Abschlussbericht

Talenta

Intelligente Asset-Management-Plattform für digital-Twinsn mittels Knowledge-Graph und maschinellen Lernens zur automatischen semantischen Auswertung und Verlinkung heterogener Daten aus der physischen Welt.

Zuwendungsempfänger:	**Förderkennzeichen**:
Vectorsoft AG (Konsortialführer)	01F2235A
Concedra GmbH	01F2235B
Constructor University Bremen gGmbH (kurz: CUB)	01F2235C

Projektträger:

Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR)

Projektzeitraum: vom 01.01.2023 bis zum 31.10.2025 (inkl. Kostenneutraler Verlängerung)

Berichtsdatum: 30.01.2026

Kurzfassung

Das TALENTA-Projekt entwickelte eine umfassende, interoperable und KMU-freundliche intelligente Asset-Management-Plattform für digital-Twins im Verkehrssektor, die zentrale Herausforderungen der unternehmensübergreifenden Zusammenarbeit und der Digitalisierung des Infrastrukturmanagements adressiert. Das Projekt begegnete persistierenden Hemmnissen bei der Einführung digitaler Zwillinge, darunter ein fehlendes gemeinsames Verständnis von Asset-Modellen, schwierige Systemintegration, Sicherheitsbedenken, eingeschränkter Datenzugang sowie ineffiziente Geschäftsmodelle, durch die Etablierung eines standardisierten, ontologiebasierten Informationsmodells für digital-Twins, das ein gemeinsames semantisches Verständnis über Organisationen und Sektoren hinweg ermöglicht.

Aufbauend auf dieser semantischen Grundlage entwickelte das Projekt Methoden zur kontinuierlichen und automatisierten semantischen Integration heterogener Datenquellen in sieben unterschiedlichen Formaten (relationale, semistrukturierte, textuelle, geometrische und bildbasierte Daten) sowie zur Verarbeitung dieser Datensätze zur Befüllung des Knowledge-Graphen durch Ontologie-Instanziierung. Darüber hinaus wurde ein umfassendes XAI-(Explainable Artificial Intelligence-)Toolset entwickelt, das mehr als zehn Methoden zur Quantifizierung der Erklärbarkeit (unter anderem SHAP, LIME, Partial Dependence Plots und Grad-CAM) sowie eine neuartige Ensemble-Metrik zur Bewertung der Interpretierbarkeit von Machine-Learning-Modellen kombiniert. Ein szenariobasierter intelligenter Such- und Entdeckungsmechanismus übersetzt natürlichsprachliche Anwendungsszenarien automatisch in Knowledge-Graph-Abfragen (SPARQL und Cypher) und ermöglicht so die effiziente Identifikation und den Abruf relevanter digitaler Assets.

Der Plattform-Prototyp umfasst ein sicheres und intuitives Asset-Management-Portal, das die Bereitstellung, Nutzung und kollaborative Wiederverwendung digitaler Assets über Organisationsgrenzen hinweg durch mehrstufige Authentifizierungs- und Autorisierungsmechanismen unterstützt. Die Plattform implementiert innovative, KMU-orientierte Geschäftsmodelle, Digital Twin as a Service (DTaaS) und Machine Learning as a Service (MLaaS), die darauf abzielen, Kosten zu senken und die Akzeptanz bei kleineren Unternehmen und Kommunen im Bereich des Verkehrs-infrastrukturmanagements zu erhöhen.

Die Umsetzung erstreckte sich über sechs inhaltliche Arbeitspakete (AP1–AP6) über einen Zeitraum von 30 Monaten. AP1 entwickelte die TALENTA-Ontologie zur Abbildung von Asset-Typen und zur Harmonisierung bestehender Standards; AP2 erarbeitete Methoden zur semantischen Anreicherung heterogener Daten und konstruierte den Knowledge-Graphen; AP3 konzipierte das XAI-Toolset mit Ensemble-Erklärbarkeitsmetriken; AP4 implementierte die szenariobasierte intelligente Asset-Suche; AP5 entwickelte intuitive Visualisierungs- und Interaktionsmethoden; AP6 integrierte alle Komponenten in eine Prototyp-Plattform.

Die Arbeiten waren systematisch an internationalen Standards ausgerichtet und nutzten offene Datenquellen, um Reproduzierbarkeit und Interoperabilität sicherzustellen. Zentrale Innovationen umfassen die ontologiebasierte semantische Datenintegration mit bidirektionaler

Verknüpfung heterogener Datenquellen und digitaler Assets, eine kombinierte XAI-Methodik zur Unterstützung transparenter Entscheidungsfindung im Infrastrukturmanagement, eine szenariobasierte intelligente Suche zur Verbesserung von Skalierbarkeit und Wiederverwendbarkeit von Assets sowie mehrstufige Sicherheitsmechanismen für den sicheren organisationsübergreifenden Asset-Austausch. Die Plattform adressiert gezielt die Bedürfnisse von KMU, indem sie Kosten reduziert, die operative Effizienz steigert und neue Erlöspotenziale erschließt, und leistet zugleich einen Beitrag zur deutschen Digitalisierungsstrategie für ein nachhaltiges und ressourceneffizientes Infrastrukturmanagement in Straßen- und Schienennetzen.

Teil 1 – Kurzdarstellung

1. Aufgabenstellung.

Die Aufgabenstellung des Projekts talenta bestand darin, eine ganzheitliche, interoperable und KMU-gerechte Asset-Management-Plattform für digital-Twins im Verkehrssektor zu entwickeln. Zentrales Ziel war der Aufbau eines standardisierten, ontologiebasierten Informationsmodells für Digital-Twin-Assets, das ein gemeinsames Daten- und Systemverständnis zwischen unterschiedlichen Akteuren und Organisationen ermöglicht.

Darauf aufbauend sollte das Projekt Methoden zur kontinuierlichen semantischen Integration heterogener Datenquellen entwickeln, sodass Daten in unterschiedlichen Formaten automatisiert nach einheitlicher Semantik angehoben und in einem Knowledge-Graphen zusammengeführt werden können.

Ein weiterer Bestandteil der Aufgabenstellung war die Entwicklung eines XAI-Toolsets, das die Interpretierbarkeit und Erklärbarkeit der durch maschinelles Lernen generierten Erkenntnisse verbessert. Ergänzend sollte ein intelligenter Such- und Entdeckungsmechanismus entstehen, der szenariobasierte Abfragen automatisiert in Knowledge-Graph-Abfragen übersetzt und relevante Assets identifiziert.

Schließlich umfasste die Aufgabenstellung die prototypische Entwicklung eines sicheren, intuitiven Asset-Management-Portals, das die Bereitstellung, Nutzung und Wiederverwendbarkeit digitaler Assets ermöglicht und innovative, KMU-orientierte Geschäftsmodelle wie Digital Twin as a Service (DTaaS) und Machine Learning as a Service (MLaaS) unterstützt.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Für die Durchführung des Vorhabens talenta lagen an der Constructor University Bremen (vormals Jacobs University Bremen) günstige organisatorische, fachliche und infrastrukturelle Voraussetzungen vor. Die zuständige Forschungsgruppe verfügte bereits vor Projektbeginn über ausgewiesene Expertise in den Bereichen semantisches Datenmanagement, Ontologieentwicklung, Linked Data, maschinelles Lernen und erklärbare KI (XAI) sowie in der Modellierung und Nutzung digitaler Zwillinge in technischen Anwendungskontexten. Diese Kompetenzen wurden u. a. im Projekt „Delfine“ sowie in einem Dissertationsvorhaben zur Messung der Erklärbarkeit von ML-Modellen im Supply-Chain-Kontext aufgebaut und bildeten eine wesentliche Grundlage für die im Projekt geplante Ontologieentwicklung und das XAI-Toolset.

Auf organisatorischer Ebene bestand ein eingespieltes, interdisziplinäres Team aus wissenschaftlichen Mitarbeitenden und Studierenden, das Erfahrungen in der Durchführung von Verbundprojekten mit Industriepartnern mitbrachte. Die im Vorhaben beteiligten Partner Vectorsoft AG und Concedra GmbH ergänzten die CUB durch ihre ausgewiesenen Kompetenzen in Backend-Datenmanagement, semantischer Integration sowie Frontend-

Entwicklung, Visualisierung und Portaldesign, sodass eine klare arbeitsteilige Umsetzung der Arbeitspakete möglich war.

Infrastrukturell standen an der CUB leistungsfähige Server- und Rechnerressourcen zur Verfügung, die im Projekt durch eine dedizierte Workstation und einen Server für die Verarbeitung von 3D-Modellen, den Aufbau des Knowledge-Graphen und die Ausführung rechenintensiver XAI-Algorithmen gezielt ausgebaut wurden.

Zudem war der Zugang zu relevanten offenen Datenquellen (mCLOUD, MDM, Open-Data-Portale der Deutschen Bahn und des DWD, Geodaten, 3D-Modelle von CGTrader/Sketchfab) sowie zu etablierten Standards und Ontologien (z. B. IFC, CityGML, SAREF, Automotive Ontology) bereits konzeptionell vorbereitet, sodass die im Projekt geplante semantische Integration und Knowledge-Graph-Erstellung auf einer belastbaren Datenbasis aufbauen konnte.

Schließlich schufen die Förderung im Rahmen der mFUND-Richtlinie und die damit verbundene strategische Ausrichtung des BMDV auf Digitalisierung und KI in der Mobilität einen förderpolitischen Rahmen, der die Bearbeitung der Forschungsfragen und die prototypische Umsetzung der talenta-Plattform in enger Abstimmung mit den programmatischen Zielen ermöglichte.

3. Planung und Ablauf des Vorhabens

Das Vorhaben talenta war als 30-monatiges Projekt im Zeitraum vom 01.12.2022 bis 31.05.2025 geplant. Die Arbeiten wurden in ein Projektmanagementpaket (AP0) und sechs inhaltliche Arbeitspakete (AP1–AP6) gegliedert. AP0 umfasste die übergreifende Projektkoordination und lief über die gesamte Projektlaufzeit.

Die inhaltliche Bearbeitung folgte einer klar abgestuften fachlichen Logik: In AP1 (12/2022–11/2023) sollte zunächst die „talenta“-Ontologie als standardisiertes, ontologiebasiertes Assetmodell entwickelt und durch typische Anwendungs-szenarien sowie Datenanforderungen im Straßen- und Schieneninfrastrukturmanagement konkretisiert werden. Darauf aufbauend war AP2 (03/2023–05/2024) für die Entwicklung der Methoden zum Semantic Uplift heterogener Datenquellen und den Aufbau des „talenta“-Knowledge-Graphen zuständig. AP3 (09/2023–11/2024) hatte die Konzeption und Implementierung des XAI-Toolsets einschließlich einer Metrik zur Messung der Erklärbarkeit von ML-Modellen sowie der Kopplung dieser Modelle mit dem Knowledge-Graphen zum Ziel.

Parallel dazu war AP5 (12/2022–02/2025) von Beginn an auf die Gestaltung und schrittweise Ausreifung des „talenta“-Portals ausgerichtet, insbesondere auf Visualisierung und Interaktion mit Assets und XAI-Ergebnissen. In einer späteren Projektphase folgte AP4 (03/2024–02/2025) mit der Entwicklung des szenarienbasierten Such- und Entdeckungsmechanismus, der auf den in AP1 und AP2 geschaffenen semantischen Grundlagen aufsetzt. In der Schlussphase wurden in AP6 (09/2024–05/2025) alle zuvor entwickelten Komponenten (Ontologie, Semantic Uplift/Knowledge-Graph, XAI-Toolset, Suchmechanismus und Portal) zu

einer prototypischen „talenta“-Plattform integriert und durch mehrstufige Authentifizierungs- und Autorisierungskonzepte sowie Geschäftsmodelle (DTaaS, MLaaS) ergänzt.

Der Projektverlauf wurde durch sechs Meilensteine strukturiert, die als inhaltliche und zeitliche Orientierungspunkte dienten: M1 „Assets-Taxonomie“ (31.05.2023), M2 „Ontologiebasiertes Assetmodell der digitalen Zwillinge“ (30.11.2023), M3 „Semantische Datenanhebung auf ‚talenta‘-Knowledge-Graph“ (31.05.2024), M4 „XAI-Toolset“ (30.11.2024), M5 „Intelligente Suche zur Entdeckung der Assets“ (28.02.2025) sowie M6 „‚talenta‘-Asset-Management-Portal“ (31.05.2025). Die Arbeitspakete und ihre zeitliche Abfolge waren so ausgelegt, dass mit jedem Meilenstein eine fachlich abgegrenzte Ausbaustufe der Plattform erreicht und als Grundlage für die jeweils nachfolgenden Aufgaben genutzt werden konnte.

4. Stand der Wissenschaft und Technik an den angeknüpft wurde

Für die Durchführung des Vorhabens talenta wurde systematisch an den nationalen und internationalen Stand der Technik zu digitalen Zwillingen, semantischen Informationsmodellen, Knowledge-Graphen sowie erklärbarer KI (XAI) angeknüpft. Grundlage bildeten etablierte Standards, existierende Plattformen für digital-Twins, veröffentlichte Patente sowie wissenschaftliche Publikationen und Ontologien.

4.1 Bekannte Konstruktionen, Verfahren und Schutzrechte

Im Bereich der digitalen Zwillinge und ihrer Referenzarchitekturen wurden insbesondere folgende Standards und Konzepte berücksichtigt:

- *Industry Foundation Classes (IFC)* und ifcOWL von buildingSMART als Standard für den Informationsaustausch im Bauwesen und für BIM-basierte Infrastrukturmodelle.
- *ISO 23247* als Referenzrahmen für digital-Twins in der Fertigung.
- Von Microsoft entwickelte Digital-Twin-Ontologien sowie die SAREF-Ontologie der ETSI für intelligente Anwendungen, die als Referenz für die semantische Modellierung technischer Assets dienten.

Gesamtvorhabenbeschreibung Tale…

- Offene Domänenstandards und -ontologien wie *railML*, *CityGML*, *Geonames*, *Automotive Ontology* und weitere transportbezogene Ontologien, die für die Modellierung von Straßen- und Schieneninfrastruktur sowie Fahrzeugen herangezogen und in der „talenta“-Ontologie wiederverwendet bzw. angebunden wurden.

