeinsame Erörterung statt, welche Datenanalysen und -aggregationen datenbankseitig durchgeführt werden müssen.

Ein flexibler Schnittstellenadapter erlaubt das Verarbeiten verschiedener Kommunikationsprotokolle. Die Schnittstellen zu einer Datenbank für Analytik-Services bieten die notwendigen Werkzeuge und Funktionen, um Daten aus der Datenbank abzurufen, zu transformieren und zu analysieren. Zu wichtigen Schnittstellenkonzepten, die für die Integration von Analytik-Services mit einer Datenbank zum Einsatz kommen, wurden für das Vorhaben genauer analysiert:

- SQL: Structured Query Language zur Ermöglichung komplexer Abfragen, Filterung, Aggregation und andere Operationen, um Daten nach bestimmten Kriterien zu extrahieren und vorzubereiten
- API: Programmatische Schnittstellen, die häufig REST (Representational State Transfer) oder GraphQL Architekturen zur Datenabfragen und -manipulation nutzen
- ETL: Unterstützung des Datenflusses von Extraktion von Daten aus der Datenbank, über die Transformation in das gewünschte Format bis hin zur anschließenden Bereitstellung für die Analyse (Einladen)
- Treiber: Datenbankspezifische Softwarekomponenten, die unterschiedliche Programmiersprachen unterstützen (bspw. Java, Python, C#) und von Datenbankentwicklungsunternehmen bereitgestellt werden
- Konnektoren: Generischere Komponenten als Treiber (s. o.) mit spezifischen Funktionen und Optimierungen für die Analyse (bspw. Parallelverarbeitung oder Abfrageoptimierung)
- Streaming: Kontinuierlicher Datenflussbereitstellung (in Bezug auf Echtzeitfähigkeit) von der Datenbank zu Analytik-Services

Generische REST API nach OpenAPI Standards boten für das Vorhaben eine umfassende, einfache und standardisierte Schnittstellenkonnektivität zu den Datenbanken, um Daten und Ressourceninformationen über ein HTTP-Protokoll entgegenzunehmen und Messereignisse abzufragen. Diese harmonierte auch mit dem für die Messeinheiten (Geräte) der integrierten Sensorbauteile spezifizierten Kommunikationsprotokoll MQTT. MQTT dient der Kommunikation und Übertragung von Daten zwischen Geräten und Anwendungen in einem Netzwerk. Damit ist es möglich, Daten von IoT-Geräten zu erfassen und anschließend diese Daten in eine Datenbank zu schreiben. Die Integration kann entweder über spezielle MQTT-Broker erfolgen, die als Vermittler zwischen MQTT-Clients (Messeinheiten) und der Datenbank fungieren, oder über Middleware- oder Integrationskomponenten (Software), die MQTT-Nachrichten verarbeiten und die Daten in die Datenbank übertragen. MQTT basiert auf dem Publish-Subscribe-Modell, bei dem Nachrichten von sogenannten Publishern an sogenannte Subscribers übermittelt werden. Es bietet damit eine leichte und effiziente Übertragung von Daten und ist besonders für Umgebungen mit begrenzter Bandbreite und hoher Latenz geeignet.

In dieser gewählten und aufgebauten Konfiguration übernahm die REST-Schnittstelle in einem MQTT-basierten IoT-Netzwerk Aufgaben zur Nachrichtenverarbeitung, Datenvalidierung, Transformation und Persistenz von Daten in der Datenbank. Zu Beginn wurden empfangenen Nachrichten des MQTT-Brokerdienstes aufbereitet, um die relevanten Nachrichten und Informationen zu extrahieren und aufzubereiten (bspw. durch Entfernung nicht benötigter Informationen). Sodann erfolgt die Datenvalidierung, um auf erwartete Standards und Vorgaben zu testen sowie das Vorhandensein erforderlicher Felder und notwendiger Berechtigungen für die Zugriffssteuerung auf die Datenbank abzuprüfen. Danach erfolgt die Datenumwandlung und Vorverarbeitung für weitere Verarbeitungsschritte der REST API für die benötigte Zielstruktur zur Persistenz in der Datenbank. Die REST API übernimmt zudem die Verwaltung von Assets, indem empfangene Daten mit einer Asset-Struktur sowie mit Metadaten oder Konfigurationsinformationen verknüpft werden.

Das Datenbanksystem wurde mit dem Ziel zur Ermöglichung zukünftiger „As a Service“-Geschäftsmodellen über Lizenzen aufgebaut. Jedoch birgt dies Herausforderungen im Management von Datenübertragungen und -speicherung und damit besteht die Notwendigkeit einer starken Authentifizierung und Zugriffskontrolle. Hierzu entwickelte vedisys AG im Berichtszeitraum notwendige Module zur Authentifizierung und Autorisierung. Diese betreffen neben einer eigens aufgebauten und modifizierbaren Benutzer- und Rechteverwaltung auch etwaige

Kommunikationsschnittstellen für Anwendungen untereinander (bspw. Rechte auf Endpunkten, Schemata, Tabellen) sowie zwischen Anwendung und Usern (Menschen) für beispielsweise Zugriffe und zur Administration.

AP5.4 Test und Evaluierung des Datenbanksystems

Erste Tests hatten die grundsätzliche Überprüfung der Funktionalität der Gesamtprozesskette als auch die Ermittlung erster Systemleistungen in Form von Performanztests zur Ermittlung von Antwortzeiten (Lasttests) sowie Haltepunkte (Stresstests) zum Ziel.

Bei Lasttests der Datenbank wurde die Antwortzeit zur Ausführung unterschiedlicher Transaktionen zum Abrufen bestimmter Datensätze (Datenpakete) überprüft. Der Stresstest identifiziert den Haltepunkt eines Datenbanksystems, indem die Anwendung so belastet wird, dass das System zum Ausfall gebracht wird. Die Ermittlung einer Datenperformanz ist mit einem nicht zu unterschätzenden Aufwand verbunden. Um nachträglich Zeit- und Kostenprobleme zu vermeiden, ist eine sorgfältige Planung erforderlich.

Das Vorgehen und Ergebnisse werden im Folgenden anhand exemplarischer Beispiele genauer erläutert:

Für **Versuch Nr. 1** wurde ein exemplarischer Zug-Fahrplan über ein Jahr mit SOLL-Zeiten der Haltepunkte sowie den realen Halteereignissen mit IST-Zeit abgefragt werden. Die Größe der Tabelle belief sich auf:

- Zeilen Fahrplan: 372.876.293 Datensätze
- Zeilen Halteereignisse: 452.973.371 Datensätze

Für die Referenzabfrage kam bei einem **Versuch Nr. 1** der folgende Datenbankserver zum Einsatz:

- Hardware: Notebook, 8 GB RAM, Intel(R) Core(TM) i7-5600U CPU @ 2.60GHz, SSD-Festplatte 500 GB
- SQL-Server: 14.0.2037.2 (**MS-SQL**)

Folgende Metriken wurden dabei erzielt:

- **Abfrage 1:** Übersicht Fahrplanabweichungen eines Haltepunkts (Haltestelle) in einem Monat
 - Filter: Datumsfilter (ein Jahr), Haltepunktnummer, Liniennummern
 - Ergebnis: 57.2 Millionen Datensätze
 - *Abfrage-Dauer: 08:14 min*
- **Abfrage 2:** Übersicht aller Fahrplanabweichungen eines Haltepunkts in einem Jahr
 - Filter: Datumsfilter (ein Monat), Haltepunktnummer, Liniennummern
 - Ergebnis: 4.4 Millionen Datensätze
 - *Abfrage-Dauer: 01:27 min*

Durch die KI-gestützten Analysefunktionalitäten der DATAbility GmbH kommen weitere Algorithmen zum Einsatz, die für aufwendigere Abfragen und (Vor-)Verarbeitung von Daten und Informationen, bspw. zur Berechnung dedizierter Stati von Komponenten (wie bspw. der Gesundheitszustand eines Klimaaggregats) verwendet werden und Antwortzeiten für die Darstellung in Web-Services auf grafischen Benutzeroberflächen optimieren sollen.

Die Antwortzeiten einer prototypischen Abfrage mit einem ersten KI-Modul konnten wie folgt ermittelt werden: Basierend auf den vorherigen verwendeten exemplarischen Betriebsdatenauszügen wurde eine weitere Datenbank aufgebaut, indem die Betriebsdatenauszüge vervielfacht und Datumsangaben emuliert wurden. Für den **Versuch Nr. 2** mit einem KI-Modul der DATAbility GmbH kam die folgende Hardware und das folgende Datenbanksystem zum Einsatz:

- Hardware: 32 GB Ram, AMD Ryzen 7 3700X 8-Core 16 threads, Samsung SSD 970 Evo (NVME) 500 GB
- PostgreSQL-Server: pg12 (TimescaleDB, optimiert für Zeitreihendaten zur Messdatenauswertung)

Die abzufragende Tabelle belief sich auf künstlich erzeugte Daten auf Basis der Pünktlichkeit mit insgesamt 60.000.000 Zeilen, was in einem realen Anwendungsfall etwa einem Zeitraum von mehr als 8 Jahren bei einer Datenlieferung von 7 Mio. Datensätzen pro Jahr entspricht.