Auf Ebene bestehender Digital-Twin-Plattformen wurden u. a. AWS IoT TwinMaker, Azure Digital Twins, Bentley iTwin, die Vertex Digital Twin Platform sowie die Verwaltungsschale von Plattform Industrie 4.0 analysiert, um Funktionalitäten, Schnittstellen und Grenzen kommerzieller Lösungen gegenüber einem Knowledge-Graph basierten, ontologischen Ansatz zu bewerten.

Im Hinblick auf Schutzrechte wurden einschlägige Patente zu Verwaltungssystemen für digital-Twins (z. B. US20170286572A1, US20190354922A1, EP3699704B1, US10564993B2) berücksichtigt, um Überschneidungen zu vermeiden und Innovationspotenziale der eigenen Lösung – insbesondere die Kombination aus Ontologie-Alignment, Semantic Uplift, XAI-Metrik und szenariobasierter Suche – klar abzugrenzen. Die im Projekt entwickelte talenta-Ontologie, das Semantic-Uplift-Verfahren, das XAI-Toolset und der szenarienbasierte Suchmechanismus wurden eigenständig konzipiert und basieren ausschließlich auf frei zugänglichen Standards, Open-Source-Bibliotheken bzw. selbst erstellten Komponenten.

4.2 Verwendete Fachliteratur und Informations-/Dokumentationsdienste

Wissenschaftlich knüpfte das Projekt insbesondere an folgende Arbeiten und Themenfelder an:

- Grundlegende Arbeiten zum Konzept des digitalen Zwillings im Infrastrukturkontext, u. a. Kuhn (2017) zur Definition, Interoperabilität und Herausforderung digitaler Zwillinge.
- Forschung zur semantischen Integration und Ontologieentwicklung, insbesondere frühere Arbeiten der CUB/JUB zu semantischer Middleware und Linked-Data-basierten Energiesystemen (z. B. Wicaksono et al. 2021; Schneider, Wicaksono & Ovtcharova 2019).
- Methoden des Semantic Uplift und der Knowledge-Graph-Erstellung, einschließlich Techniken zur Abbildung von relationalen, halbstrukturierten und unstrukturierten Daten auf RDF/OWL (R2RML, JSON-LD, RML; OntoCAD, YOLO für geometrische und bildbasierte Daten).
- Arbeiten zu erklärbarer KI und Evaluationsmetriken, insbesondere der Überblick von Zhou et al. (2021) zu XAI-Methoden und Qualitätsmaßen, der als fachliche Grundlage für die im Projekt entwickelte Ensemble-Metrik zur Erklärbarkeitsbewertung diente.
- Beiträge zu transformerbasierten Sprachmodellen (z. B. Semantics-aware BERT) und NLP-zu-SPARQL/Cypher-Ansätzen, die für die szenariobasierte Übersetzung natürsprachlicher Abfragen in Knowledge-Graph-Abfragen herangezogen wurden.

Zur Informations- und Dokumentationsrecherche wurden insbesondere die Online-Portale der Standardisierungsorganisationen (W3C, buildingSMART, ETSI, railML, OGC/CityGML, Geonames), die Open-Data-Plattformen mCLOUD, MDM, die offenen Datenportale der Deutschen Bahn sowie weitere Open-Data-Dienste (z. B. DWD, kommunale Geodatenportale) genutzt. Diese Quellen bildeten die Grundlage für die Identifikation geeigneter Szenariodaten, für die Auswahl und Wiederverwendung bestehender Datenmodelle sowie für die praktische Ausgestaltung der Semantic-Uplift-Verfahren.

Insgesamt wurden somit etablierte wissenschaftliche Erkenntnisse, Standards und Plattformen systematisch ausgewertet und gezielt mit eigenen, neu entwickelten Verfahren (Ontologie-Alignment, Semantic Uplift, XAI-Metrik, szenariobasierte Suche, mehrstufige

Authentifizierung) kombiniert, um den Innovationsgehalt der talenta-Lösung klar über den Stand der Technik hinauszuführen.

5. Zusammenarbeit mit anderen Stellen

Im Rahmen des Talenta Projektes wurden unter anderem Folgende Organisationen ein Informationsaustausch durchgeführt, primär um an Daten bezüglich der hiesigen Infrastruktur zu gelangen. Diese sind unter anderem die A+S Consult GmbH, Autobahn.de, DB InfraGo AG (ehem. DB Netze) sowie die IZ Plan.

Teil II – Eingehende Darstellung

1. Der Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

Die nachfolgend dargestellten technischen Details dienen exemplarisch der Verdeutlichung des methodischen Ansatzes und erheben keinen Anspruch auf Vollständigkeit.

1.1 1. Ziel: Entwicklung eines standardisierten semantischen DA-Modells und von Beziehungen zw. den DA, um ein gemeinsames Verständnis von Daten und Modellen zwischen Unternehmen/Organisationen zu fördern

Dieses Ziel wird durch AP1 – das ontologiebasierte Asset-Modell – erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP1 dargestellt.

1.1.1 Anwendungsszenarien und Datenanforderungen

Anhand von drei Anwendungsszenarien wurden Daten ermittelt:

Szenario 1: Bahnschwellen. Durch sogenannten Betonkrebs als mögliche Ursache von Zugunglücken werden die verbauten Bahnschwellen in den Fokus genommen. Nutzungsdauer, sowie Intensität im gesamten Streckennetz werden abgefragt.

Szenario 2: Shared Mobility. Eine Möglichkeit den öffentlichen Nah- und Fernverkehr für Nutzer*innen attraktiver zu gestalten ist die Verknüpfung mit Shared Mobility.

Szenario 3: Autoreisezüge. Der Bereich der Autoreisezüge stellt einen weiteren Ansatz dar. Durch Einsicht in bestehende Kapazitäten könnte diese Fortbewegungsart interessanter für eine größere Nutzerbasis werden. Hierbei sind die Gegebenheiten der verschiedenen Waggons und die Variationen der Fahrzeuge und Fahrzeugtypen zu berücksichtigen.

1.1.2 Modellierung der Assets in der Ontologie

Die Talenta-Ontologie wurde als ein standardisiertes und formalisiertes semantisches Modell entwickelt, das die Strukturen, Funktionalitäten und Beziehungen von Digital-Twin-Assets beschreibt. Die Konstruktion der Ontologie begann mit der detaillierten Definition von Use-Case-Szenarien sowie einer Analyse der Datenanforderungen für das Management von Verkehrsinfrastrukturen. Die Ontologie wurde zudem in Übereinstimmung mit der konzeptionellen Architektur von Talenta entwickelt, wie sie in Abbildung 1.1 dargestellt ist.

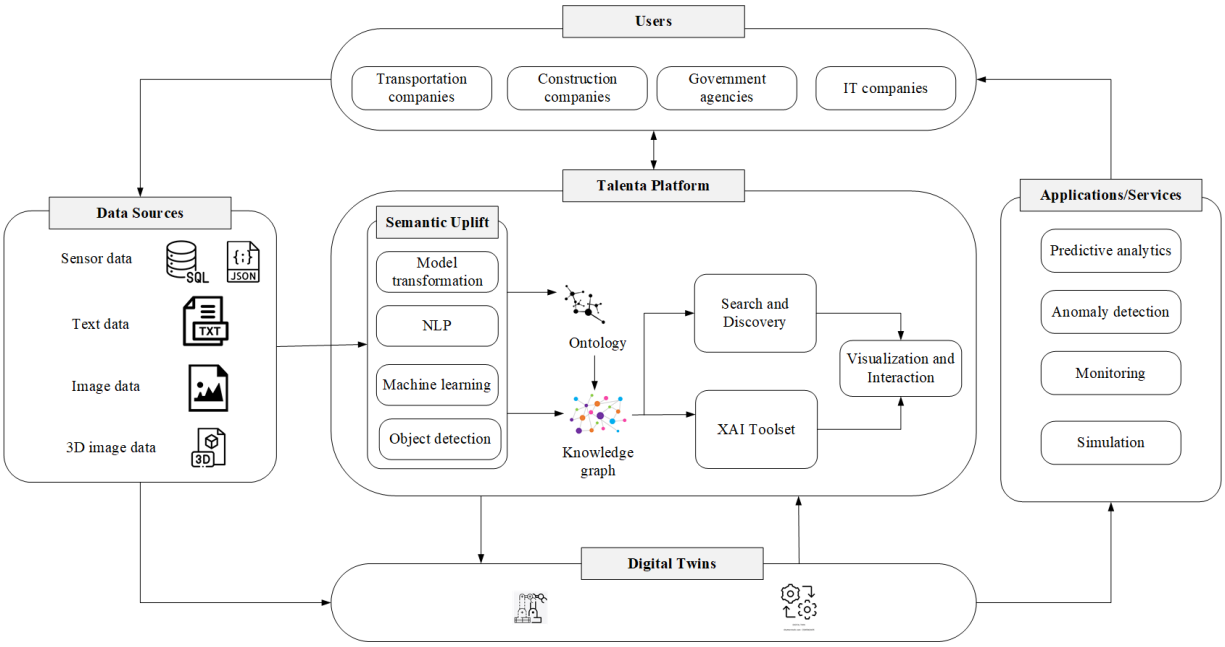


Abb 1.1. Die Talenta-Plattformarchitektur

Die Ontologie repräsentiert ein umfassendes Asset-Modell, das Verkehrsinfrastruktur-Assets und deren wesentliche Attribute umfasst:

1. Zeitstempel (Timestamp): Informationen zur Nachverfolgung der Erstellung, Änderung und Entfernung von Assets
2. Geometrie (Geometry): Verweise auf geometrische Informationen (2D-Zeichnungen, 3D-Modelle)
3. Struktur (Structure): Eigenschaften und strukturelle Spezifikationen von Assets (Materialien, Prozessspezifikationen, Funktionalitäten, Komponenten)
4. Standort (Location): Tatsächlicher Standort physischer Objekte, die als digitale Assets repräsentiert werden
5. Datenquelle (DataSource): Verweise (URLs) auf erforderliche Datenquellen (Sensordaten, Prozesszustandsdaten)
6. Berechtigungen (Credential): Informationen über Autoren, Eigentümer und Wartungsverantwortliche der Assets.

Abbildung 1.2 zeigt die partielle Klassenhierarchie der Talenta-Ontologie.

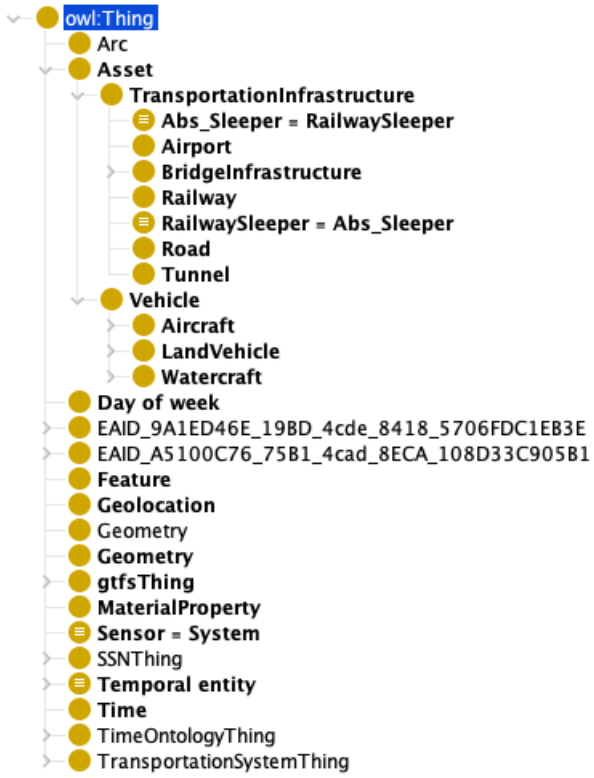


Abb 1.2. Die Talenta-Ontologie

1.1.3 bestehender Standards und Ontologien aus verschiedenen Domänen

Um eine breite Systeminteroperabilität zu gewährleisten und die Integration bestehender (Legacy-)Modelle zu erleichtern, identifizierte und integrierte das Projekt systematisch relevante Begriffe und Strukturen aus internationalen Standards und etablierten Ontologien, wie Railway, IFC, Geodaten, ISO 21972, GTFS, SOSA, SSN und Bridge, um Wissen aus heterogenen Domänen zu konsolidieren. Detaillierte Beziehungen dieser Ontologien zur Talenta-Ontologie sind in Tabelle 1.1 dargestellt.

Tabelle 1.1. Ausrichtung der Talenta-Ontologie mit anderen anerkannten Ontologien.

Talenta-Ontologieklasse	Quelle	Beschreibung
gtfsThing	https://lov.linkeddata.es/dataset/lov/vocabs/gtfs	Übersetzt GTFS in URIs, um eine Datenaustauschplattform zu ermöglichen, in der das Linked-GTFS-Modell als Grundlage für die Bereitstellung korrekt formatierter und interoperabler Daten dient.
Sensor	https://www.w3.org/TR/vocab-sosa/	Stellt ein modulares Framework zur Beschreibung von Sensoren, Beobachtungen, Aktuatoren und zugehörigen Prozessen bereit und unterstützt ein breites Anwendungsspektrum von wissenschaftlicher Überwachung bis hin zu industriellen Infrastrukturen und dem Web of Things.
SSNThing	https://www.w3.org/TR/vocab-ssn/	Bietet zusammen mit seinem leichtgewichtigen Kern SOSA ein modulares semantisches Framework zur Modellierung von Sensoren, Beobachtungen, Prozeduren und Aktuatoren und unterstützt eine Vielzahl von Anwendungen von großskaliger

		Überwachung bis hin zu industriellen Infrastrukturen und dem Web of Things.
TimeOntologyThing	https://www.w3.org/TR/owl-time/	Stellt ein standardisiertes Vokabular zur Repräsentation zeitlicher Positionen, Dauern sowie Ordnungsbeziehungen zwischen Zeitpunkten und Zeitintervallen unter Verwendung mehrerer zeitlicher Referenzsysteme bereit.
TransportationSystemOntology	http://ontology.eil.utoronto.ca/icity/TransportationSystem/1.2/	Verknüpft das Verkehrsnetz und die Verkehrsinfrastruktur, indem ein Arc mit einem Transportation Complex (oder einem anderen Straßensegment) in Beziehung gesetzt wird.
BridgeInfrastructure	https://github.com/Alhakam/bridgeOntology [2]	Stellt ein generisches und erweiterbares Framework zur Modellierung von Brückenkonstruktionen bereit, einschließlich Komponenten, Zonen sowie deren topologischen und vertikalen Beziehungen, anwendbar auf unterschiedliche Brückentypen.

1.1.4 Ontologie-Verlinkung

Die Ontologieverknüpfung bzw. eigenschaftsbasierte Alignment wurde unter Verwendung einer fortgeschrittenen Ensemble-Technik durchgeführt. Dieser Prozess kombiniert lexikalische, strukturelle und semantische Ähnlichkeitsmaße, um formale Korrespondenzen herzustellen. Dabei wurden bidirektionale Abbildungen zwischen Klassen und Eigenschaften der Talenta-Ontologie und Elementen externer Ontologien etabliert, wie in Tabelle 3.1 dargestellt.

1.2 2. Ziel: Entwicklung eines Ansatzes zur kontinuierlichen semantischen Integration von Daten mit verschiedenen Formaten und aus heterogenen Quellen

Dieses Ziel wird durch AP2 – das Semantic-Uplift mittels des Ontologien-Modells – erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP2 dargestellt.

1.2.1 Analyse und Auswahl der Datenquellen

Relevante Datenquellen wurden systematisch identifiziert und ausgewählt, basierend auf spezifischen Use-Case-Szenarien im Kontext des Managements von Verkehrsinfrastrukturen. Der Auswahlprozess nutzte mehrere Kriterien, um eine umfassende und qualitativ hochwertige Datenintegration sicherzustellen. Die Datenquellen wurden hinsichtlich ihrer Relevanz für Szenarien des Straßen- und Schieneninfrastrukturmanagements, der Qualität und Vollständigkeit der verfügbaren Daten, der Zugänglichkeit über APIs oder automatisierte Datenfeeds sowie der Abdeckung aller Phasen des Asset-Lebenszyklus – einschließlich Planung, Betrieb und Instandhaltung – bewertet.

Die Analyse identifizierte sieben Hauptkategorien von Datenquellen, die für das Projekt geeignet sind. Relationale Datenbanken, bestehend aus CSV- und SQL-Datenbanken von Verkehrsmanagementsystemen, lieferten strukturierte operative Daten. Semi-strukturierte

Daten in JSON- und XML-Formaten aus Telemetrie- und Leitsystemen boten Flexibilität für unterschiedliche Datenrepräsentationen. Unstrukturierte Textdaten aus Wartungsprotokollen, Störungsberichten und operativer Dokumentation erforderten spezialisierte Verarbeitungstechniken. Bild- und Geometriedaten aus CAD-Zeichnungen (DWG-, DXF-Formate) sowie dreidimensionalen Modellen (OBJ, FBX) stellten räumliche Informationen zu Assets bereit. Echtzeit-Fahrplaninformationen im GTFS-Format (General Transit Feed Specification) erfassten dynamische Betriebsdaten. Schließlich lieferten Open-Data-Quellen wie mCLOUD, MDM, die Open-Data-Portale der Deutschen Bahn sowie Wetterdaten des DWD wertvolle Kontextinformationen für das Infrastrukturmanagement.

Bestimmte Datenquellen erforderten spezielle methodische Ansätze, insbesondere im Hinblick auf Datenerhebungs- und Vorverarbeitungstechniken, um eine Weiterverarbeitung innerhalb des Machine-Learning- und Evaluierungsrahmens des Projekts zu ermöglichen (Abbildung 1.3). Die methodische Strategie basierte auf der Architektur der Talenta-Plattform (Abbildung 1.1) und stellte die Übereinstimmung zwischen den Datenverarbeitungspipelines und dem Systemdesign sicher.

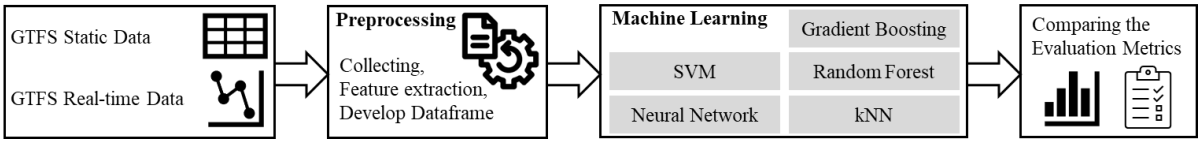


Abb 1.3. Methodischer Ansatz zur GTFS-Datenerhebung und -vorverarbeitung für die Weiterverarbeitung im Talenta-Projekt.

Beispielhafte GTFS-Daten wurden von der nationalen Eisenbahngesellschaft Deutschlands, der Deutschen Bahn (https://developer-docs.deutschebahn.com), sowie von GTFS für Deutschland (https://gtfs.de) bezogen. Diese Quelle stellt täglich generierte Fahrplandaten für den gesamten Fern- und Regionalverkehr der Deutschen Bahn sowie für den lokalen Verkehr aller Verkehrsverbünde und Anbieter in Deutschland bereit.

- Datenerhebung: Die GTFS-Daten umfassten statische und Echtzeit-Komponenten. Die statischen Daten beinhalteten geplante Haltestellen, Routen und Fahrzeiten, während die Echtzeitdaten Aktualisierungen zu Ankunfts- und Abfahrtszeiten, Fahrten und Haltestellen enthielten.
- Merkmalsextraktion: Die Merkmalsextraktion umfasste die Identifikation und Auswahl relevanter Attribute aus den GTFS-Daten, die die Vorhersage von Verspätungen beeinflussen können, wie geplante Ankunfts- und Abfahrtszeiten sowie tatsächliche Ankunfts- und Abfahrtszeiten. Dieser Schritt war entscheidend, um die Lernfähigkeit der Modelle zu verbessern und präzise Vorhersagen zu ermöglichen. Die aus den statischen und Echtzeitdaten ausgewählten Merkmale sind in Tabelle 1.2 dargestellt.
- Entwicklung eines Dataframes: Nach der Datenerhebung und Merkmalsextraktion wurde ein umfassender Dataframe entwickelt. Dieser Prozess begann mit der Zusammenführung der statischen und Echtzeit-Datensätze anhand gemeinsamer Attribute, wodurch sichergestellt wurde, dass die geplanten und tatsächlichen Zeiten jeder Fahrt korrekt ausgerichtet sind.

Tabelle 1.2. Ausgewählte Merkmale aus dem statischen und Echtzeit-GTFS-Datensatz.

Aus statischem Datensatz	Aus Echtzeit-Datensatz
trip_id	trip_id
service_id	StartDate
stop_sequence	stop_sequence
stop_id	stop_id
arrival_time (scheduled)	arrival_time (actual)
departure_time (scheduled)	departure_time (actual)
stop_name	

Der Datenerhebungsprozess begann mit dem Abruf statischer GTFS-Daten, die geplante Haltestellen, Routen und Fahrzeiten enthielten. Da der Echtzeit-Datensatz tägliche Aktualisierungen erhält, wurde ein dynamisches System implementiert, das diese Daten kontinuierlich abruft und speichert. Zu diesem Zweck wurde ein Python-Skript (Algorithmus 1) entwickelt, das täglich ausgeführt wird, um die Echtzeitdaten automatisch herunterzuladen und im JSON-Format zu speichern. Ein ergänzendes Skript konvertierte diese JSON-Daten in das CSV-Format und führte sie mit dem statischen GTFS-Datensatz zusammen, wodurch eine einheitliche Datenstruktur für die nachfolgende Analyse entstand.

Die verschiedenen Dateien des statischen GTFS-Datensatzes wurden anschließend in einer Reihe systematischer Schritte, die in Algorithmus 2 beschrieben sind, zu einem einzigen Dataframe zusammengeführt. Dieser Organisationsprozess erwies sich als wesentlich für die anschließende Zusammenführung mit dem Echtzeit-Datensatz, die anhand der drei gemeinsamen Attribute Trip-ID, Stop-ID und Stop-Sequenz durchgeführt wurde.

Nach der Zusammenführung der statischen und Echtzeit-Datensätze wurde eine neue Spalte mit der Bezeichnung „Delay“ in den Dataframe eingeführt. Diese Spalte wurde als Differenz zwischen der tatsächlichen Ankunftszeit und der geplanten Ankunftszeit einer Fahrt an einem bestimmten Datum berechnet und diente als Zielvariable für die Machine-Learning-Modelle. Die detaillierten Schritte dieses Prozesses sind in Algorithmus 3 beschrieben.

Der finale Dataframe wurde so strukturiert, dass er eine umfassende Analyse verschiedener Faktoren ermöglicht, die Verspätungen beeinflussen. Das Organisationsschema umfasste Spalten für Ursprungs- und Zielinformationen, Wochentage, geplante Zeiten sowie berechnete Verspätungswerte. Eine neue Spalte mit dem Namen „Origin“ wurde hinzugefügt, um die Haltestellennamen zu enthalten, von denen jedes Fahrzeug startete, während die bestehende Spalte „stop_name“ in „Destination“ umbenannt wurde, um die Zielhaltestelle des Fahrzeugs darzustellen. Diese in Algorithmus 4 beschriebene Reorganisation ermöglichte eine systematische Analyse des Einflusses verschiedener Variablen – einschließlich Haltestellen, Routen, Fahrzeiten und Wochentagen – auf Verkehrsverspätungen im gesamten Netzwerk.

Algorithm 1 Download and Process Real-Time Data

Send request to `https://realtime.gtfs.de/realtime-free.pb` to download real-time data.
if request is successful (status code 200) **then**
    Parse the protocol buffer content using `gtfs_realtime_pb2.FeedMessage()`.
    Convert the parsed content into a dictionary.
    Open the CSV file where real-time data is stored.
    for each trip_update in the parsed content **do**
        Extract `trip_id`, `start_date`, and stop_time updates.
        for each stop_time update **do**
            Extract arrival and departure times.
            Write trip details to the CSV file where real-time data is stored.
        end for
    end for
end if

Algorithm 2 Constructing Data Frame from Static GTFS Dataset

for each file inside the unzipped file of the static dataset **do**
    Extract the name of the file.
    Read the file using Python pandas `read_csv` function.
    Store the data from the file in a dictionary with the file name as the key.
end for
From the dictionary:
  Take columns: `trip_id`, `route_id`, `service_id` from `dictionary['trips']`.
  Take columns: `trip_id`, `arrival_time`, `departure_time`, `stop_id`, and `stop_sequence` from `dictionary['stop_times']`.

Algorithm 3 Merging Real-Time and Static Data Frames

Read both the static data frame constructed from 2 and real-time data frame constructed from 1.
Merge these two data frames using Python pandas `merge()` function on the features: `trip_id`, `stop_id`, and `stop_sequence`.
Add a column `delay` to the merged data frame.
for each row in the merged data frame **do**
    Calculate the value for `delay` by subtracting the real-time arrival time from the scheduled arrival time.
end for
Output the final merged data frame into another CSV file.

Algorithm 4 Rearrange Data Frame for 'Origin' and 'Destination'

Load the merged data frame from 3.
for each row within the data frame **do**
    if the `stop_sequence` is 0 **then**
        Set `Origin` and `Destination` columns to the same stop name.
    else
        Set the `Origin` for the current stop to the stop name at the previous `stop_sequence` (current `stop_sequence` - 1).
    end if
end for

1.2.2 Datenschema-Ontologie-Mapping

Client auf basis von NEO4J entwickelt und mit einer Ontologie initialisiert (s. Abb.8)
NEO4J-Client erweitert, sodass Ontologien und Daten hochgeladen werden können, und sind durch CYPHER interagierbar (s. Abb.7)
Entwicklung einer Anwendung für den GTFS-Datenkonverter, die das Erfassen dynamischer GTFS-Daten, die Integration statischer und dynamischer GTFS-Daten sowie das Hinzufügen einer Verzögerungsspalte umfasst.

1.2.3 semantischen Erhebung von halbstrukturierten Daten

Konzept erstellt auf der Basis von JSON und CSV. Semantischer Erfassungsansatz umgesetzt und in Client integriert. Ermöglicht das Importieren von halbstrukturierten Daten als Dateien.

1.2.4 semantischen Erhebung von Textdaten mittels Natural Language Processing (NLP)

Das Talenta-Projekt implementiert zwei komplementäre Frameworks der natürlichen Sprachverarbeitung (Natural Language Processing, NLP) zur Extraktion semantischer Informationen aus unstrukturierten Dokumentationen der Verkehrsinfrastruktur. Diese ermöglichen eine robuste Entitätserkennung über verschiedene Dokumenttypen und sprachliche Variationen hinweg. Die Dual-Strategie nutzt sowohl spaCy-basierte als auch Flair-basierte Systeme zur Named Entity Recognition (NER), die jeweils für unterschiedliche Anwendungsfälle optimiert sind.

Der spaCy-Ansatz verwendet spacy.load(‘en_core_web_lg’) für Englisch und spacy.load(‘de_core_news_lg’) für Deutsch und bietet schnelle Inferenzzeiten mit Wort-Einbettungen, die sich für den produktiven Einsatz eignen. Diese Methode verarbeitet PDF-Dokumente mittels PyMuPDF (fitz) und kombiniert schlüsselwortbasierte Entitätsabgleiche mit statistischer NER, wodurch eine schnelle Entitätsextraktion bei geringem rechnerischem Aufwand erreicht wird.