Für den Versuch Nr. 2 konnten folgende Metrik erzielt werden:

-
- **Abfrage 3:** Abrufen aller Ankunftspunktlichkeitsmessungen im gesamten Zeitraum (8 Jahre)
 - Filter: kein Datumsfilter, Haltepunktnummer, Liniennummern
 - Ergebnis: 530.000 Datensätze für einen auszuwertenden Haltepunkt
 - *Abfrage-Dauer:* 00:05 min

Bei diesen Versuchsauswertungen ist der Einfluss der unterschiedlich performanten Systeme zu beobachten. Zudem zeigen die Versuche unterschiedliche Technologien der Datenbanksysteme und Ansätze bei der Datenbankabfrage. Die Ergebnisse geben einen Überblick über zu erreichende Grenzen bei der Angabe von Performanzen (Abfragezeiten). Durch den Aufbau eines Cloud-Hostings mit Web-Services der Demonstratorlösung im Projekt wurde festgestellt, dass die Performanz durch geeignete Skalierung der IT-Ressourcen weiter gesteuert (optimiert) werden konnte. Auf der Basis dieser Ergebnisse wird es jedoch aus Performancegründen nicht möglich sein, eine direkte Speicherung von Messparametern der Fahrzeuge mit Übertragungsraten im Kilohertz-Bereich aus dem Fahrzeug heraus (online) in die Datenbank übertragen zu können. Hier bietet sich eine vorherige Vorverarbeitung der Daten an, um die zu verarbeitende Datenlast auf ein Minimum reduzieren zu können, ohne den Informationsgehalt der verarbeiteten Messparameter unbrauchbar zu machen. Diese Funktionalitäten sollen im weiteren Verlauf des Vorhabens umgesetzt werden.

Über einen sogenannten SOST, Englisch für System Operation Self Test, baute die vedisys AG einen Monitoring-Service auf, welcher automatisiert die grundsätzliche Erreichbarkeit des Application Servers und die Verbindung zur Datenbank testet und damit die Verfügbarkeit bewerten kann. Sollte eine dieser funktionalen Anforderungen nicht gegeben sein, löst dieser Service Notifikationen (Fehlermeldungen) aus, um auf Administratorebene als auch Nutzungsebene informiert zu werden. Zudem lassen sich damit (automatisiert) Maßnahmen einleiten, um die erneute Verfügbarkeit sicherstellen zu können. Die Testautomation wird beim Start des Systems sowie manuell bzw. durch abgelegte Routinen regelmäßig durchgeführt. Der Test überprüft verschiedene Teilsysteme und deren Funktionen, wozu neben Hardwarekomponenten wie beispielsweise Prozessoren, Speicher, Festplatten und Grafikkarten auch Softwarekomponenten wie Treiber (insbesondere Datenbanktreiber), Betriebssystemdienste und andere Systemressourcen zählen. Potenzielle Probleme sollen damit im Betrieb und nach Updates frühzeitig erkannt werden, bevor sie zu essentiellen Fehlfunktionen oder Ausfällen führen.

In Bezug auf die Sicherheit wurden neben der grundsätzlichen Überprüfung bestehender Sicherheitsmechanismen für den autorisierten Zugriff auf das Datenbanksystem die Wiederherstellungsfähigkeit für den Fall eines Gesamtausfalls oder eines Datenverlusts überprüft, um das entwickelte Backup- und Wiederherstellungskonzept zu testen. Diese Arbeiten stellen die Basis zur Weiterentwicklung des Entscheidungsunterstützungssystems dar. Es konnten eine Reihe von Optimierungspotentialen identifiziert werden, um im Rahmen der weiteren Entwicklung die Qualität des Systems zu verbessern.

AP5.5 Inbetriebnahme des Datenbanksystems

Die Inbetriebnahme des Datenbanksystems umfasste das Aufsetzen, die Einstellung und die Konfiguration der Serverarchitektur mit der Installation der Datenbanksoftware, von Komponenten und Treibern, Speicherzuweisungen, Festlegung von Sicherheitseinstellungen und Netzwerkkonfigurationen. Nachdem die Datenbank erstellt wurde, wurden die bisher zur Verfügung stehenden Datenquellen in die Datenbank importiert.

Timescale wurde als Zeitreihen optimierte Datenbankerweiterung verwendet, um eine leistungsfähige und effiziente Verarbeitung und Abfrage von Sensordaten sicherzustellen. Timescale ist dabei als Open-Source verfügbar und ist eine installierbare Erweiterung zu PostgreSQL. Damit können die Vorteile einer relational strukturierten Datenbank auch bei sehr großen Datenmengen genutzt werden. Sogenannte Hypertables ermöglichen einen schnellen Zugriff auf Datenchunks (Datenpartitionierungen über automatisch verwaltete Subtabellen), welche die Daten auf Basis eines definierten Zeitintervalls gruppiert und damit die Selektionseffizienz gewährleistet. Selbst bei einem Datenset von mehreren Milliarden Testdatensätzen kann Timescale effizient arbeiten und schnelle Abfragen ermöglichen. Zudem erweitern die integrierten Funktionen zur zeitlichen Aggregation und Datenanalyse den Funktionsumfang einer PostgreSQL deutlich, sodass bereits kondensierte Datenpakete zu den Abnehmerdiensten transferiert werden können. Durch die Möglichkeit zur effizienten

zentralisierten Durchführung von Standardberechnungen verbessert sich auch die Leistung (Reaktionsfähigkeit) der Endabnehmer.

AP5.6 Anforderungsanalyse für die Datenanalytik

Auf die Gesamtarchitektur des Datenbanksystems setzten die entwickelten Analytikservices der DATAbility GmbH auf. Zusammengefasst ist dies in Abbildung 9 schematisch dargestellt ist.

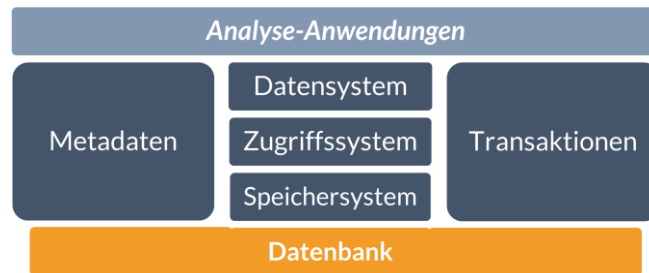


Abbildung 9: Schematische Gesamtarchitektur des Datenbanksystems

Neben der Zusammenführung und Speicherung unterschiedlicher Datensysteme stellt deren geeignete Aufbereitung und Verarbeitung zur Beantwortung spezifischer Analysefragen eine entscheidende Rolle für das Vorhaben. Für die Entwicklung eines PHM-Systems beschäftigte sich die DATAbility GmbH mit der Erörterung von Anforderungen an die Analysefunktionalitäten und Algorithmik für die in Arbeitspaket AP2 identifizierten Anwendungsfälle (AP5.6) sowie mit der Identifikation benötigter Daten und Informationen für eine Datendeskription (AP5.7). Die Tätigkeiten hatten zum Ziel, eine Basis für die Entwicklung von KI-Modellen und für das Entscheidungsunterstützungssystem zu schaffen, mit denen bspw. Beanspruchungen (Lastkollektive) und Gesundheitszustände von Bauteilen durch die integrierte Sensorik bewertet werden können.

Workshops zur Analyse der Datengrundlage und Zusammenstellung verfügbarer und zusätzlich benötigter Daten und Formate wurden zusammen mit den Konsortialpartnern durchgeführt. Die Arbeiten wurden in enger Kooperation mit der vedisys AG zum Aufbau des Datenbanksystems durchgeführt. In Bezug auf das dritte Element der zuvor vorgestellten Grundstruktur des Datenbankkonzeptes beschäftigte sich die DATAbility GmbH mit den folgenden Tätigkeiten zur Anforderungsermittlung des Output-Layers:

Durchführung von Datenanalysen und Reporting

- Ableitung von möglichen Funktionen und Anwendungsfällen, wie bspw.:
 - Automatische Ableitung und Bewertung von Lastkollektiven
 - Automatische Auswertung der Messdaten und Reporting (DBS) für
 - Früherkennung von Heißläufern
 - Erkennung von Lastpeaks
 - Früherkennung von Fehlern und Schädigungsverläufen
 - Bewertung der Dauerfestigkeit und Systemzuverlässigkeit
 - Bewertung der Laufstabilisierung (Flachstellen, Drehdämpfung)
 - Bewertung von Versagensrisiken bei Zugtüren
 - Bewertung der Infrastruktur
- Erhebung, Auswertung und Bereitstellung von Qualitätsdaten
 - Ableiten von Kennzahlen und Qualitätskriterien (z.B. objektive Pünktlichkeit)
 - Bonus-Malus Auswertung
 - Verspätungsverlaufsanalysen und Prognosen
 - Verspätungsursachenauswertung, Trends und Korrelationen
 - Erkennung von Betriebsabweichung (Anomalien)

- Entwicklung eines Dashboards mit geeigneten Aufbereitungen und Visualisierungen der Daten (Informationen zur Speicherbelegung, Inhalten und Datenraten für das Data Warehouse)

AP5.7 Datendeskription

Für eine erste Realisierung und für die Diskussion unterschiedlicher Monitoringkonzepte von Zugbauteilen und Systemen hat der Anwendungsgeber DB Systemtechnik einen exemplarischen Datensatz zur Verfügung gestellt. Es handelte sich dabei um Sensoraufzeichnung aus dem Daten- und Informationssystem eines Zuges, die im Rahmen einer Datendeskription hinsichtlich der Themen Informationsgehalte und Datenqualität analysiert wurden.