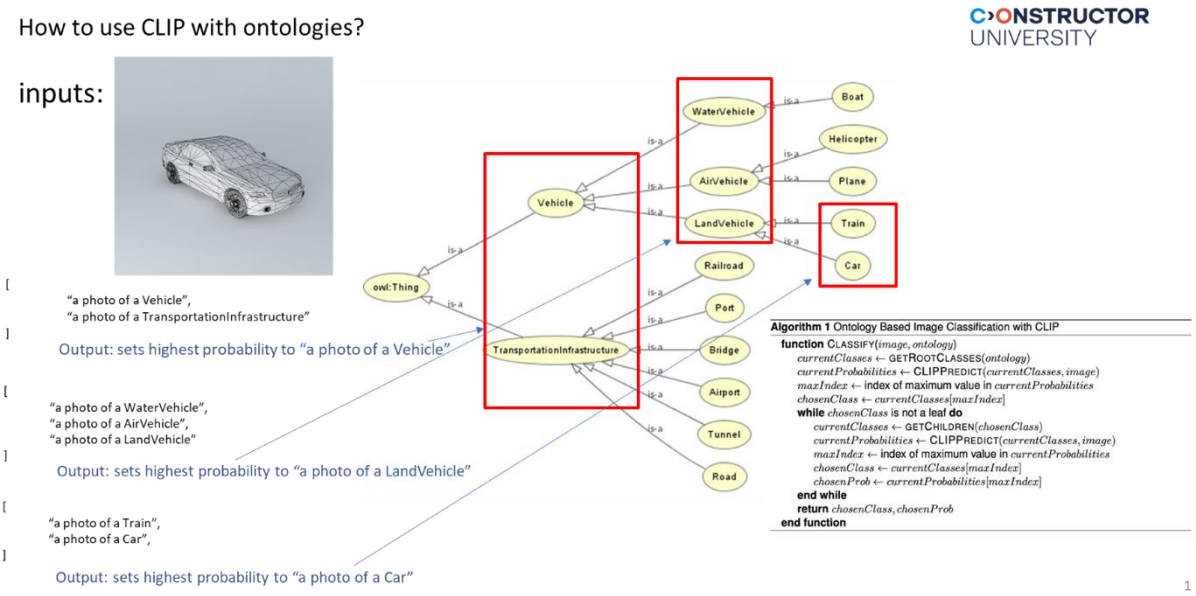
Im Gegensatz dazu nutzt der Flair-basierte Ansatz kontextuelle String-Einbettungen über SequenceTagger.load(‘ner-large’) und SequenceTagger.load(‘de-ner-large’), was ein überlegenes Kontextverständnis ermöglicht und mehrdeutige Terminologie im Verkehrsbereich mit höherer Genauigkeit verarbeitet. Beide Ansätze integrieren langdetect zur automatischen Sprachzuordnung, regex-basierte Attributextraktion für infrastrukturspezifische Muster (z. B. Bauzeiträume, Brückenlängen) sowie Sentence-BERT (paraphrase-multilingual-MiniLM-L12-v2) zur semantischen Abbildung auf die Talenta-Ontologie. Beide Systeme teilen sich identische RDF-Serialisierungspipelines, ein einheitliches Namespace-Management für Brückenklassen innerhalb der Ontologie (BRCOMP, BMAT, BRIDGE, PROP) sowie die Integration in den Knowledge-Graphen über Neo4j.

1.2.5 semantischen Annotation von Geometrie und Bilddaten mittels Geometriemustererkennung

Die Methodik adressiert die Integration von 3D-Objekten in die Talenta-Ontologie, ohne ein erneutes Training des Modells zu erfordern, indem CLIP genutzt wird – das Zero-Shot-Klassifikationsmodell von OpenAI, das mit 400 Millionen Bild-Text-Paaren trainiert wurde. Während der Vorhersage berechnet CLIP Ähnlichkeitswerte zwischen einem Bild und textuellen Beschreibungen mittels innerer Produkte und wählt die Klasse mit dem höchsten Wert aus, wie in Abbildung 1.4 dargestellt.

Das System führt den Algorithmus unabhängig für jedes gerenderte Bild aus und erzeugt dabei mehrere Klassifikationsergebnisse. Anstatt die Wahrscheinlichkeiten zu mitteln (was häufig korrekte Klassifikationen fälschlicherweise benachteiligen würde), summiert der Algorithmus alle Wahrscheinlichkeiten für jede vorhergesagte Klasse und wählt die Klasse mit dem höchsten kumulativen Wert aus. Diese Gewichtung stellt sicher, dass sowohl die Häufigkeit der Vorhersagen als auch die Konfidenzwerte angemessen berücksichtigt werden.

Das klassifizierte Objekt wird in die Ontologie integriert, indem eine individuelle Instanz unter der vorhergesagten Klasse erzeugt wird. Die Instanzmetadaten speichern den Erstellungszeitstempel, den ursprünglichen Dateipfad des 3D-Modells sowie alle vorhergesagten Datenproperty-Werte. Die Namenskonvention ergänzt den Klassennamen um einen numerischen Index, um eindeutige Identifikatoren zu erzeugen. Dieser vollständige Workflow – bestehend aus Multi-View-Rendering, hierarchischer Ontologie-Traversal, Zero-Shot-Klassifikation mittels CLIP und automatisierter Erstellung von Ontologie-Instanzen – erfordert kein zusätzliches Modelltraining und passt sich nahtlos an jede korrekt benannte Ontologiestruktur an, wodurch er sich in hohem Maße für Digital-Twin-Anwendungen in der Verkehrsinfrastruktur und darüber hinaus eignet.



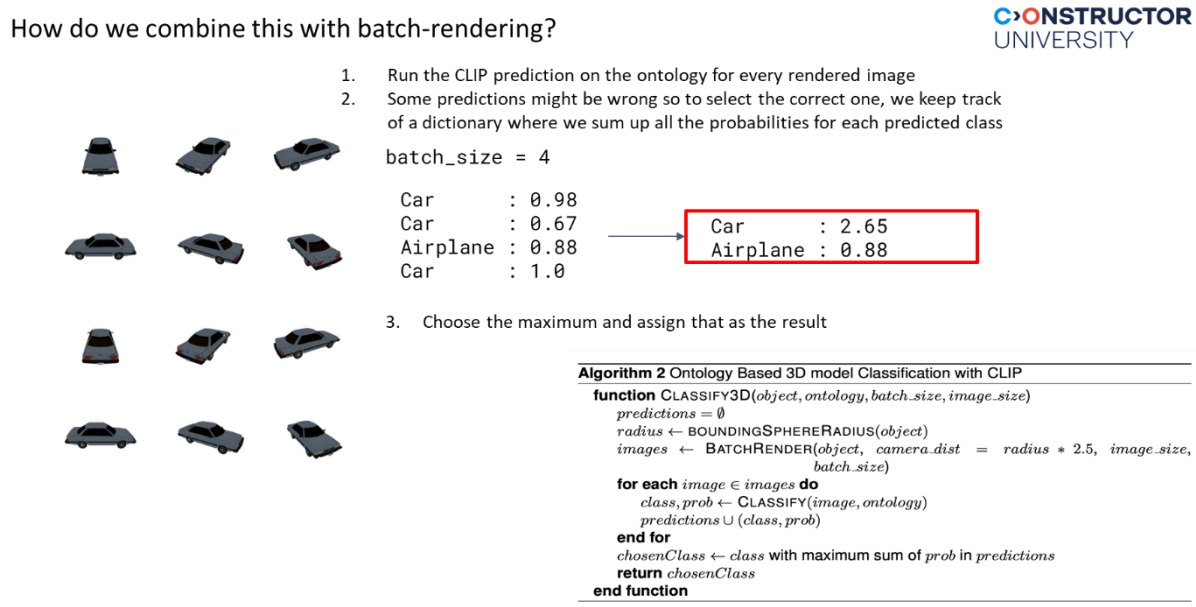


Abb 1.4. Semantischen Annotation von 3D-Bilddaten

1.2.6 Automatische Verlinkung zwischen Datenquellen und „talenta"-Ontologie

Die Entwicklung einer automatischen Verknüpfung zwischen Datenquellen und der „Talenta"-Ontologie sowie die Befüllung/Population der „Talenta"-Ontologie werden mithilfe von APIs für die jeweiligen Anwendungsfälle umgesetzt, darunter:

- Entwicklung einer Anwendung für den GTFS-Datenkonverter, die das Erfassen dynamischer GTFS-Daten, die Integration statischer und dynamischer GTFS-Daten sowie das Hinzufügen einer Verspätungsspalte umfasst.
- Entwicklung von Anwendungen zur Verspätungsvorhersage unter Verwendung von Methoden des maschinellen Lernens (ML).

- Methoden zur Erkennung geometrischer Muster, wie z. B. CLIP, Grad-CAM und Patch Detector.
- Eine Methode zur semantischen Extraktion von Textdaten mittels Natural Language Processing (NLP) befindet sich in Entwicklung (Abschnitt 3.1.2.4).
- Ein Konzept zur semantischen Annotation von Geometrie- und Bilddaten unter Verwendung geometrischer Mustererkennung, wie z. B. CLIP, Grad-CAM und Patch Detector (Abschnitt 1.1.2.5).

1.3 3. Ziel: Entwicklung eines Toolsets/ Modellbibliothek zur Verbesserung der Interpretierbarkeit und Erklärbarkeit der erhobenen Daten.

Dieses Ziel wird durch AP3 – das XAI-Toolsets/Modellbibliothek zur Verbesserung der Interpretierbarkeit und Erklärbarkeit – erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP3 dargestellt.

1.3.1 Spezifikation der abhängigen und unabhängigen Variablen für die ML-Algorithmen

Für tabellarische Daten wurden die unabhängigen Variablen aus den GTFS-Daten extrahiert, einschließlich geplanter Ankunfts- und Abfahrtszeiten, Haltestellenfolgen und Fahrzeiten (siehe Tabelle 3.2). Die abhängige Zielvariable ist die Verspätung, definiert als Differenz zwischen tatsächlicher und geplanter Ankunftszeit.

1.3.2 Definition von Anforderungen an den "Explainer" der ML-Modelle

Moderne öffentliche Verkehrssysteme erzeugen große Datenmengen aus unterschiedlichen Quellen. Die Datenqualität in Bezug auf Vollständigkeit und Genauigkeit ist entscheidend für den Aufbau zuverlässiger Vorhersagemodelle. Ungenaue oder unvollständige Daten können die Zuverlässigkeit von Verspätungsvorhersagen erheblich beeinträchtigen. Daher ist die Sicherstellung hochwertiger Daten grundlegend, um die Genauigkeit der Verspätungsvorhersagen zu verbessern.

Selbst bei genauen Daten und Vorhersagen steht die Einführung von Machine-Learning-(ML)-Modellen jedoch vor einer weiteren großen Herausforderung: ihrer „Black-Box“-Natur. Verkehrsplaner und Entscheidungsträger müssen die Begründung hinter Modellvorhersagen verstehen, um fundierte Entscheidungen treffen zu können. Ohne Interpretierbarkeit könnten diese Akteure den Modellen misstrauen und zögern, sie in der Praxis einzusetzen. Die Verbesserung der Transparenz und Erklärbarkeit von ML-Modellen ist daher entscheidend für deren effektive Implementierung. Dies umfasst die Entwicklung von Modellen, die nicht nur genaue Vorhersagen liefern, sondern auch klare Einblicke in die Faktoren geben, die diese beeinflussen.

Die Anforderungen an einen Erklärer (Explainer) umfassen sowohl globale Einblicke (Bedeutung von Merkmalen im gesamten System) als auch lokale Erklärungen (Begründung

für eine spezifische Vorhersage). Explainer müssen robust gegenüber geringfügigen Änderungen der Eingabedaten sein und intuitive Visualisierungen bereitstellen, um Vertrauen bei den Stakeholdern aufzubauen.

1.3.3 Evaluierung und Auswahl geeigneter Algorithmen

Im Projekt wurden mehrere ML-Algorithmen verglichen, darunter Gradient Boosting, Random Forest, Support Vector Machines (SVM), neuronale Netze (MLP) und k-Nearest Neighbors (kNN).

Gradient Boosting erstellt robuste Modelle, indem mehrere schwache Lerner kombiniert werden, um den Fehler zu minimieren, und ist bekannt für seine hohe Genauigkeit sowie die Fähigkeit, komplexe, nichtlineare Zusammenhänge zu erfassen. Random Forest, eine weitere Ensemble-Methode, verwendet mehrere Entscheidungsbäume, um Overfitting zu reduzieren und die Vorhersagegenauigkeit zu verbessern; es arbeitet gut mit großen Datensätzen und komplexen Interaktionen. SVM sind effektiv für Klassifikationsaufgaben und eignen sich gut für hochdimensionale Daten, da sie die optimale Trennlinie (Hyperplane) finden, die Datenpunkte in unterschiedliche Klassen unterteilt. Neuronale Netze, insbesondere Deep-Learning-Modelle, können automatisch Merkmale aus Rohdaten extrahieren und komplexe Muster lernen; sie erfordern jedoch erhebliche Rechenressourcen und sorgfältige Hyperparameter-Abstimmung, bieten dafür aber hohe Flexibilität und Vorhersagekraft. kNN ist zwar einfacher, eignet sich jedoch in Situationen, in denen ähnliche historische Fälle wertvolle Informationen liefern können; es ist leicht implementierbar und interpretierbar, kann aber bei komplexen Zusammenhängen Schwierigkeiten haben.

Die Leistung der verschiedenen ML-Modelle wurde anhand von drei Schlüsselmetriken bewertet: MAE, MSE und RMSE. Die betrachteten Modelle umfassen Gradient Boosting, Random Forest, neuronale Netze, SVM und kNN. Die Ergebnisse sind in Tabelle 1.3 zusammengefasst.

Tabelle 1.3. Ergebnisse der ML-Modelle aus GTFS-Daten.

ML-Algorithmen	MAE	MSE	RMSE
Gradient Boosting	1.34	67.6	3.59
Random Forest	3.39	80.1	4.09
neuronale Netze	1.3	67.1	3.59
SVM	1.1	79.4	3.46
kNN	4.35	102.5	5.12

Zum Vergleich erzielte das SVM-Modell den niedrigsten MAE von 1,1, was darauf hindeutet, dass die Vorhersagen des SVM-Modells im Durchschnitt den tatsächlichen Werten am nächsten lagen. Die Modelle Neural Network und Gradient Boosting erzielten ebenfalls gute Ergebnisse mit MAEs von 1,3 bzw. 1,34. Die Modelle Random Forest und kNN wiesen deutlich höhere MAEs auf, wobei kNN mit 4,35 am schlechtesten abschnitt. Dies deutet darauf hin, dass SVM, Neural Network und Gradient Boosting in ihren Vorhersagen genauer sind als Random Forest und kNN.

Das Neural-Network-Modell erzielte den niedrigsten MSE von 67,1, dicht gefolgt von Gradient Boosting mit einem MSE von 67,6. Dies zeigt, dass diese Modelle effektiver darin sind, größere Fehler zu minimieren als die anderen. Trotz des niedrigsten MAE hatte das SVM-Modell einen leicht höheren MSE von 79,4, was auf das Vorhandensein einiger größerer Fehler in seinen Vorhersagen hinweist. Random Forest und kNN wiesen höhere MSE-Werte von 80,1 bzw. 102,5 auf, was erneut zeigt, dass diese Modelle weniger effektiv bei der Minimierung signifikanter Fehler sind.