Die Daten stammen aus dem sogenannten Multifunction Vehicle Bus, kurz MVB. Der MVB beinhaltet dabei wichtige Daten und Informationen der Leittechnik sowie der Diagnosesysteme eines Zuges. Zu diesen zählen beispielsweise Antriebssysteme, Zugsicherungskomponenten, Führerstandanzeigen, zentrale Steuergeräte, dezentrale Ein-/Ausgabegeräte, Türsteuerungen, Klimaanlage, Hilfsbetriebsumrichter und Datenrekorder.

Der Datenextrakt bestand beispielsweise aus Zeitreihendaten eines 3-Achs-Beschleunigungssensors über einen Erfassungszeitraum von 12 Tagen und beinhaltet unterschiedliche Belastungsverläufe über unterschiedliche Fahrtrouten.

Die nachfolgende Abbildung im Zwischenbericht zeigt eine exemplarische Auswertung eines Anwendungsszenarios der DB Systemtechnik AG zur Erkennung und Bewertung von Last-Peaks aus den Beschleunigungsmessdaten an der Radachse.

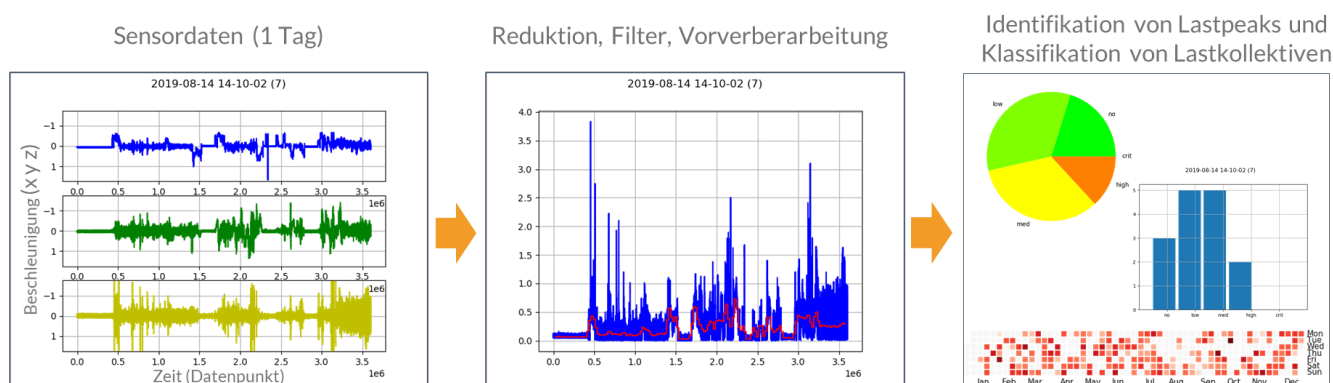


Abbildung 10: Verarbeitung von Beschleunigungsmessdaten zur Identifikation und Bewertung von Last-Peaks zu Lastkollektiven

Auf der Basis dieser Ergebnisse wurden mit dem Anwendungsgeber Anforderungen in Bezug auf die Darstellung der Bewertung für Nutzer*innen im Dashboard sowie Charakteristiken zur Ableitung von Lastkollektiven und Früherkennung von Fehlern (Schädigungsverläufen) diskutiert.

Im Datensatz sind weiterhin über 33.000 Datenpunkte aus 12 unterschiedlichen Fahrzeugen bzw. Wagons mit etwa 8.000 unterschiedlichen Parametern enthalten. Die Datenlieferungen wurden in sogenannten Events gebündelt. Abbildung 11 zeigt die aufgezeichnete Anzahl an Datenpunkten über den Betriebstag.

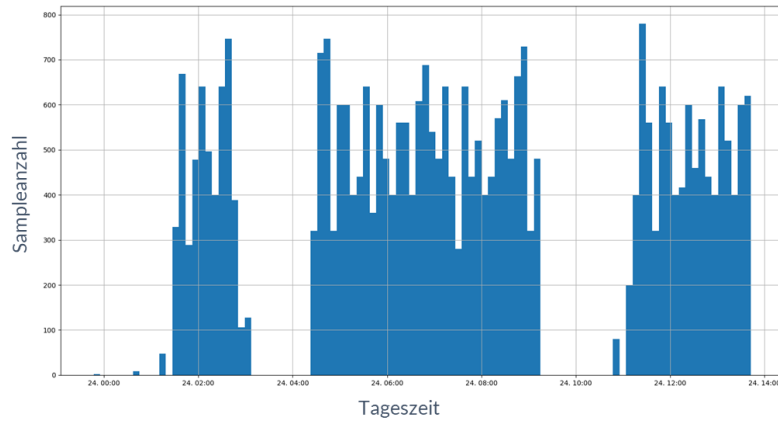


Abbildung 11: Anzahl der aufgezeichneten Events (Data Samples) über den Aufzeichnungstag

Jedes Event enthält dabei unterschiedliche Parameter, die beispielsweise über einen Zeitraum gemittelt wurden oder aktuelle Istwerte darstellen. Neben der Geolokalisierung des Zuges finden sich in den Events zusätzlich noch Informationen über die mittlere Außentemperatur, die Ist-Fahrgeschwindigkeit sowie mittlere Batteriespannung des Zuges und Fahrdrahtspannung. Insgesamt besteht der bereitgestellte Datensatz aus über 140 Events solcher Events. Mittels deskriptiver Analysen konnten Interdependenzen in den Datenauszügen hinsichtlich Plausibilität und Datenqualität bewertet werden.

AP5.8 Algorithmenentwicklung PHM

Die anhand der zuvor vorgestellten Datenauszüge bewerteten Charakteristiken zeigen, dass hinsichtlich der Subsysteme Klimaanlage und Türen mit MVB Auszügen genauere Analysen und der Aufbau von Services in Bezug auf die Ermittlung von Lastkollektiven und Schätzung zukünftiger Systemzustände vielversprechend erscheint.

Eine im Rahmen der Tätigkeit aufgebaute prototypische Applikation zeigt Abbildung 12 mit der Visualisierung bereitgestellter Daten (bspw. Messdaten vom Projektpartner DB Systemtechnik aus dem MVB-Bus). Diese ermöglicht eine Filterung für unterschiedliche Messparameter wie Beschleunigungs-, Kraft- sowie Temperaturdaten. Für die Anwendung konnten gemeinsam mit der vedisys AG ebenfalls Erweiterungen und Optimierungen direkt an der Schnittstelle zwischen Datenbanksystem und Web-Anwendung vollzogen werden. Die Arbeiten zielten auf die Reduktion der Abfragezeiten bei höheren Datenmengen ab, indem Datenabfrageprozesse sowie das Daten-Handling für den praxistauglichen Einsatz der Anwendung im Betrieb und der Instandhaltung von Bahntechnikkomponenten und -systemen optimiert wurden. Dadurch können lange Responsezeiten bei Nutzung vermieden werden.

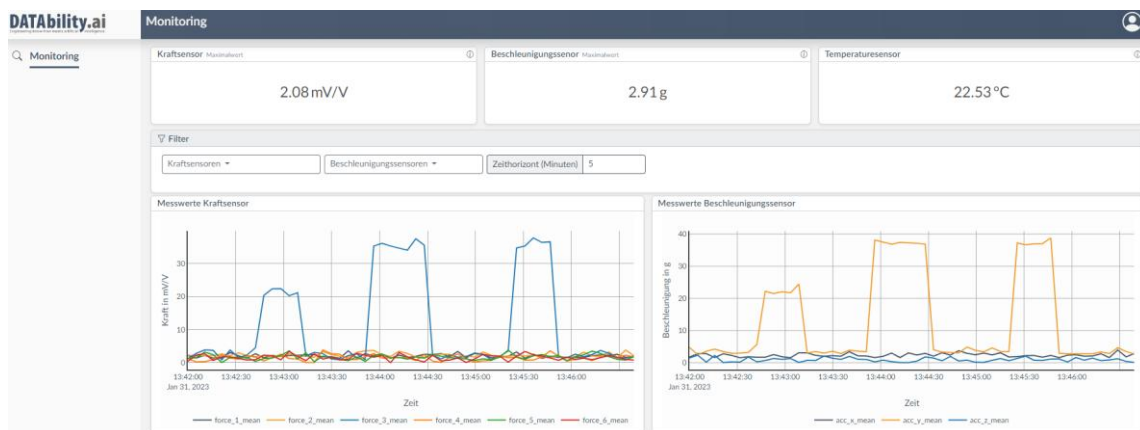


Abbildung 12: Prototypische Web-Anwendung (erster Demonstrator) für die Darstellung und Auswertung von Zugbetriebsdaten

Auf der Basis der identifizierten Anwendungsfälle wurde die Einbeziehung weiterer Datenquellen (bspw. neben Event- auch Zeitreihendaten) detaillierter untersucht. Hierzu wurde ein prototypischer Algorithmus aufgebaut und getestet. Dieser zielt auf die Überwachung eines Linearantriebes, bspw. beim Einsatz eines solchen in einem Türaufzug ab (siehe Liste identifizierter Anwendungsfälle und relevanter technischer Komponenten aus dem ersten Zwischenbericht). Dieser kann zur Fehlerdiagnose eines Linearantriebs auf Basis von Strommessungen genutzt werden. In Abbildung 13 ist der Analyseablauf und die Ergebnisse der Klassifikationsergebnisse visualisiert.

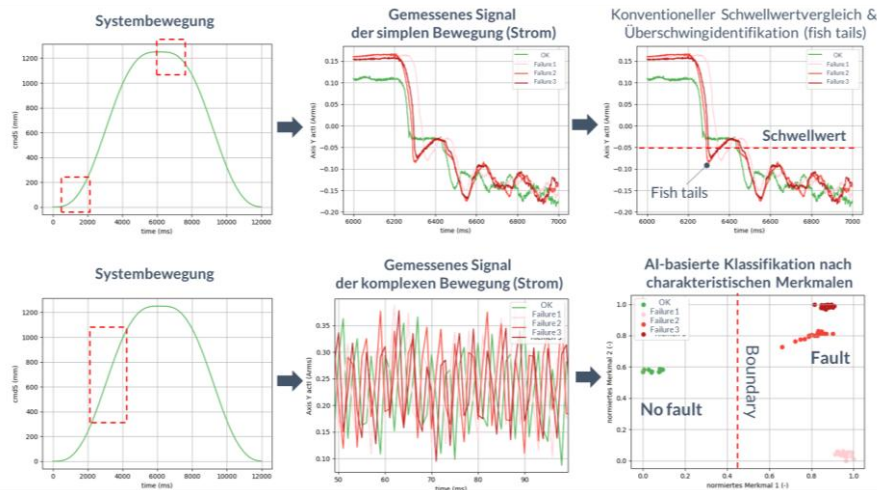


Abbildung 13: Prototypische Diagnose-Ergebnisse (Analyse-Demonstrator) für die Identifikation von Fehlerzuständen anhand des exemplarischen Anwendungsfall zur Überwachung eines Linearantriebes für den Türaufzug

Zur Entwicklung des Diagnosealgorithmus wurden vorliegende Strommessungen eines Linearantriebs herangezogen und für die Analyse vorbereitet. Der Algorithmus nutzt die Strommessungen eines Linearantriebs zum Training eines Klassifikators, um zwischen normalem und fehlerbehaftetem Betrieb anhand neuer Messdaten zu unterscheiden. Für das Training muss eine Aufteilung in diese Kategorien und ggfs. weiterer Fehlerarten vorgenommen werden. Verschiedene Algorithmen wie Support Vector Machines (SVM), Random Forest, Naive Bayes oder künstliche neuronale Netze wurden getestet, um die beste Leistung in Bezug auf Genauigkeit und Robustheit zu erzielen. Das Beispiel zeigt eindrücklich, dass für einzelne Bewegungsphasen des Linearantriebs, hier Anfahren, konventionelle Analysemethoden mittels Schwellwerten zur Erkennung von Überschwingen des Stromsignals ausreichen, um Systemfehler wie Verschleiß der Kraftübertragung zuverlässig identifizieren zu können.

Neben den technischen Bewertungen von Fahrzeugsystemen anhand von Realdaten wurde im Projekt SenseTrAIIn das Vehicle Routing Problem (VRP) als einer der zentralen Herausforderung in der Logistik und im Transportwesen, speziell für die multimodale Mobilität – insbesondere für die Individualbeförderung- untersucht. Es beschreibt die Optimierung von Routen für Fahrzeuge und Flotten für die Durchführung von Fahrten von Start zu Ziel mit minimalen Kosten unter Berücksichtigung verschiedener Restriktionen wie Kapazitäten, Zeitfenster oder Reisebedingungen. Die Entwicklung eines Services zur Erstellung optimaler Routenempfehlungen für das VRP erfordert eine durchdachte Architektur sowie effiziente Planungsalgorithmen. Hierzu zählten im vorliegenden Fall der Einsatz genetischer Algorithmen. Inspiriert von der biologischen Evolution werden Populationen von Lösungen erzeugt, die sich durch Mutation und Rekombination weiterentwickelt werden, bis unter den gegebenen Start- und Randbedingungen möglichst optimale Routenempfehlungen gefunden werden. Sie können für Probleme eingesetzt werden, für die es keine exakten Lösungsverfahren gibt.

Das VRP konzentriert sich primär auf die Optimierung von Routen für eine homogene Fahrzeugflotte. Beim multimodalen Routing hingegen müssen verschiedene Verkehrsmittel und deren spezifische Eigenschaften berücksichtigt werden. Hierzu zählen beispielsweise weitere Informationen wie Kapazitäten sowie Fahrzeugmerkmale wie Möglichkeit zum Transport von bestimmten Gütern (z. B. Fahrräder, Rollstühle). Insbesondere ist die Kostenfunktion um zusätzliche Faktoren wie Umsteigezeiten, Fahrpreise, Komfort und

Umweltbelastung zu erweitern. Reichert man diese zusätzlichen Randbedingungen dem Algorithmus an, entstehen multimodale Routenempfehlungen wie in Abbildung 14 dargestellt. Die Visualisierung erfolgte hierbei mit dem Open Source Software Open TripPlanner¹.

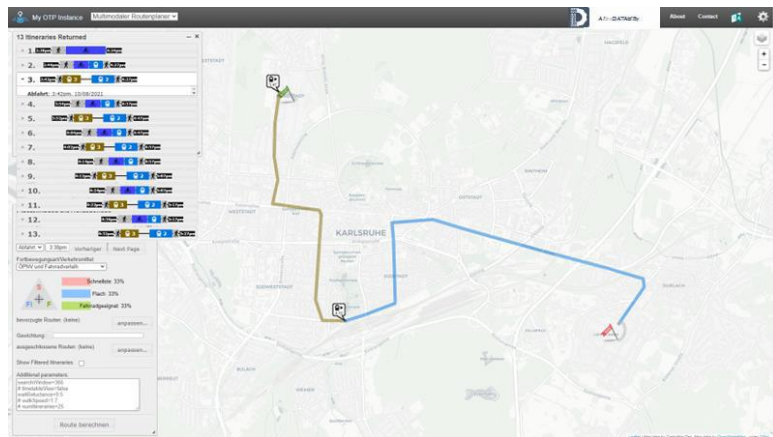


Abbildung 14: Beispiel für ein multimodales Routing durch Karlsruhe

Im Rahmen des entwickelten und verprobten Services erfolgte stets ein Mapping zwischen den relevanten Fahrplandaten einer Anfrage sowie einem Haltestellenverzeichnis zur Analyse und Berechnung der Routenempfehlungen. Zusätzlich erfolgt ein Mapping des betroffenen Haltestellenverzeichnisses mit Tarifpunkten. Daraus werden Mapping-Tabellen erstellt, die für zur Berechnung von Routenanfragen abgerufen werden können. Dies ist als Zwischenergebnis der Routenberechnung als Beispiel in dargestellt.