Das SVM-Modell erreichte den niedrigsten RMSE von 3,46, was darauf hindeutet, dass es insgesamt die genauesten Vorhersagen liefert, wenn sowohl durchschnittliche als auch größere Fehler berücksichtigt werden. Die Modelle Neural Network und Gradient Boosting hatten beide einen RMSE von 3,59, was ihre starke Leistung im Einklang mit ihren MSE-Werten zeigt. Random Forest und kNN wiesen höhere RMSE-Werte von 4,09 bzw. 5,12 auf, was ihre geringere Vorhersagegenauigkeit im Vergleich zu den anderen Modellen unterstreicht.

Die vergleichende Analyse liefert wichtige Erkenntnisse über die Leistung verschiedener Machine-Learning-Modelle basierend auf den Metriken MAE, MSE und RMSE. Die Ergebnisse zeigen, dass das SVM-Modell in Bezug auf MAE und RMSE durchweg bessere Ergebnisse erzielte, was seine Robustheit bei der Bereitstellung genauer Vorhersagen verdeutlicht. Trotz eines leicht höheren MSE im Vergleich zu Neural Network und Gradient Boosting deutet die Gesamtleistung von SVM auf seine Effektivität bei der Minimierung von Fehlern über den gesamten Datensatz hinweg hin. Diese Erkenntnis stimmt mit der Fähigkeit von SVM überein, die optimale Trennlinie (Hyperplane) zu finden, die Datenpunkte am besten in unterschiedliche Klassen oder Gruppen trennt, was es insbesondere für Regressionsaufgaben nützlich macht.

Das Neural-Network-Modell zeigte eine wettbewerbsfähige Leistung mit dem niedrigsten MSE, was seine Fähigkeit zur Minimierung größerer Vorhersagefehler verdeutlicht. Neuronale Netze sind bekannt für ihre Fähigkeit, komplexe Muster in Daten zu erfassen, was ihre Effektivität in dieser Studie erklären könnte. Gradient Boosting erzielte ebenfalls gute Ergebnisse über alle Metriken hinweg und zeigt seine Zuverlässigkeit und Robustheit bei prädiktiven Aufgaben. Gradient Boosting verbessert die Modellleistung iterativ, indem mehrere schwache Lerner kombiniert werden, was es für eine Vielzahl von Problemen effektiv macht.

Random Forest und kNN zeigten hingegen eine schlechtere Leistung im Vergleich zu SVM, Neural Network und Gradient Boosting. Obwohl Random Forest eine leistungsstarke Ensemble-Lernmethode ist, erzielte es in diesem Kontext keine so guten Ergebnisse, was darauf hindeutet, dass seine Entscheidungsbäume die zugrunde liegenden Muster in den Daten möglicherweise nicht effektiv erfassen. Ebenso war die Leistung von kNN deutlich schlechter, was auf seine Empfindlichkeit gegenüber Rauschen und die geringere Eignung für diese spezifische Vorhersageaufgabe hinweist.

Die Leistungsunterschiede dieser Modelle verdeutlichen die Bedeutung, ihre Stärken und Schwächen in Bezug auf spezifische Datensätze und Vorhersageziele zu verstehen. Faktoren wie Datencharakteristika, Modellkomplexität und Interpretierbarkeit spielen eine entscheidende Rolle bei der Modellauswahl und der Leistungsbewertung.

1.3.4 Entwicklung eines neuartigen „Explainers“

Das Projekt entwickelte einen integrierten Explainer-Ansatz, der Shapley Additive Explanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), Partial Dependence Plot (PDP) und Gradient-weighted Class Activation Mapping (Grad-CAM) kombiniert.

Der Grund für die Kombination dieser vier Methoden liegt darin, dass sie mehrere zentrale Dimensionen des XAI-Designs repräsentieren. Erstens decken sie sowohl modell-agnostische als auch modell-spezifische Ansätze ab, da SHAP und LIME als modell-agnostische Erklärer eingesetzt werden können, während Grad-CAM eine modell-spezifische Technik ist, die auf Convolutional Neural Network (CNN)-Architekturen zugeschnitten ist. Zweitens umfassen sie sowohl lokale als auch globale Erklärungsebenen: LIME und instanzbasiertes SHAP liefern lokale Erklärungen pro Vorhersage, während PDP eine globale Sicht darauf bietet, wie Merkmale die Modellvorhersagen auf Bevölkerungsebene beeinflussen, und SHAP kann ebenfalls aggregiert werden, um globale Wichtigkeitsprofile abzuleiten. Drittens adressieren sie unterschiedliche Datenmodalitäten: SHAP, LIME und PDP werden üblicherweise auf tabellarische Daten angewendet, die in vielen Entscheidungsunterstützungssystemen dominieren, während Grad-CAM für Bilddaten entwickelt wurde und somit visuelle, hochdimensionale Eingaben repräsentiert. Schließlich profitieren alle vier Methoden von ausgereiften, aktiv gepflegten Open-Source-Bibliotheken, was die Reproduzierbarkeit und den praktischen Einsatz in industriellen Anwendungen erleichtert. Beispielsweise sind SHAP und LIME als dedizierte Python-Pakete verfügbar, PDP und seine Erweiterungen werden in Bibliotheken wie scikit-learn und Alibi Explain implementiert, und Grad-CAM wird in allen gängigen Deep-Learning-Ökosystemen unterstützt. Diese Kombination aus methodischer Vielfalt und Reife der Werkzeuge macht die vier Methoden zu einem repräsentativen und praktisch relevanten Teil des breiteren XAI-Landschafts für die Zwecke dieser Studie.

Abbildung 1.5 zeigt ein Beispiel von Objekten im Verkehrssektor und das Ergebnis gemäß dem Zielkonzept.

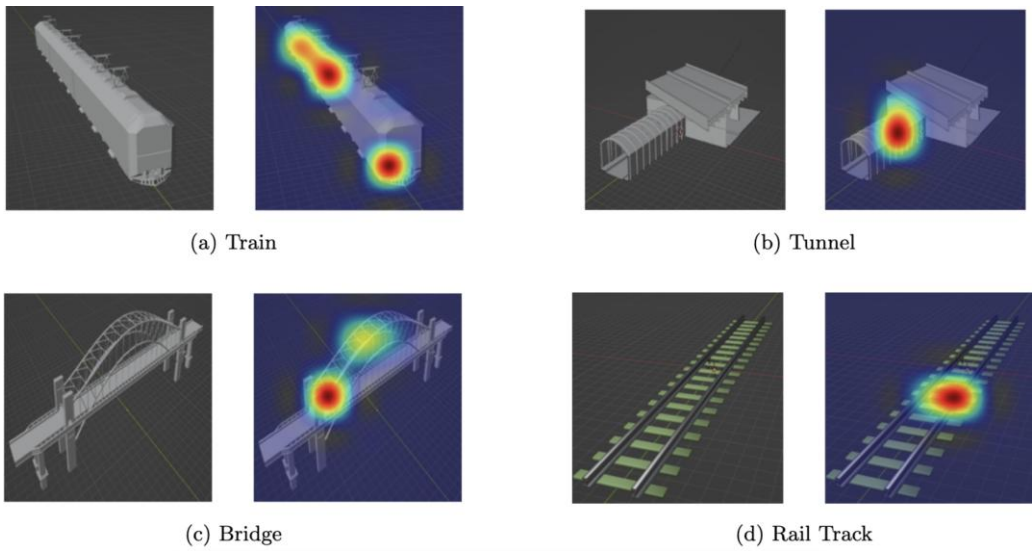


Abb 1.5. Beispiel für die Anwendung von Grad-CAM in der Studie.

1.3.5 Entwicklung einer Metrik zur Bewertung der Interpretierbarkeit und Erklärbarkeit der ML-Modelle

Um die Erklärbarkeit objektiv zu bewerten, wurde eine auf dem Analytic Hierarchy Process (AHP) basierende Metrik entwickelt. Elf Indikatoren über drei Dimensionen wurden von Experten gewichtet. Tabelle 1.4 zeigt die Dimensionen und die elf Indikatoren. Sie wurden auf Basis der Literaturrecherche ausgewählt, wobei zwei aus der eigenen Taxonomieerstellung des Autors stammen.

Tabelle 1.4. Ausgewählte Dimensionen und Kriterien.

Dimension	Kriterien	Beschreibung
Erklärmethoden	Translucency	Drückt aus, wie tief ein Erklärungsansatz das Modell untersucht.
	Complexity	Die rechnerische Komplexität des Erklärungsalgorithmus.
	Completeness	Das Ausmaß, in dem ein zugrunde liegendes Schlussfolgersystem durch die Erklärung beschrieben wird.
Einzelne Erklärungen	Faithfulness	Dem Rat des Systems zu vertrauen, auch wenn der Benutzer nicht mit Sicherheit weiß, dass er korrekt ist.
	Consistency	Das Ausmaß, in dem verschiedene Modelle, die dasselbe Problem gelernt haben, ähnliche Erklärungen liefern.
	Accuracy	Eine Erklärung einer spezifischen Entscheidung auf zuvor unbekannte Situationen zu verallgemeinern.
	Versatility	Das Ausmaß, in dem das Modell für Stakeholder und Branchen anwendbar ist.
Benutzerfreundlichkeit	Comprehensibility	Die Lesbarkeit und Länge der Erklärungen.
	Fairness	Ob die Vorhersagen keinerlei implizite oder explizite Voreingenommenheit gegenüber den Zielnutzern enthalten.
	Reliability	Sicherstellen, dass geringfügige Eingabeänderungen keinen signifikanten Einfluss auf die Modellvorhersage haben.
	Accessibility	Wie einfach es für den Benutzer ist, die relevanten Konzepte zu erlernen und anzuwenden.

Tabelle 1.5 zeigt das Gewicht und den Rang jedes Kriteriums, wie vom AHP-OS bestimmt. Abbildung 3 zeigt das Radardiagramm der AHP-Gewichte. Das Ergebnis ist eine Aggregation der individuellen Eingaben von sieben Teilnehmern.

Tabelle 1.5. Ausgewählte Dimensionen und Kriterien

Dimension	Kriterien	Gewicht	Rank
Erklärmethoden	Translucency	0.066	9
	Complexity	0.033	11
	Completeness	0.072	7
Einzelne Erklärungen	Faithfulness	0.046	10
	Consistency	0.118	3
	Accuracy	0.202	1
	Versatility	0.069	8
Benutzerfreundlichkeit	Comprehensibility	0.100	4
	Fairness	0.083	6
	Reliability	0.128	2
	Accessibility	0.084	5

Die am höchsten gewichteten Indikatoren in dieser Studie: Genauigkeit (0,202), Zuverlässigkeit (0,128) und Konsistenz (0,118), spiegeln eine Verschiebung hin zu operativen Zuverlässigkeitsmetriken wider, anstelle rein subjektiver oder theoretischer Konstrukte. Diese Ausrichtung auf stakeholderzentrierte Entscheidungsfindung ist bemerkenswert, da sie darauf hindeutet, dass Praktiker die Robustheit und Wiederholbarkeit von Modellen höher priorisieren als andere Dimensionen, die in früheren Frameworks betont wurden. Zum Beispiel erhielt die Komplexität in dieser Studie das niedrigste Gewicht (0,033), im Gegensatz zu Frameworks, die Einfachheit als wesentlich für das Verständnis durch den Nutzer hervorheben. Dies deutet darauf hin, dass Einfachheit zwar wünschenswert sein mag, Stakeholder jedoch bereit sind, komplexere Erklärungen zu akzeptieren, wenn das zugrunde liegende Modell genauer und zuverlässiger ist – ein praxisrelevanter Kompromiss, der in unterschiedlichen Branchenkontexten weiter untersucht werden sollte.

Die Ergebnisse können aus verschiedenen Perspektiven interpretiert werden. Die Teilnehmenden betrachteten die Genauigkeit als am wichtigsten mit einem Wert von 0,202, gefolgt von Zuverlässigkeit (0,128) und Konsistenz (0,118), die zusammen 44,8 % des gesamten Indikatorgewichts ausmachen. Diese Dominanz operativer Robustheitsmetriken spiegelt mehrere kontextuelle Faktoren in entscheidungskritischen Umgebungen wider. Erstens erfordern Verkehrssysteme, in denen ML-Modelle operative Entscheidungen informieren (z. B. Verspätungsvorhersagen), dass Stakeholder-Verantwortung und regulatorische Vorgaben zuverlässige, wiederholbare Erklärungen sicherstellen, die in operativen und rechtlichen Überprüfungen standhalten. Zweitens zeigen empirische Befunde zur organisatorischen Einführung von KI-Systemen, dass Entscheidungsträger bei hohen Einsätzen Vertrauenswürdigkeit und Konsistenz über andere Dimensionen priorisieren –

Stakeholder müssen sicher sein, dass kleine Änderungen der Eingabedaten nicht zu radikal unterschiedlichen Modellentscheidungen führen. Drittens korreliert die Genauigkeit von Erklärungen direkt mit der Modelltreue: Wenn die Erklärungen einer XAI-Methode die tatsächliche Entscheidungslogik des Modells falsch darstellen, wird selbst eine einfache und intuitive Erklärung kontraproduktiv. Im Gegensatz dazu erhielt die Komplexität das niedrigste Gewicht (0,033), was darauf hindeutet, dass Stakeholder bei der Wahl zwischen einer komplexen, aber genauen Erklärung und einer einfachen, möglicherweise irreführenden Erklärung die erstere bevorzugen, ein Befund, der praktische Implikationen für die Gestaltung von XAI-Tools in hochverantwortlichen Bereichen hat.

1.3.6 Verlinkung der ML-Modelle mit Knowledge-Graph

ML-Modelle wurden in die Knowledge-Graphen integriert, indem Eingabevariablen und Vorhersageergebnisse als Ontologie-Instanzen modelliert wurden. Diese ermöglicht automatisierte Abfragen, und das System ruft die beitragenden Faktoren direkt aus der verknüpften Graphstruktur ab (siehe Abbildung 4, Anhang 3.5.3).

1.3.7 Validierung der Integration der ML-Algorithmen und „Explainer“

Die vorhandenen und im Zuge der Projektverlängerung zu ergänzenden oder veränderten ML-Algorithmen und Explainer sind evaluiert und dokumentiert, sowie in die Suche beziehungsweise in das Gesamtportal integriert.

1.4 Ziel 4: Entwicklung einer intelligenten Suche zur Entdeckung der Assets von DZ, um die Skalierbarkeit und Wiederverwendbarkeit der Assets von DZ zu verbessern.