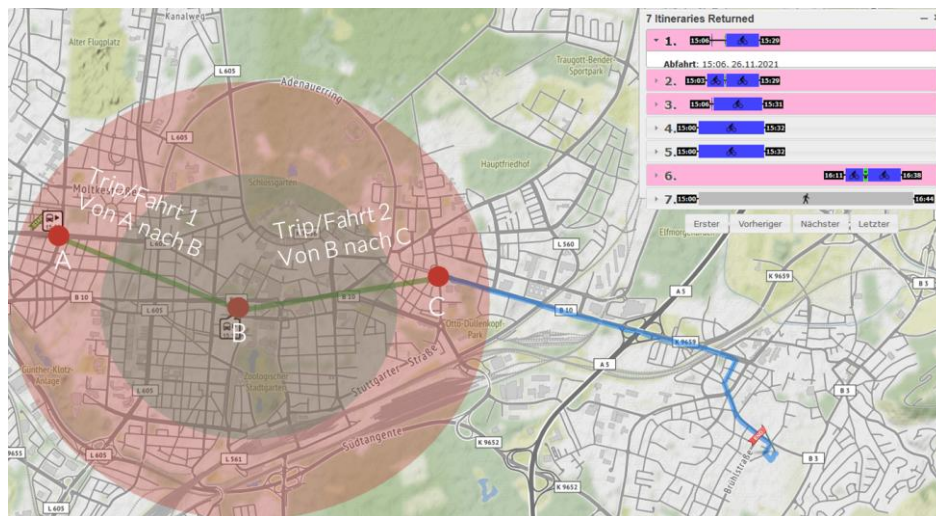


Abbildung 15: Beispiel zur Tarifberechnung zwischen Start und Ziel einer Route (hier A→C)

Das multimodale Routing selbst berechnet darauf basierend mögliche Graphen. Diese werden entlang ihrer Kanten durch Ihre Rahmenbedingungen wie bspw. Preis, Zeit oder Entfernung gewichtet. Nach Bewertung und Priorisierung der geeignetsten Routen anhand der Vorgaben wird sodann eine oder mehrere Routenempfehlung ausgegeben.

¹ Siehe <https://www.opentripplanner.org/>

AP5.9 Verifikation und Validierung der Datenanalytik

Für die Nachrichtenverarbeitung von Messparametern zukünftiger integrierter Sensor-Bauteile wurde für das Zusammenspiel aus Analytik-Services und Datenbank in ersten Tests validiert. Dahingehend wurde ein umfassender Penetrationstest, um die Leistungsfähigkeit des MQTT-Kommunikationsprotokolls zu validieren, durchgeführt. Zudem wurden Test hinsichtlich des Abfangens von Lastpeaks bei er Analyse untersucht. Die Ergebnisse werden im Folgenden dargelegt.

Der Penetrationstest umfasste sowohl die Verwendung von simulierten als einzelnen realen IoT-Geräten, um eine realistische Umgebung zu schaffen. Der Untersuchungsfokus lag auf der Untersuchung von Latenzzeiten und des Nachrichtenverlustes. Um die Belastbarkeit des Systems zu überprüfen, wurden mehrere Tausend simulierte IoT-Geräte (ca. 5.000 Stück) in den Penetrationstest einbezogen. Dies ermöglichte es, die Skalierbarkeit der Nachrichtenverarbeitung zu bewerten und mögliche Schwachstellen in Bezug auf die Anzahl der verbundenen Geräte aufzudecken. Der Test simulierte eine Vielzahl von Szenarien, um eine breite Nutzungspalette zu berücksichtigen.

Ein wichtiger Aspekt der Penetrationstestdurchführung war die Messung der Latenzzeiten zwischen dem Abschicken eines Messwerts von einem IoT-Gerät und dem Schreibvorgang in der Datenbank. Die Ergebnisse zeigten eine durchschnittliche Latenzzeit von 3 Sekunden, wobei der längste gemessene Wert bei 15 Sekunden lag. Dies deutet darauf hin, dass die Nachrichtenverarbeitung insgesamt effizient war und die gewünschte Leistung erfüllte. Dennoch sollte bei Latenzzeiten von mehr als 10 Sekunden eine weitere Optimierung in Betracht gezogen werden, um die reibungslose Funktion des Systems sicherzustellen. Ein weiterer wichtiger Faktor bei der Durchführung des Penetrationstests war die Messung des Nachrichtenverlustes während der Kommunikation über das MQTT-Kommunikationsprotokoll. Das Ergebnis zeigte einen vernachlässigbaren Nachrichtenverlust von weniger als 0,01%. Dies ist ein ermutigendes Ergebnis, das darauf hinweist, dass die Nachrichtenintegrität während des Kommunikationsprozesses gewahrt bleibt.

Der durchgeführte Penetrationstest zur Validierung der Nachrichtenverarbeitung mittels MQTT-Kommunikationsprotokoll zeigte insgesamt positive Ergebnisse. Basierend auf diesen Ergebnissen kann davon ausgegangen werden, dass das MQTT-Kommunikationsprotokoll für die beabsichtigten Anwendungen geeignet ist, jedoch sollten die Latenzzeiten überwacht und gegebenenfalls optimiert werden, um eine optimale Leistung zu gewährleisten.

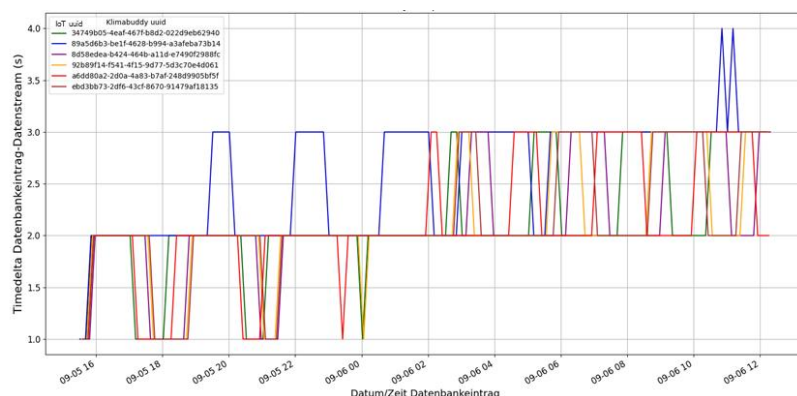


Abbildung 16: Exemplarische Ergebnisdarstellung zur Messung der Antwortzeiten in Abhängigkeit der Geräteanzahl

Um sicherzustellen, dass die Analyseseite bei Lastspitzen effizient und zuverlässig in Bezug auf Analytik-Services funktioniert, wurde eine Überwachung der Serverarchitektur implementiert. Das Abfangen von Lastpeaks kann seitens der Serverarchitektur erfolgen, um eine kontinuierliche Überwachung und Anpassung von Ressourcen zu gewährleisten sowie eine zuverlässige und reaktionsschnelle Datenbank und IoT-Datenanalyse zu ermöglichen. Das Monitoring-Tool Prometheus wurde eingesetzt, um wesentliche Performancekennwerte wie CPU- und RAM-Auslastung, freien Festplattenspeicher sowie eingehende und ausgehende Datenpakete zu erfassen. In Abbildung 17 ist eine solche Auswertung dargestellt. Diese Metriken ermöglichen es, potenzielle Engpässe oder Überlastungen frühzeitig zu erkennen und entsprechende Maßnahmen einzuleiten. Darüber hinaus wurden auch

eingehende und ausgehende Datenpakete erfasst, um den Netzwerkverkehr und die Kommunikation mit den IoT-Geräten zu analysieren. Im Folgenden werden die Hauptaspekte des Monitorings und die angewandten Maßnahmen zur Lastbewältigung beschrieben.

Bei erhöhter Auslastung besteht die Option, zusätzliche Rechnerknoten hinzuschalten, um die Ressourcenkapazität zu erweitern. Dadurch konnten die erforderlichen Ressourcen an die aktuellen Anforderungen angepasst werden. Zusätzlich wurde ein Load Balancing mit Hilfe von HAProxy eingeführt, um die Last dynamisch auf die verfügbaren Knoten zu verteilen. Dies gewährleistete eine effiziente Nutzung der vorhandenen Ressourcen und verbesserte die Gesamtpformance der Analyseseite. Ein Alarmmanagement-System wurde zudem implementiert, um bei Überschreiten definierter Grenzwerte automatisch Warnmeldungen zu generieren. Beispielsweise wurde eine hohe CPU-Auslastung als kritischer Faktor betrachtet. Bei Erreichen eines bestimmten Schwellenwerts wurde ein Alarm ausgelöst, der das Service-Operation-Team benachrichtigt, um geeignete Maßnahmen zur Lastverteilung manuelle oder automatisiert ergreifen zu können. Dadurch konnten potenzielle Leistungsengpässe proaktiv identifiziert und behoben werden, um Ausfälle oder Beeinträchtigungen der Analyseseite zu vermeiden.

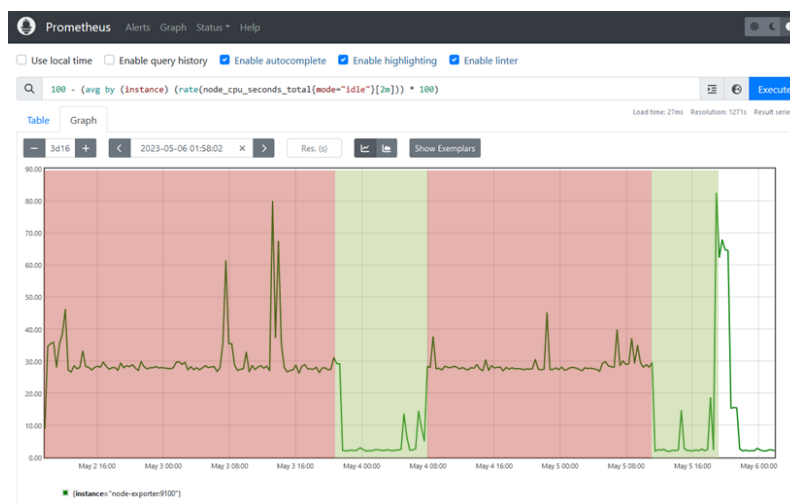


Abbildung 17: Exemplarische CPU-Auslastung eines Knotens während einer Datenanalyse - Steuerung der Auslastung durch Hinzunahme weitere Rechnerknoten

Die Implementierung des Monitorings-Tools sowie der genannten Lastbewältigungsstrategien führte zu einer verbesserten Performance und Robustheit der Analyseseite.

AP6 Demonstrator

Die multimodale Routenplanung ist ein wichtiger Baustein für eine zukunftsfähige und nachhaltige Mobilität, da sie die Berücksichtigung verschiedener Transportmoden und Verkehrsmittel wie Individualverkehr (Auto, Sharing), öffentliche Mobilitätsangebote sowie persönliche Fortbewegungsmittel (Fußstrecken, Fahrrad) bei der Planung einer Reise berücksichtigt. Dies kann zu effizienteren und nachhaltigeren Reisemöglichkeiten führen. Die Nutzung verschiedener Mobilitätsmoden ermöglicht es, auf individuelle Bedürfnisse und sich ändernde Umwelt- und Rahmenbedingungen einzugehen. So kann je nach Tageszeit, Wetter oder Gepäckmenge das passende Transportmittel gewählt werden. Durch die Kombination verschiedener Verkehrsmittel können oft schnellere und günstigere Verbindungen gefunden werden, als mit einem einzigen Verkehrsanbieter. Multimodale Routen tragen zur Reduzierung des Verkehrsaufkommens, insbesondere des Individualverkehrs, besonders in urbanen Gebieten sowie als Erschließungsmöglichkeit ländlicher Regionen bei. Dies wiederum führt zu weniger Emissionen, weniger Lärm und einer besseren Luftqualität. Zudem lassen sich Städte und Regionen damit attraktiver gestalten.

Die Tätigkeiten in Use Case 6 beschäftigten sich mit dem Aufbau eines Mobility-as-a-Service zur Bündelung verschiedener Verkehrsangebote (bspw. ÖPNV, Bike- & Carsharing, etc.) auf Datenbankseite, zur Berechnung und

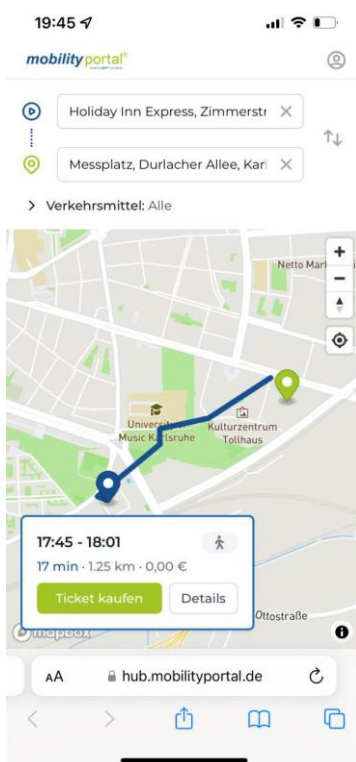
Analyse von Routenempfehlungen auf Analytik-Seite sowie mit der Konzeption zur einfachen Buchung und Bezahlung solcher multimodalen Reiseangebote. Für die vedisys AG standen vor allem der Aufbau des notwendigen Daten-Handlings sowie der Verfügbarmachung des Services auf einer zentralen Mobilitätsplattform, die auf dem Datenbanksystem basiert, im Vordergrund.

Zur Erweiterung der Datenbank wurden Echtzeit-Verkehrsdaten berücksichtigt, die bspw. über Sensoren in Fahrzeugtelematiksystemen erhoben werden, um die aktuellen Positionen der Verkehrsmittel (Auto, Bus, Zug, Bikes, Roller etc.) einbeziehen zu können. Weiterhin sind Fahrplandaten von den jeweiligen Verkehrsbetrieben notwendig, um Planungsstände einbeziehen zu können. Alle beschriebenen Stammdaten werden von der vedisys AG zusätzlich im entwickelten Datenmanagementsystem verwaltet. Ausgenommen hiervon sind die ausschließlich für das Routing relevanten Stammdaten, wie Fahrplan und Haltestellen – diese werden durch den Routenermittlungsservice von DATAbility GmbH direkt in den Routing-Service bezogen bzw. abgefragt und gegebenenfalls für das Routing aufbereitet. Die SOLL-Fahrplandaten und das Haltestellenverzeichnis können bspw. über die öffentliche Schnittstelle DELFI² bezogen werden.

Für die Analyse und Tests zum Anwendungsfall wurde im Service eine multimodale Reiseanfrage vorgesehen. Diese bestand aus Start- und Zielangabe, Datum und Uhrzeit sowie nach Bedarf den gewünschten Reisepräferenzen (zu Fuß, ÖPNV, PKW, Car Sharing, Bike/Scooter etc.). Das Ergebnis der Reiseanfrage aus dem Service musste für User mindestens folgende Informationen enthalten:

- Routeninformation mit allen Teilabschnitten unter Berücksichtigung der gewünschten Modalitäten
- Datum/Uhrzeit jedes Ein-/Aus-/Umstiegs
- Gesamtlänge (zurückzulegende Wegstrecke in Kilometer) und gesamte Reisedauer
- Optional: Ermittlung der Gesamtkosten unter Berücksichtigung unterschiedlicher Tarifzonen
- Gegebenenfalls Alternativroutenvorschläge (absteigend sortiert nach Relevanz gemäß Präferenzen)

Für den Aufbau des Demonstrators wurden definierte Demotarife verwendet. Dabei wurden kleinflächige Tarifgebiete definiert, die zur Komplexitätsreduktion vorerst keine hierarchischen Strukturen aufwiesen.



Nach Konzepterstellung, Spezifikation und Implementierung wurde eine Testumgebung als Web-Service aufgebaut. Die dafür entwickelte Benutzeroberfläche für den Usertest des Services ist in **Fehler! Verweisquelle konnte nicht gefunden werden.** dargestellt. Diese bietet alle beschriebenen Services über den Application Server an.

Der Service „travelrequest“ auf dem Application Server nimmt die Reiseanfrage vom Client entgegen, führt optional eine Anreicherung der Daten durch und leitet diese Anfrage schließlich an den Routing Service weiter (siehe weitere Erläuterung im nächsten Kapitel). Als Response wird das vom Routing Service gelieferte und wiederum angereicherte Ergebnis an den ursprünglichen Client zurückgeliefert. Die Anreicherung erfolgt hierbei insbesondere um Abrechnungsdaten. Falls der Anwender bei der Erstellung seiner Reiseanfrage Präferenzen auswählt, können diese direkt im Serviceaufruf mitgegeben werden. Sollte der Anwender registriert sein, kann auch dessen Account-ID übergeben werden. In diesem Fall werden gegebenenfalls im Account Profile hinterlegte Präferenzen automatisch ermittelt. Eventuell manuell vom Anwender ausgewählte Präferenzen haben dabei immer Vorrang vor den im Account Profile gespeicherten Angaben. Zu den Präferenzen gehört auch die gewünschte Sortierung des Ergebnisses, wie Zeitdauer, Streckenlänge oder Preis. Die eigentliche Sortierung für die Anzeige liegt jedoch in der Verantwortung des Clients, daher führt der Service selbst keine Sortierung der Daten durch.

² <https://www.delfi.de/de/leistungen-produkte/daten-dienste/>

Unterschiedliche Verkehrsmittel, Ridepooling-Angebote und Tarifoptionen wurden zusammengeführt und individualisierte Reiseketten unter Berücksichtigung personalisierter Zielvorgaben wie Reisezeit, Nachhaltigkeit oder Kosten vorschlagen.

AP7 Feldtest & Nutzenbewertung

Für den Feldtest wurde die Einbeziehung und Auswertung von Zusatzinformationen in den Analytik-Services vorgesehen, die im Bereich der Instandhaltung und im Betrieb von Schienenfahrzeugen anfallen können. Es handelt sich dabei um Texte, die bspw. verschriftlich in Wartungsprotokollen abgelegt werden oder im Beschwerdemanagement von Kunden (Reisegästen oder auch Fahrzeugbetreiber (bspw. Techniker, Fahrpersonale) rückgemeldet werden (bspw. telefonische Transkription oder Feedback-Formular auf der Webseite des EVUs bzw. des Verbundes). Diese Informationen können einerseits für das Labeling von Datensätzen unterstützen, um beispielsweise Zeiträume von Fehler-Events besser eingrenzen zu können und andererseits zusätzliche Merkmale bieten, sodass Fehlerfälle (beispielsweise in Ort, Art und Ausprägung) detaillierter bzw. differenzierter analysiert werden können.