Dieses Ziel wird durch AP4 – die Intelligente Suche zur Entdeckung der Assets– erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP4 dargestellt.

1.4.1 Standardisierte Struktur zur Formulierung der Anwendungsszenarien

Der methodische Rahmen zur Formulierung von Anwendungsszenarien im Talenta-Knowledge-Graphensystem etabliert eine standardisierte Pipeline, die Reproduzierbarkeit und Transparenz bei der Konstruktion von Knowledge-Graphen sicherstellt. Der in Abbildung 1.6 dargestellte Rahmen baut auf vorausgehenden Arbeitspaketen auf und nutzt insbesondere Spezifikationen aus dem Dokument zur Szenario-Daten-Ontologie-Abbildung, das in den Arbeitspaketen 1 bis 3 erarbeitet wurde. Diese systematische Grundlage ist essenziell, da sie eine semantische Ausrichtung zwischen Anwendungsszenarien und den zugrunde liegenden Datenstrukturen gewährleistet.

Der Workflow schreitet in einer methodischen Identifikationsphase voran, in der abhängige und unabhängige Variablen aus den Szenariospezifikationen extrahiert werden. Dieser Prozess ist grundlegend für die nachfolgenden Phasen, da er unmittelbar die Auswahl der

Machine-Learning-Modelle sowie die Qualität des Feature Engineerings bestimmt. Durch die explizite Definition der Variablenbeziehungen in dieser frühen Phase stellt der Rahmen sicher, dass die ausgewählten Algorithmen auf einer theoretisch fundierten und empirisch begründeten Variablenstruktur basieren.

Nach der Variablenidentifikation (abhängige und unabhängige Variablen) integriert der Rahmen ein Evaluations- und Auswahlprotokoll für Machine-Learning-Algorithmen, eine Phase, die sowohl für die Vorhersageleistung als auch für die Interpretierbarkeit der Modelle von zentraler Bedeutung ist. Der Rahmen veranschaulicht diesen Prozess anhand einer konkreten Instanziierung, etwa unter Verwendung von GTFS-Daten. Die praktische Anwendung mit GTFS-Daten demonstriert die operative Umsetzbarkeit und Effektivität, wobei der resultierende Knowledge-Graph, wie in Abbildung 3 (Anhang 3.5.3) dargestellt, Beziehungen, Entitäten und Inferenzstrukturen kodiert, die aus dem GTFS-Datensatz abgeleitet sind. Insgesamt etabliert dieser Rahmen einen standardisierten Ansatz zur Konstruktion von Knowledge-Graphen.

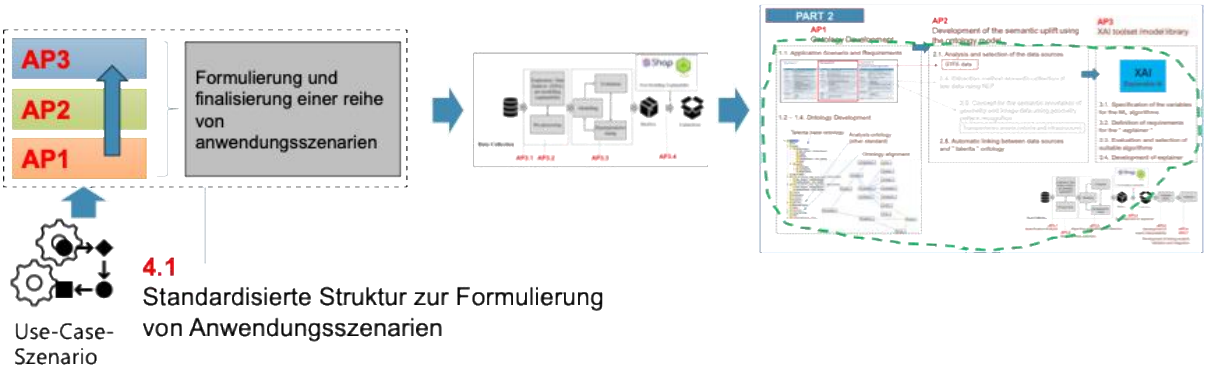


Abb 1.6. Standardisierter Rahmen zur Formulierung von Anwendungsszenarien im TALENTA-Knowledge-Graphen

1.4.2 Entwurf und prototypische Implementierung der Eingabemaske zur Szenariendefinition und Suche.

Alle für die Prozesse nötigen Masken sind definiert, gestaltet und im Verlauf der Projektverlängerungen mit den verbesserten Strukturen ergänzt, so dass Assets über die Masken eingestellt und im semantischen Uplift verarbeitet werden können. Ebenso können neu eingestellte Assets über die Suchfunktionen gefunden werden. Zusätzlich können relevante ergänzende Informationen angezeigt, sowie mit diesen verknüpfte, passende Assets/ Asset-Gruppen, aber auch passende ganze digital-Twins, angezeigt werden.

1.4.3 Übersetzungsprogramms, das die definierte Such- und Anwendungsszenarieneingabe auf die (Abfragesprache des Knowledge-Graphen, z.B. SPARQL, Cypher, übersetzt.

Das Programm übersetzt die natürliche, „menschliche“ Sprache dahingehend, dass die spezifischen „technischen“ Abfragen auf dem Knowledge-Graphen durchgeführt und die Ergebnisse wieder in „menschliche“ Sprache rückübersetzt werden können.

1.4.4 Methode zur Organisierung und Darstellung der Abfrageergebnisse

Zur Darstellung der Abfrageergebnisse wurde, je nach Szenario, eine auf den jeweiligen Merkmalen basierende Darstellung entwickelt. Sobald neue Szenarien angelegt und gepflegt werden, passen sich Organisationsstruktur und die daraus resultierenden Masken automatisch an.

1.4.5 Datenschnittstelle des neuen Such- und Entdeckungsmechanismus zum in AP5 entwickelnden Portal

Der Datenaustausch zwischen Portal, der Suche und den Backend-Applikationen wurde so entwickelt, dass neue Assets auf den Graphen geschickt und über den Suchmechanismus und seine grafische Darstellung im Portal abgerufen und dargestellt werden.

1.5 Ziel 5: Entwicklung eines Asset-Management-Portals, das sichere Verwendung und Austausch von DA sowie intuitive Interaktion erlaubt

Dieses Ziel wird durch AP5 – die Interaktions- und Visualisierungsansätze für das „talenta“-Portal – erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP5 dargestellt.

1.5.1 Entwurf einer graphischen Benutzeroberfläche des „talenta“-Portals

Für alle Ebenen des Portals sowie der bewerbenden Marketing-Webseite wurde ein grafisches Interface entwickelt und auf die verschiedenen _Usecases und die damit verknüpften Rollen und Rechte abgestimmt. Das im Verlauf der Projektverlängerung entwickelte weitere Szenario wurde abgeglichen und notwendige Anpassungen an den verschiedensten Oberflächen angebracht. Darüber hinaus wurden ein Markenname und ein Claim entwickelt und abgestimmt. Darauf basierend sind mehrere Designs prototypischen entwickelt, abgestimmt und dann beispielhaft ausgearbeitet und auf das Portal übertragen worden. (siehe Anhang).

1.5.2 Visualisierung und Interaktion der XAI-Ergebnisse.

Es wurde ein neuartiges Konzept zur Visualisierung und Interaktion der XAI-Ergebnisse entwickelt. Dies ist notwendig, um die Erklärungen der ML-Modelle für die Endnutzer im Portal verständlich darstellen zu können.

1.5.3 Visualisierungs- und Interaktionsmethode von DA anhand der abgebildeten Verlinkung zwischen Geometrie-/Bilddaten und Objekten

Eine Methode zur Anzeige von Vorschau- und Metainformationen ist realisiert und eine Verknüpfung mit Echtdaten wurde prototypisch umgesetzt. Ein Objekt, zum Beispiel eine Brücke, kann gefunden werden. Alle Metadaten zur Brücke werden angezeigt, ebenso die auf dem Knowledge-Graph darüber hinaus direkt verknüpften Daten, die mit dem Objekt in Verbindung stehen. Ebenso werden die Metadaten zu möglicherweise relevanten weiteren Objekten mit Bezug (ähnlichen Brücken, in der Nähe befindliche Brücken, aus gleichem Material oder zur gleichen Zeit erbaute Objekte) aber auch Objekte, die einen „entfernteren“ Bezug haben (Bahnverbindungen, die über diese Brücke führen, Schienensysteme, die bei solchen Brücken-Typen verwendet werden, Informationen über bekannte Fehlfunktionen, die bei ähnlichen Objekte in der Vergangenheit auftraten, etc.) werden im Kontext der Objektdarstellung mit angezeigt.

1.6 Ziel 6: Entwicklung KMU-gerechter innovativer Geschäftsmodelle durch die Verwendung der “talenta“-Plattform

Dieses Ziel wird durch AP6 – Prototypische „talenta“-Plattform – erreicht. In den folgenden Abschnitten werden die Ergebnisse und Leistungen der Teilarbeitspakete innerhalb von AP6 dargestellt.

1.6.1 Authentifizierung- und Autorisierungsmechanismen

Sowohl die einzelnen API Endpunkte, sowie die Talenta plattform an sich verfügen über rudimentäre Authentifizierungs- und Autorisierungsmechanismen.

Diese gelten zum jetzigen Zeitpunkt nicht plattformweit.

Die Arbeiten konnten innerhalb der Projektlaufzeit nicht vollständig abgeschlossen werden und bilden Ansatzpunkte für weiterführende Entwicklungen.

1.6.2 Integrationstests innerhalb des Gesamtsystems

Die Komponenten wurden teilweisen Integrationstests unterzogen

Da kein integriertes Gesamtsystem erstellt werden konnte waren auch keine Gesamt-Integrationstest möglich

Die Arbeiten konnten innerhalb der Projektlaufzeit nicht vollständig abgeschlossen werden und bilden Ansatzpunkte für weiterführende Entwicklungen.

1.6.3 Charakterisierung der entwickelten Methoden zur Datenerhebung und Optimierung von Datenübertragungs- und -verarbeitungslaufzeiten

Die einzelnen Komponenten wurden simplen Laufzeit Tests unterzogen, wobei darauf hinzuweisen ist, dass diese direkt auf Entwickler Maschinen ausgeführt wurden. Aufgrund dessen sind die erhaltenen Ergebnisse als Qualitativ und Statistisch bedenklich einzustufen. Nichts destotrotz wurden Optimierungen auf Basis der Testergebnisse getroffen.

1.6.4 Modifizierung der entwickelten ML-Modelle

Auf Grund von zeitlichem Mangel bei der Bearbeitung des Gesamtprojekts kam es zu keiner abschließenden modifizierung der ML-Modelle.

Die Arbeiten konnten innerhalb der Projektlaufzeit nicht vollständig abgeschlossen werden und bilden Ansatzpunkte für weiterführende Entwicklungen.

1.6.5 Optimierung des Such- und Entdeckungsmechanismus

Auf Grund von zeitlichem Mangel bei der Bearbeitung des Gesamtprojekts kam es zu keiner Finalen Optimierung des Such- und Entdeckungsmechanismus.

Die Arbeiten konnten innerhalb der Projektlaufzeit nicht vollständig abgeschlossen werden und bilden Ansatzpunkte für weiterführende Entwicklungen.

1.6.6 Finale Optimierung des Gesamtsystems

Auf Grund von zeitlichem Mangel bei der Bearbeitung des Gesamtprojekts kam es zu keiner Finalen Optimierung des Gesamtsystems. Das System wurde aufgrund von Laufzeit-Tests optimiert, dies ist eher Komponentenspezifisch passiert.

Die Arbeiten konnten innerhalb der Projektlaufzeit nicht vollständig abgeschlossen werden und bilden Ansatzpunkte für weiterführende Entwicklungen.

1.6.7 neuartige Geschäftsmodelle (DTaaS und MLaaS)

Für die Arbeiten aus AP 6.7 wurde ein detailliertes Verwertungskonzept erstellt, das die im FuE-Vorhaben entwickelten Ansätze für DTaaS- und MLaaS-Geschäftsmodelle aufgreift, konkretisiert und strukturiert weiterführt. Damit werden die Geschäftsmodelle erstmals vollständig ausgearbeitet, wirtschaftlich bewertet und in ein skalierbares Gesamtgeschäftsmodell überführt.

Siehe Anlage: “Verwertungskonzept zur Wirtschaftlichen Nutzung der talenta Plattform” (3.5.3.2)

1.6.8 Nachhaltigkeits- und Transferkonzepten

Das erstellte Verwertungskonzept operationalisiert die begonnenen Transfer- und Nachhaltigkeitsansätze systematisch. Es definiert einen klaren Phasenplan, organisatorische Strukturen und wirtschaftliche Rahmenbedingungen, um die Projektergebnisse in einen nachhaltigen, marktfähigen Betrieb zu überführen.

2. der wichtigsten Positionen des zahlenmäßigen Nachweises

Für die Entwicklung sind keine ungeplanten kosten Angefallen. Da bis auf ein SAN-System, keine zusätzliche Infrastruktur (Server, Entwickler-Maschinen, etc.) eingekauft werden musste. Das SAN-System hat die geplanten Kosten sogar unterschritten. Bis auf das mFund Netzwerktreffen in Aachen im Oktober 2023 wurden alle Meetings der Projektpartner online abgehalten. Lediglich der Partner CUB hat die Reisekosten durch die Vortragsreisen vollkommen ausgeschöpft, dadurch wurden auch die Reisekosten des Projektes unterschritten. Der Großteil der Aufwendungen des Gesamtvorhabens ist demnach in Form von Personalkosten angefallen.

3. der Notwendigkeit und Angemessenheit der geleisteten Arbeit

des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans, explizit auch im Hinblick auf die Daten- und Dienstebereitstellung in der Mobilithek. Stellen Sie deutlich dar, welche Daten Sie in die Mobilithek geladen haben und wie diese einen Nutzen für Dritte darstellen können. Gehen Sie bei Nichtveröffentlichung von erhobenen bzw. veredelten Daten ausführlich auf die Gründe der Nichtveröffentlichung ein.