Als Basis für den Aufbau und Tests der Analysefunktionalität wurde auf exemplarische Rückmeldung aus dem Beschwerdemanagement eines EVUs durch anonymisierte Daten der vedisys AG zurückgegriffen. Diese wurden ebenfalls als Datenquelle (bisher noch ohne Zusammenhang zu den analysierten MVB-Daten) über das Datenbanksystem der vedisys AG bereitgestellt. Es wurde ein Analyseservice aufgebaut, welcher auf Basis der Meldungstexte als Lerndatenbasis neue Freitexteingaben in dedizierte Kategorien klassifizieren kann. Diese beziehen sich beispielweise auf die Kategorien Pünktlichkeit oder Fundsachen.

Dahingehend erfolgte der Aufbau einer geeigneten Datenprozesskette für unterschiedliche Vorverarbeitungsschritte für die Textverarbeitung – beispielsweise Lemmatisierung sowie das Entfernen sogenannter „stop words“ in den Meldungstexten. Darauf aufbauend konnten verschiedene NLP-Klassifikationsmodelle für die Kategorisierung neuer Nachrichten (Freitexteingaben) aufgebaut, modifiziert und validiert werden. Ziel der Validierung war insbesondere die Durchführung einer Sensitivitätsanalyse auf Basis der Meldungstexten zu klassifizierten Kategorien. Die Ergebnisse daraus wurden zur Optimierung der Modelle durch ein sogenanntes Hyperparametertuning verwendet.

In Abbildung 18 ist das Front-End des entwickelten Services für die automatische Kategorisierung von Nachrichten dargestellt. Auf der Basis der für das Training verwendeten Daten werden die durch Nutzereingabe neu eingegebenen Texte (in das Eingabefeld) der Analyse durch das NLP-Klassifikationsmodell (im Hintergrund) zugeführt. Das Ergebnis der Klassifikation ist unmittelbar (live) unterhalb des Eingabefeldes als sich veränderter Prozentwert der vom Modell zugeschriebenen Kategorie dargestellt.

Im Rahmen der weiteren Tätigkeiten der DATAbility GmbH wurden die notwendigen Schnittstellen (auf Basis OpenAPI) für die Bereitstellung des Services und der Kommunikation mit dem Datenbanksystem zusammen mit der vedisys AG definiert und die Anwendung als Cloud-Service als Container mit Hilfe der Amazon Web Service Umgebung verfügbar gemacht. Im weiteren Verlauf des Vorhabens im Folgejahr sollen dann zusammenhängende Datensätze zwischen MVB-Bus sowie Meldungstexte aus der Instandhaltung oder dem Beschwerdemanagement für eine kombinierte Fehleranalyse genutzt werden und die Funktionalität für das Prognostics & Health Management weiter ausgebaut werden.

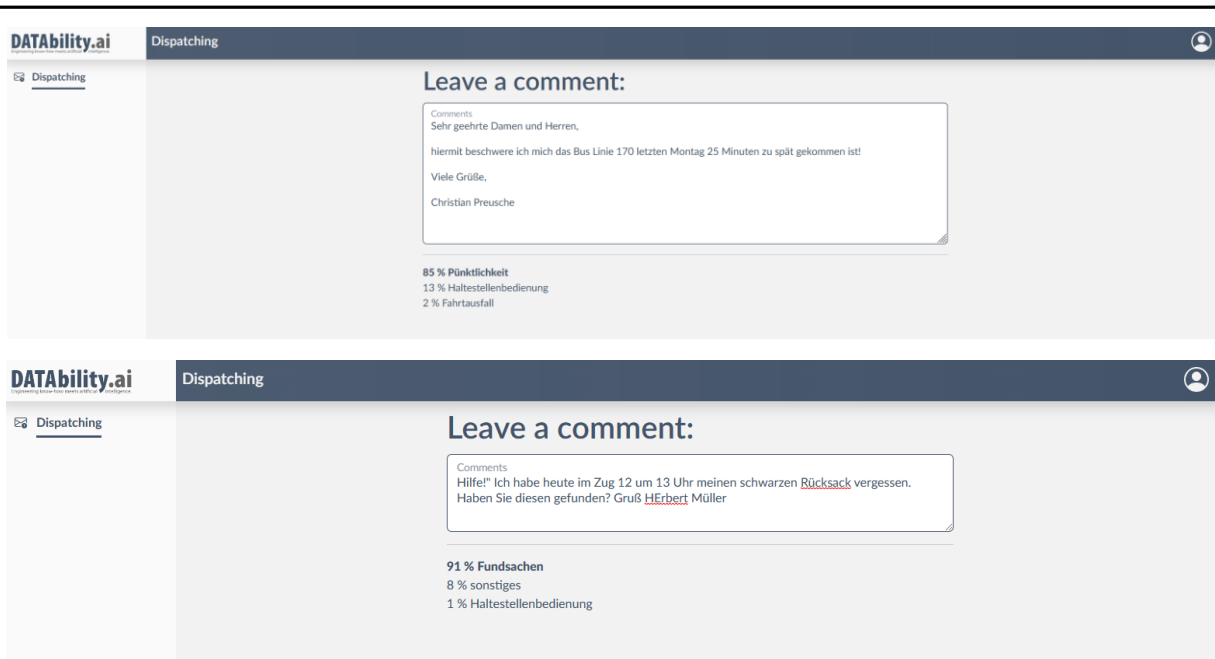


Abbildung 18: Prototypische Eingabefläche (Front-End) der entwickelten Anwendung zur Klassifikation von Meldungstexten für die Instandhaltung und das Beschwerdemanagement (Kunden-Feedback)

Für einen weiteren Feldtest des aufgebauten Datenbank- und Analysesystems wurde ein zu dieser Zeit aktueller Bedarf im Mobilitätsbereich aufgegriffen: Das Deutschlandticket, auch bekannt als 49-Euro-Ticket, bot bundesweite Fahrten im öffentlichen Personennahverkehr (ÖPNV). Mit der Einführung des Deutschlandtickets waren jedoch Betrugsfälle aufgetreten. Insbesondere wurde in einigen Fällen berichtet, dass Nutzer versuchten, Buchungsdaten zu manipulieren. Damit wurde versucht, das Ticket beispielsweise zu einem niedrigeren Preis zu erhalten oder um es mehrfach zu bestellen und dann weiterzuverkaufen. Eine weitere Variante ist die Manipulation der Zahlungsinformationen. Durch diesen Missbrauch in einer nicht zu geringen Anzahl von Fällen wurden Zahlungsinformationen angegeben, die nicht gedeckt sind oder es wurde die Zahlungsabwicklung derart manipuliert, dass Tickets ausgestellt wurden, ohne tatsächlich das angefallene Lastschriftmandat zu bedienen. In solchen Fällen ist es möglich, dass das Ticket zunächst als gültig ausgegeben wird, die Zahlung aber später noch fehlschlägt. Als Abstellmaßnahmen wurden bereits strengere Zahlungsprüfungen durchgeführt beziehungsweise auf andere Bezahlmethoden umgestellt, um sicherzustellen, dass nur vollständig bezahlte Tickets ausgegeben werden. Durch den Einsatz fortschrittlicher Technologien zur Betrugserkennung können verdächtige Bestellungen schneller erkannt und gestoppt werden. Für das Projekt wurden Bestellmustern analysiert, um auffällige Aktivitäten zu identifizieren.

Für die datenseitige Verarbeitung solcher transaktionellen Datensätze wurde das Datenmanagementsystem entsprechend erweitert und ein weiterer Datenanalytik-Service im Umfeld des Application Servers aufgebaut. Um eine Transaktion auf Betrug zu überprüfen, wurde ein sogenannter Validation Request aus dem Datenmanagementsystem an den Analytikservice gestellt. Die Nachricht dafür enthielt alle dafür benötigten Informationen. Sie wird an die Anwendung weitergeleitet und löste dort automatisch eine Validierung der gegebenen Buchung aus. Nachdem eine Anfrage für eine Betrugsüberprüfung eingegangen und die dazugehörige Überprüfung abgeschlossen wurde, wurde eine Antwort vom Service an das Datenmanagementsystem zurückgesendet. Diese Antwort, die sogenannte Validation Response, übersendet das Ergebnis der Überprüfung zusammen mit einer Begründung am Ende der Nachricht. Dabei erfolgte die Identifizierung der Betrugsversuche so, dass die Anonymität der Bestellenden gewahrt wurde und keinerlei Bezug zu Name, Wohnort, Alter, Herkunft oder Geschlecht hergestellt werden konnte.