Im Verkehrssektor ist die Implementierung digitaler Zwillinge Teil der Digitalisierungsmaßnahme zur Verbesserung der Ressourceneffizienz im Management der Infrastrukturen. Jedoch ist der Einsatz von digitalen Zwillingen aufgrund von Herausforderungen wie dem fehlenden gemeinsamen Verständnis von digitalen Zwillingsmodellen, schwieriger Modellintegration, Sicherheitsproblemen, fehlendem Zugriff auf wichtige Daten und hohen Kosten aufgrund ineffizienter Geschäftsmodelle immer noch begrenzt. Die Arbeiten an talenta zielten darauf ab, diese Herausforderungen mit Hilfe neuer Technologien wie z.B. xAI in Angriff zu nehmen und dabei eine benutzerfreundliche Management-Platform zu schaffen. Eine Veröffentlichung veredelter Projekt- oder Betriebsdaten in der Mobilithek erfolgte nicht, da es sich überwiegend um synthetische Testdaten, proprietäre Datenverarbeitungsprozesse sowie geschützte Softwareartefakte handelt. Die entwickelten Konzepte API’s und Ontologien sind jedoch grundsätzlich wiederverwendbar.

4. des während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Im Laufe des Projektes sind leider keine Neuerungen in den Bereichen des Digitalen-Assets Managements beziehungsweise der Ontologien bekannt geworden.

5. der erfolgten oder geplanten Veröffentlichungen des Ergebnisses

5.1 Erfolgte Veröffentlichungen

Fekete, T., Mengistu, G., & Wicaksono, H. (2025). Leveraging causal AI to uncover the dynamics in sustainable urban transport: A bike sharing time-series study. *Sustainable Cities and Society*, *122*, 106240. https://doi.org/10.1016/j.scs.2025.106240

Vijaya, A., Bhattarai, S., Angreani, L. S., & Wicaksono, H. (2024). Enhancing transparency in public transportation delay predictions with SHAP and LIME. *2024 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 1285–1289. https://doi.org/10.1109/IEEM62345.2024.10857000

Vijaya, A., Gudissa, B. L., Angreani, L. S., & Wicaksono, H. (2024). Predictive analysis of public transportation delays using machine learning models on GTFS data. *2024 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 450–454. https://doi.org/10.1109/IEEM62345.2024.10857111

Wicaksono, H., Nisa, M. U., & Vijaya, A. (2023). Towards intelligent and trustable digital twin asset management platform for transportation infrastructure management using knowledge graph and explainable artificial intelligence (XAI). *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 0528–0532. https://doi.org/10.1109/IEEM58616.2023.10406401

Pilz, M. (2023): Die Entwicklung der KI-gestützten Asset-Management-Plattform „talenta“ wird als Verbundprojekt im Rahmen der Innovationsinitiative mFUND mit insgesamt 643.240 Euro durch das Bundesministerium für Digitales und Verkehr (BMDV) gefördert. Vectorsoft Blog. https://www.vectorsoft.de/blog/2023/09/forschungsprojekt-talenta-gestartet-ki%e2%80%91asset%e2%80%91management-fuer-digitale-zwillinge/

5.2 Geplante Veröffentlichungen

An Analytic Hierarchy Process (AHP)-Based Metric for Explainable AI (XA) Evaluation: Empirical Evidence for Data-Driven Decision Making, *International Journal of Information Management Data Insights*

A Semantic and Explainable AI-Driven Platform for Collaborative Digital Twin Management in Transportation Infrastructure, *Automation in Construction*

5.3 Anhang

5.3.1 Abbildungen

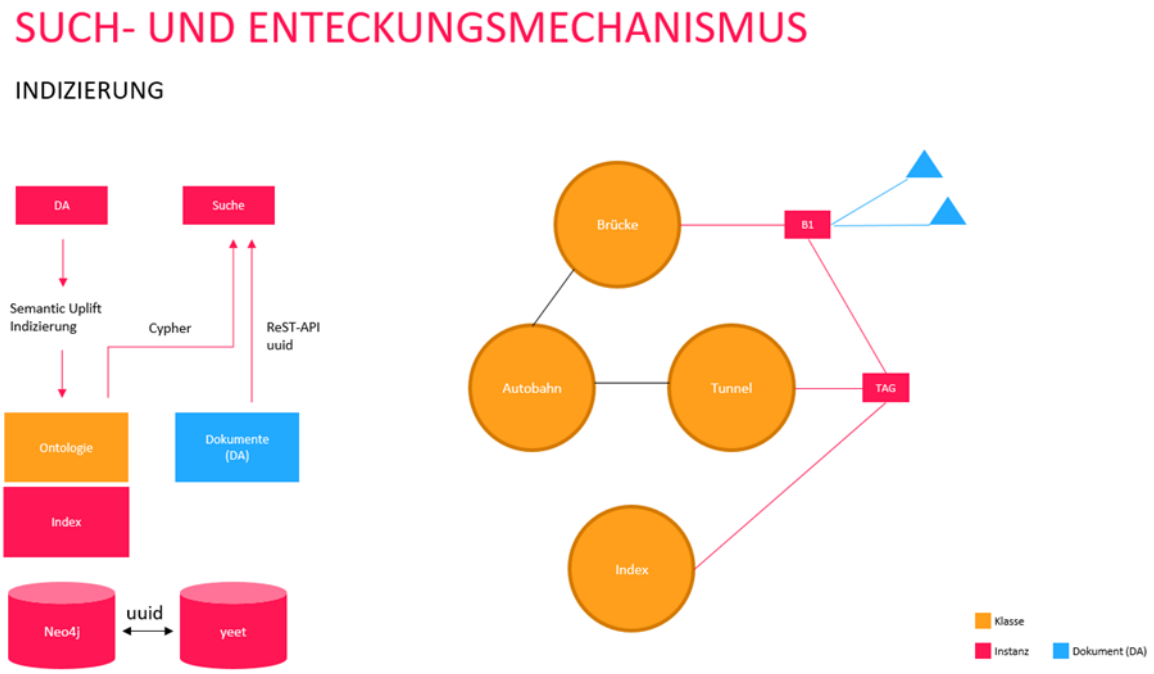


Abb. 1 Interaktion der verschiedenen Komponenten welche das Suchen ermöglicht

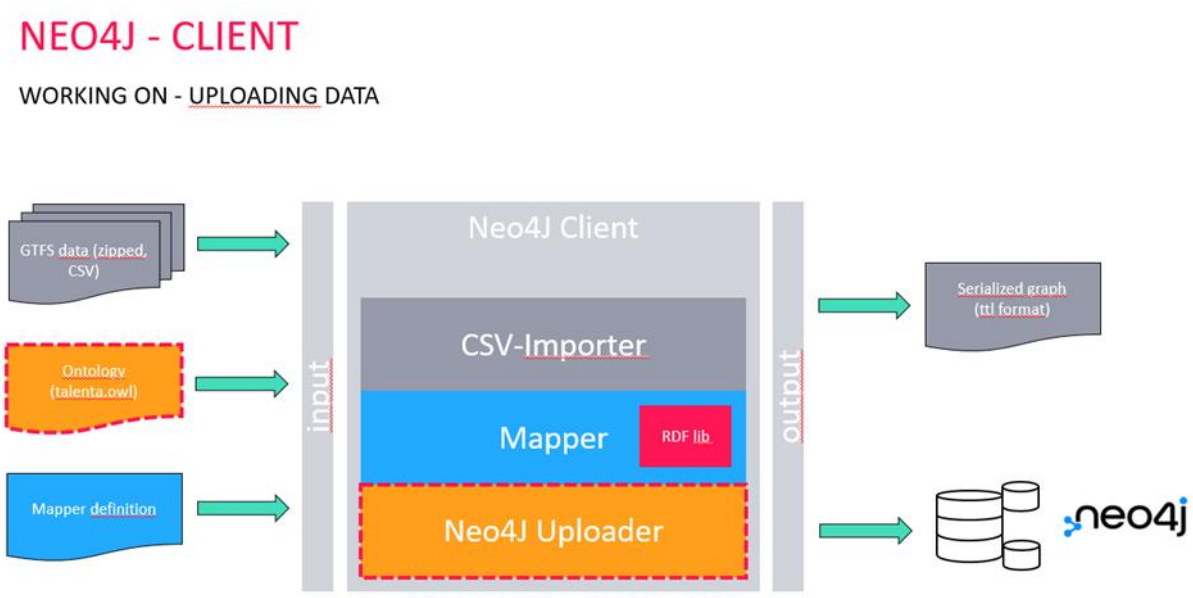


Abb. 2 Überblick über die Komponenten welche beim Uploadvorgang benötigt werden

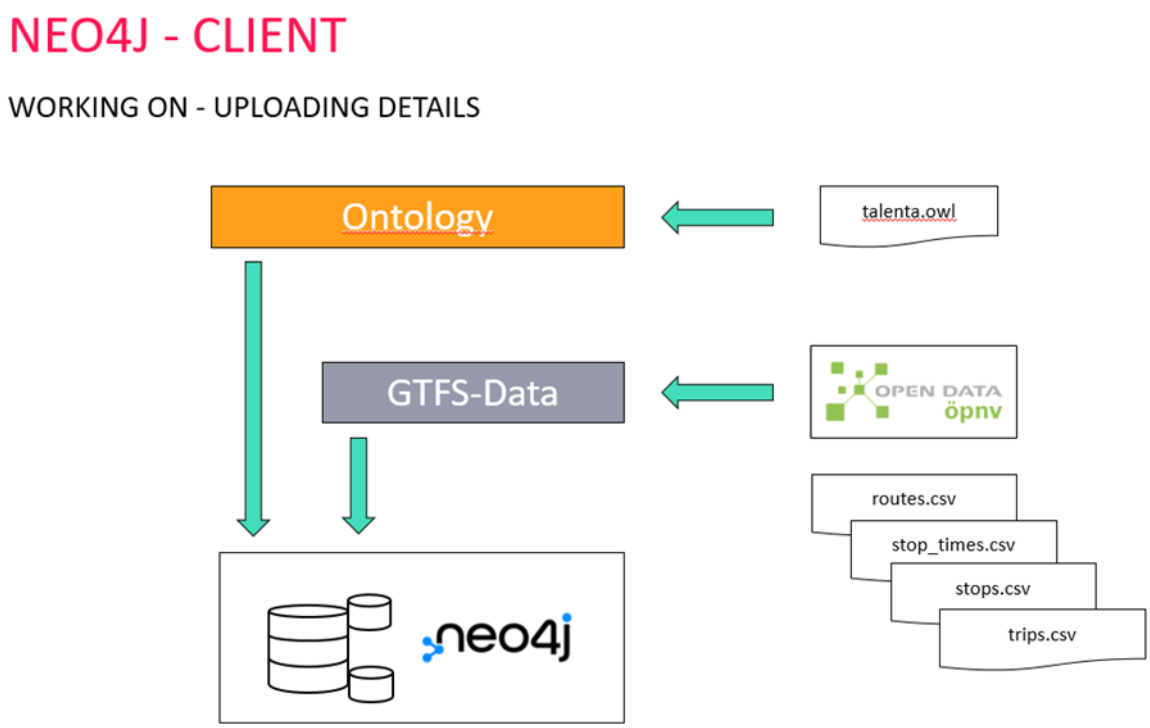


Abb. 3 Überblick über die unterschiedliche Verarbeitung von GTFS und GBFS Daten

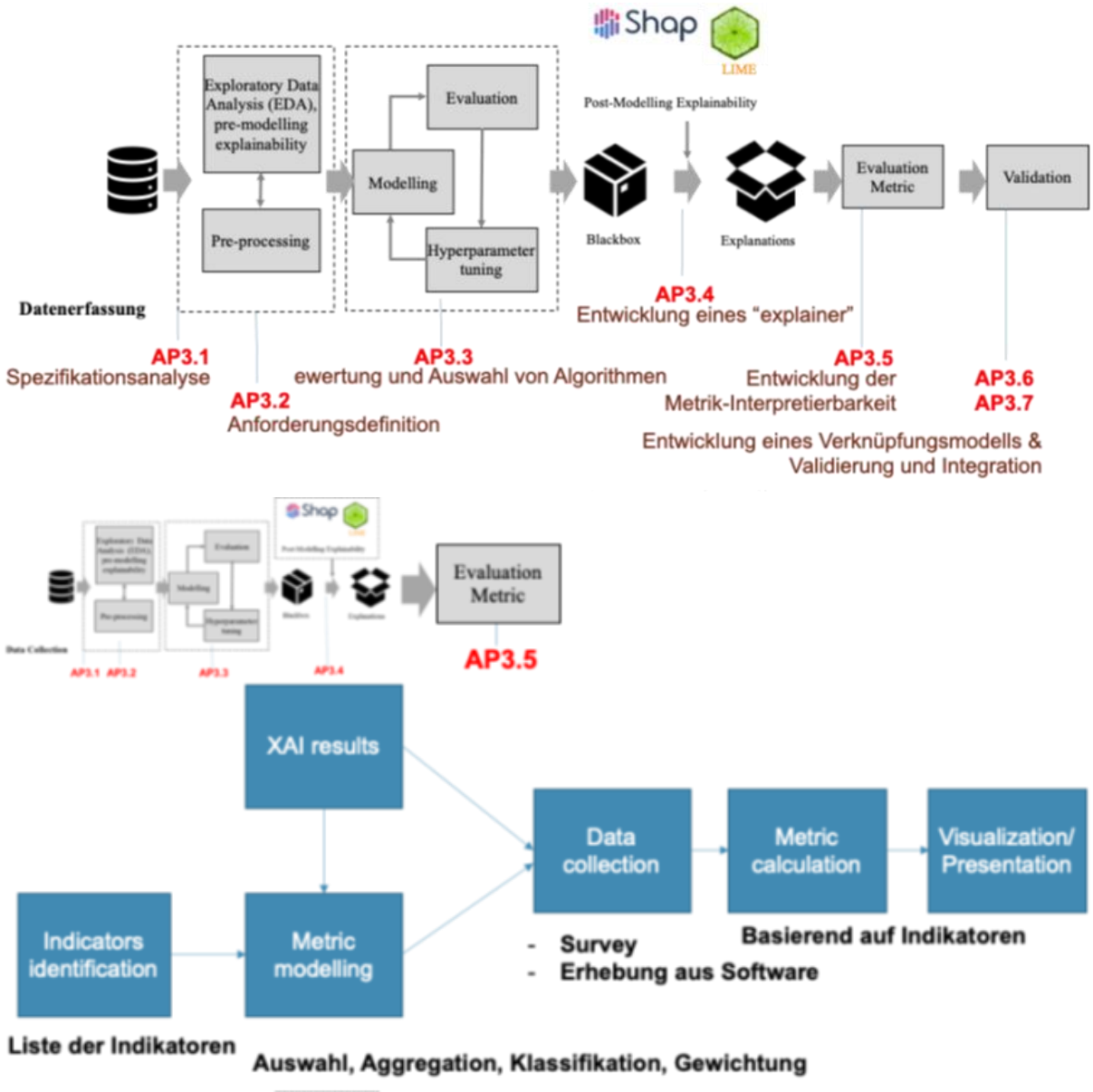


Abb. 4 Benötigte Komponenten bei der Datenerfassung, sowie Semantic-Uplift jener. Sowie die Zuordnung der Arbeitspaket Einteilung.