Zunächst wurde eine Methodik entwickelt, die darauf abzielt, verdächtige Muster und Anomalien bei Ticketbuchungen zu identifizieren. Diese Daten beinhalteten legitime als auch betrügerische Transaktionen und wurden durch Mitarbeit betroffener Unternehmen entsprechend markiert. Danach wurden Schlüsselfaktoren

identifiziert, die auf Betrug hindeuten könnten, wie z. B. ungewöhnliche Bestellmengen, auffällige Zahlungsmethoden oder abweichende bzw. irreführende Angaben. Diese Merkmale wurden in Form von Features in das Modell integriert. Basierend auf der Datenanalyse wurde ein geeignetes Modell zur Betrugserkennung ausgewählt und in Zusammenarbeit mit Regel-basierten Modellen für die Prozessabwicklung aufgebaut.

Nachdem die Methodik entwickelt wurde, erfolgte die praktische Umsetzung. Dies umfasste die Implementierung des Modells in einen Micro-Service, um diesen auf dem Application Server verfügbar zu machen. Abschließend wurde das Modell und der Service unter verschiedenen Szenarien getestet, beispielsweise bei stark erhöhtem Buchungsvolumen oder bei gezielten Betrugsversuchen, um seine Robustheit zu prüfen. Neben den technischen Metriken wurde auch die wirtschaftliche Auswirkung der verschiedenen Methoden betrachtet, z. B. die Kosten durch Fehlalarme (False Positives) und damit zurückgehaltene Bestellungen (ggfs. dann durch Buchungsausfall) oder nicht erkannte Betrugsfälle (False Negatives). Durch dieses strukturierte Vorgehen konnte ein robustes System zur Erkennung von Betrugsversuchen bei Ticketbuchungen entwickelt und kontinuierlich verbessert werden. Der von DATAbility entwickelte Algorithmus identifizierte auffällige Bestellungen in Echtzeit, um Betrugsversuche bei Bestellungen direkt nach Eingang unterbinden und vor einem vollständigen Prozessdurchlauf stoppen bzw. pausieren zu können. Es konnten Tests und Verifikationen des Services mit realen Anwendungsunternehmen von der vedisys AG vorgenommen werden.

Die Funktionsfähigkeit der aufgebauten Analysefunktionalitäten, darunter für das Prognostics and Health Management (PHM), zum multimodalen Routing sowie für das Beschwerde- und Bestellmanagement sowie die Transformation der erzeugten Ergebnisse in konkrete Entscheidungsregeln wurden validiert. Durch die enge Zusammenarbeit von DATAbility GmbH und vedisys AG konnte sichergestellt werden, dass diese komplexen Analysedaten in verständliche und handlungsrelevante Regeln übersetzt werden. Die Validierung erfolgte durch umfassende Tests, unter anderem an Demonstratoren im realitätsnahen Bereich der DB Systemtechnik GmbH sowie weiterer Akteure in diesem Umfeld. Damit wurde der Projektmeilenstein MS 5 für das Teilvorhaben erzielt.

II.2 Erläuterung der wichtigsten Positionen des zahlenmäßigen Nachweises

Im Projektverlauf sind wesentlich Kosten für Personal sowie F&E-Fremdleistungen entstanden. Dabei wurden die vom Projektträger genehmigten vorkalkulierten Kosten in keiner Kostenart überschritten. Die Personalkosten sind zur Bearbeitung der im Projektplan beschriebenen Arbeiten aufgelaufen. Die eingebundenen Unteraufträge für die Frankfurt Consulting Engineers war notwendig, um das Gesamtziel zu erreichen und notwendiges externes Know-how in das Projekt zu integrieren.

II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die geleisteten Arbeiten waren notwendig zur Erreichung der dargestellten Ergebnisse. Das im Rahmen des Projektes alternative Use Cases im vorliegenden Teilvorhaben bearbeitet werden mussten, war den enormen Aufwänden bei der Erzeugung realer Messdaten aus Testfahrten geschuldet.

II.4 Voraussichtlicher Nutzen

Die vedisys AG wird das entwickelte Datenbanksystem zukünftig für den Ausbau und Einsatz von Micro-Service Architekturen für das bestehende Produktportfolio weiterentwickeln und einsetzen. Vorgesehen sind dabei die beschriebenen Einsatzszenarien, die zukünftig eine Reisenden-zentrierte Entscheidungsunterstützung sowie die Einbeziehung von Fahrzeugzuständen ermöglicht. Alle beschriebenen Möglichkeiten der Verwertung wurden bereits in ersten Gesprächen mit Kunden angesprochen, die das Interesse und eine Zahlungsbereitschaft hierfür signalisierten.

Die im Projekt SenseTrAln entwickelten Technologien und Methoden für das Datenbanksystem und die KI-Analyseservices haben ein hohes Transferpotenzial in andere Industriezweige, die ebenfalls von der Kombination

aus Entscheidungsunterstützung durch prädiktive Analytik, der Zusammenführung heterogener Daten- und Informationssysteme sowie dem Einsatz intelligenter Sensorik profitieren können (z.B. Luftfahrt, Maschinenbau, Energieanlagen).

Zusammenfassend lässt sich sagen, dass SenseTrAIIn dazu beitragen wird, den Betrieb von EVUs sicherer, effizienter und wirtschaftlicher zu gestalten, indem moderne Software mit Künstlicher Intelligenz kombiniert.

II.4 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Das Bestreben, moderne und modulare Datenmanagementsysteme mit KI-gestützter Analytik für EVUs aufzubauen, ist ein zentraler Pfeiler der Digitalisierung der Bahnbranche weltweit. SenseTrAIIn bildet ein wichtiges Puzzleteil in diesem größeren Bild, das sich durch eine Vielzahl von Projekten, Plattformen und spezialisierten Lösungen auszeichnet. Der Fokus liegt dabei immer stärker auf der Transformation von reaktiver oder festgetakteter Systemanalyse, bspw. in der Wartung und Instandhaltung, hin zu einer proaktiven, datengetriebenen und vorausschauenden Entscheidungsunterstützung, um die Sicherheit, Effizienz und Wirtschaftlichkeit des Schienenverkehrs zu maximieren.

Die Deutsche Bahn (DB) ist ein treibender Akteur bei der Digitalisierung der Schiene in Deutschland. Das Programm "Digitale Schiene Deutschland" umfasst zahlreiche Initiativen, darunter auch den Einsatz intelligenter Sensorik. Innerhalb dieses Programms gibt es zahlreiche KI-Projekte, die sich mit der Optimierung des Zugverkehrs, der Planung und Disposition von Zugfahrten, der Baustellenplanung und der Entwicklung von "Digitalen Zwillingen" des Schienennetzes befassen.

Bahntechnikhersteller wie Siemens Mobility, Alstom, und Stadler Rail integrieren zunehmend intelligente Sensoren in ihre neuen Fahrzeuggenerationen und entwickeln eigene Lösungen für die digitale Wartung und Überwachung der technischen Systeme. Bisher handelt es sich jedoch um proprietäre Lösungen, die den genannten Herausforderungen von EVUs zur Harmonisierung der Systeme entgegenstehen, sodass hier das Anschlusspotential der Projektpartner weiter bestehenden bleibt.

II.6 Erfolgte oder geplanten Veröffentlichungen der Ergebnisse

Die DATAbility steuerte mit Ihren gemeinsamen Arbeiten mit der vedisys AG aus dem Vorhaben SenseTrAIIn ein Praxisbeispiel zur Entwicklung und dem Einsatz eines Data Mesh für eine Veröffentlichung bei. Diese Veröffentlichung wurde aus der Arbeitsgruppe "Big Data & Advanced Analytics" des Digitalverband der deutschen Informations- und Telekommunikationsbranche bitkom e. V. unter Beteiligung der DATAbility GmbH als Mitgliedsunternehmen im Mai 2022 unter dem Titel „Data Mesh – Datenpotenziale finden und nutzen“ veröffentlicht. Diese Veröffentlichung gibt praktische Einblicke in die Arbeit des F&E Unterauftragnehmers und greift die mit der vedisys AG erarbeiteten Ansätze für dieses Vorhaben auf und fasst diese weiter für Anwendungen bezogen auf ein Mobilitätsportal.³

In den Projektjahren 2023 und 2024 veröffentlichte die vedisys AG zusammen mit der DATAbility GmbH News-Bulletins über den eigenen Newsletter sowie im Format itcs-info (www.itcs-info.de) zum Status und Verlauf des Projektes.

³ Baumann, S. et al. (2022): Data Mesh – Datenpotenziale finden und nutzen. Ein Praxisleitfaden vom Digitalverband Bitkom. Bitkom e. V., Berlin (Hrsg.). Online-Dokument, abrufbar unter: https://www.bitkom.org/sites/main/files/2022-06/220531_LF_Data_Mesh.pdf (zuletzt abgerufen am 10.07.2022)