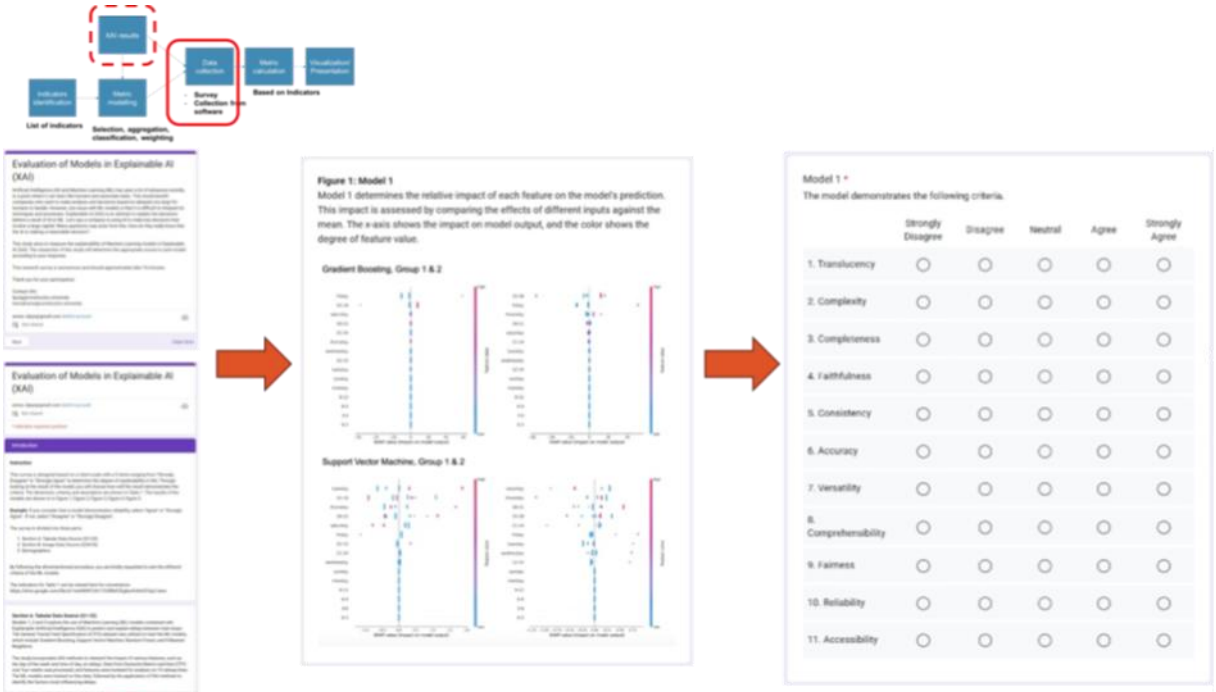


Abb. 5 Erklärung / Beispiel, wie die Ki-Ergebnisse mithilfe von XAI erklärt werden, sowie die möglichkeit das System durch User-Bewertung für die Zukunft Dynamisch zu verbessern

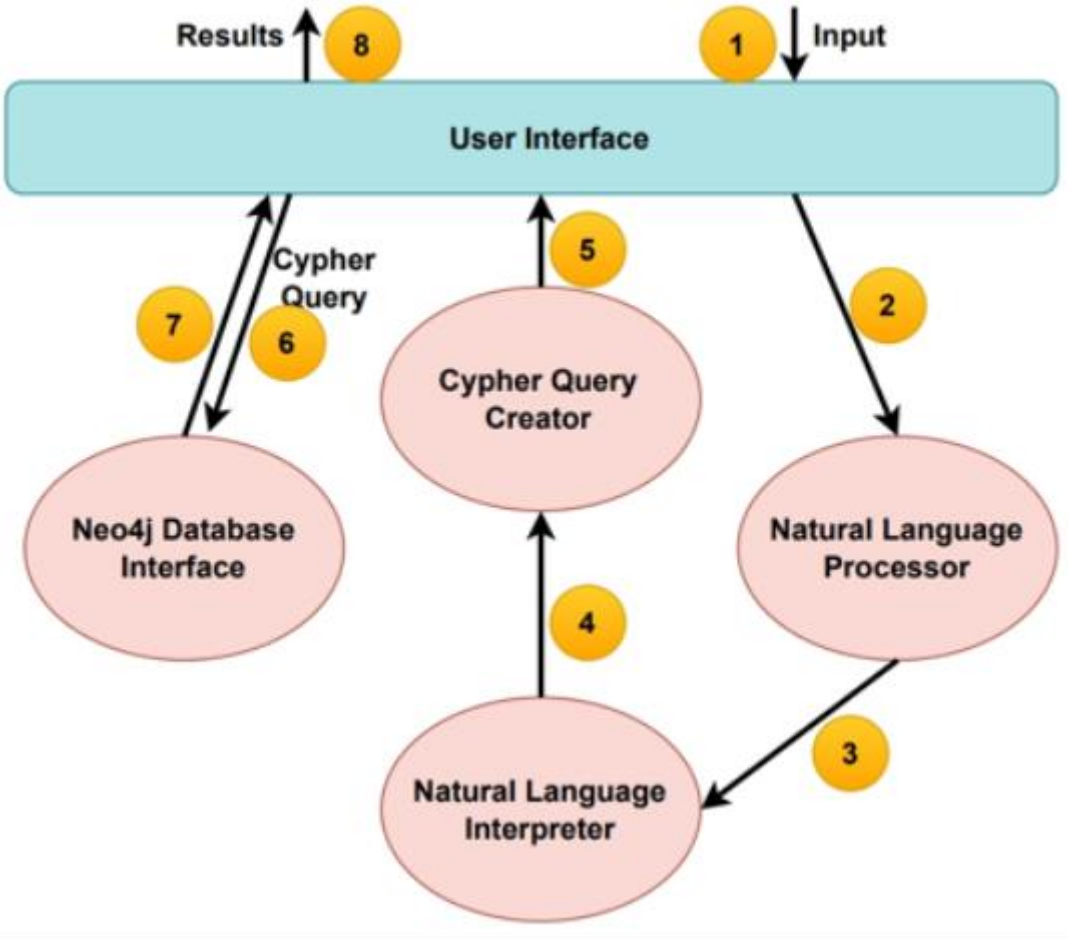


Abb. 6 Initiales Schaubild über die Komponenten und Informationsfließrichtung, damit eine Intelligente Suche über die Daten ermöglicht wird

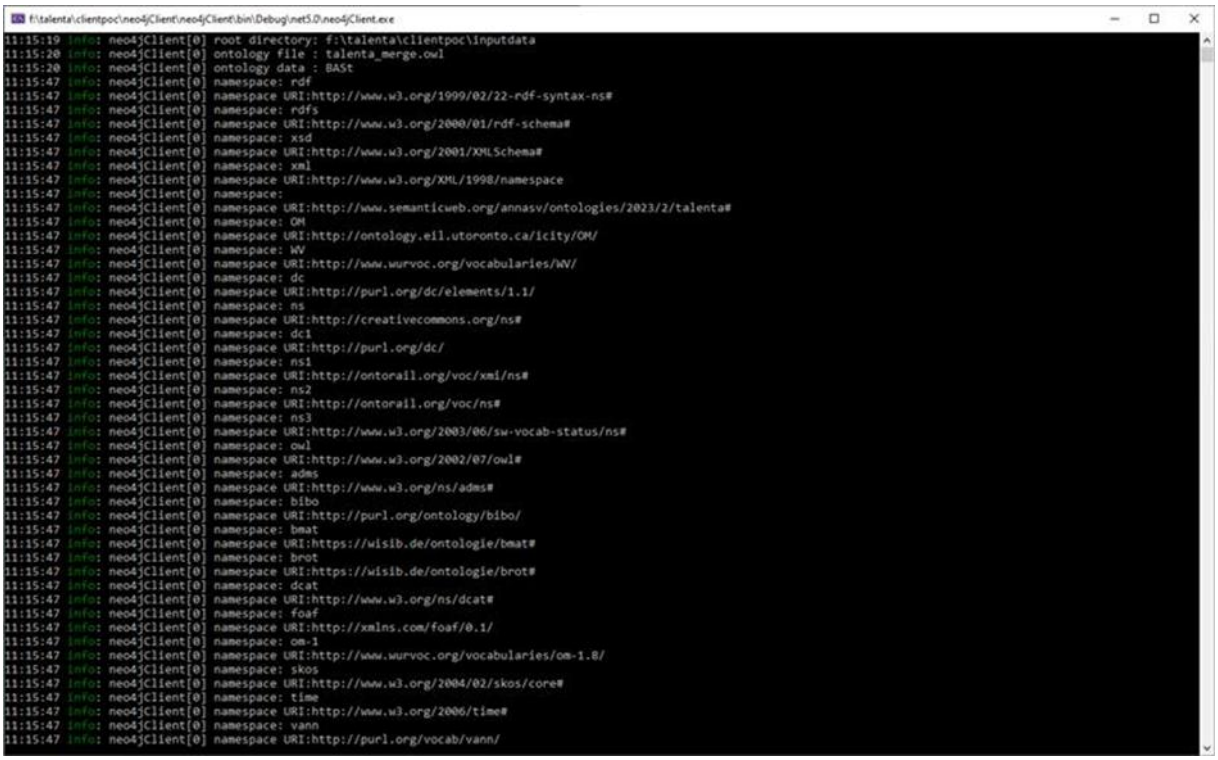


Abb. 7 Interne Verarbeitung von Brückendaten (BASt) innerhalb der Talenta Ontologie.

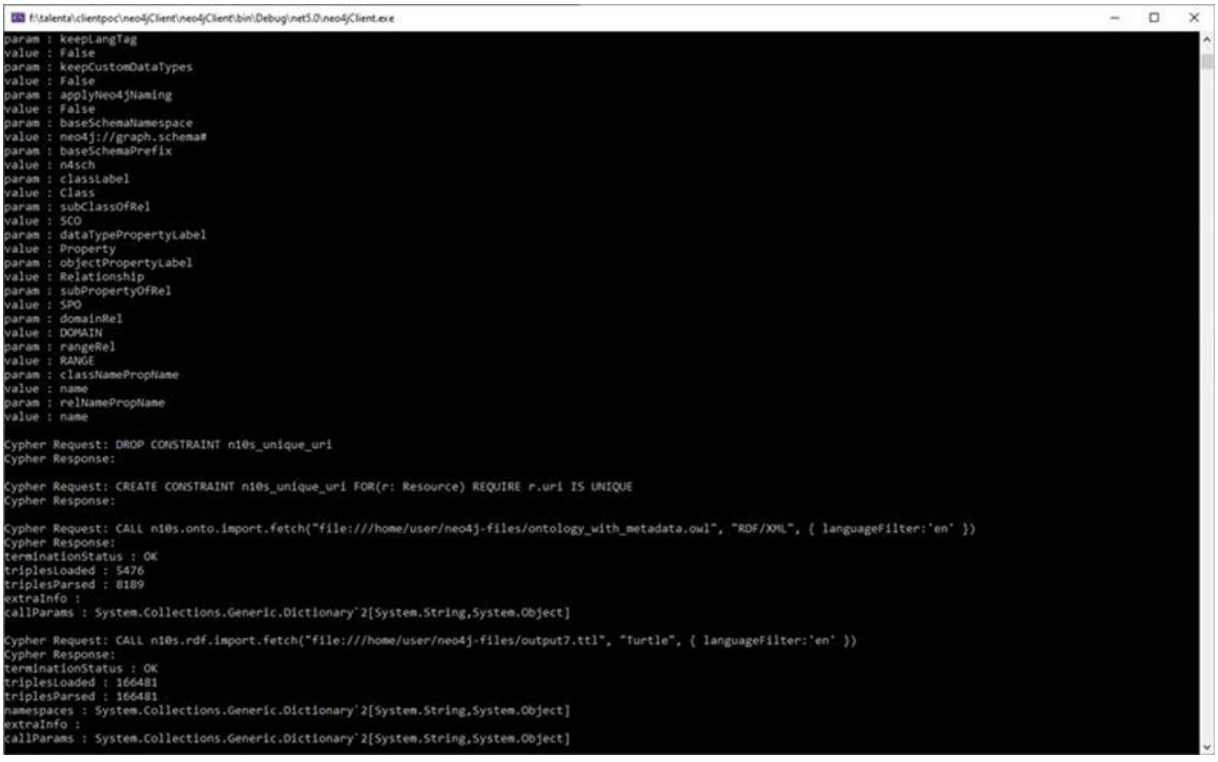


Abb. 8 Ausschnitt der Cypher-Requests welche übersetzt und auf dem vectorsoft Neo4J-Server durchgeführt werden

Rolling Stock (Perspektive)

Infrastruktur (Fixed Assets)

Übertragbar auf weitere Verkehrsträger

Verkehrsträger	Transportmittel	Verkehrsweg	Verkehrsinfrastruktur
Straßen	PKW, LKW	Straßennetz	Straßenbrücken, Straßen-tunnel, Parkplätze
Schienen	Eisenbahnen	Schienennetz	Bahnhöfe, Eisenbahn-brücken, Eisenbahntunnel
Luftraum	Flugzeuge	Luftstraßen	Flughäfen
Wasserwege	Schiffe, Fähren	Flüsse, Kanäle, Seewege	Häfen, Schleusen
Rohrleitungen	Pipelines, Versorgungsleitungen	Leitungsnetze	Speicher, Pumpen
Seile	Kabinen, Sessel, Schienenfahrzeuge (?)	Luftseilbahnen, Sesselifte, Standseilbahnen	Seilbahnstationen, Seilstützen
Weltraum	Raketen, Satelliten	Flugbahn, Umlaufbahn	Weltraumbahnhof

Abb. 9 Übertragbarkeit nach Laufzeitende

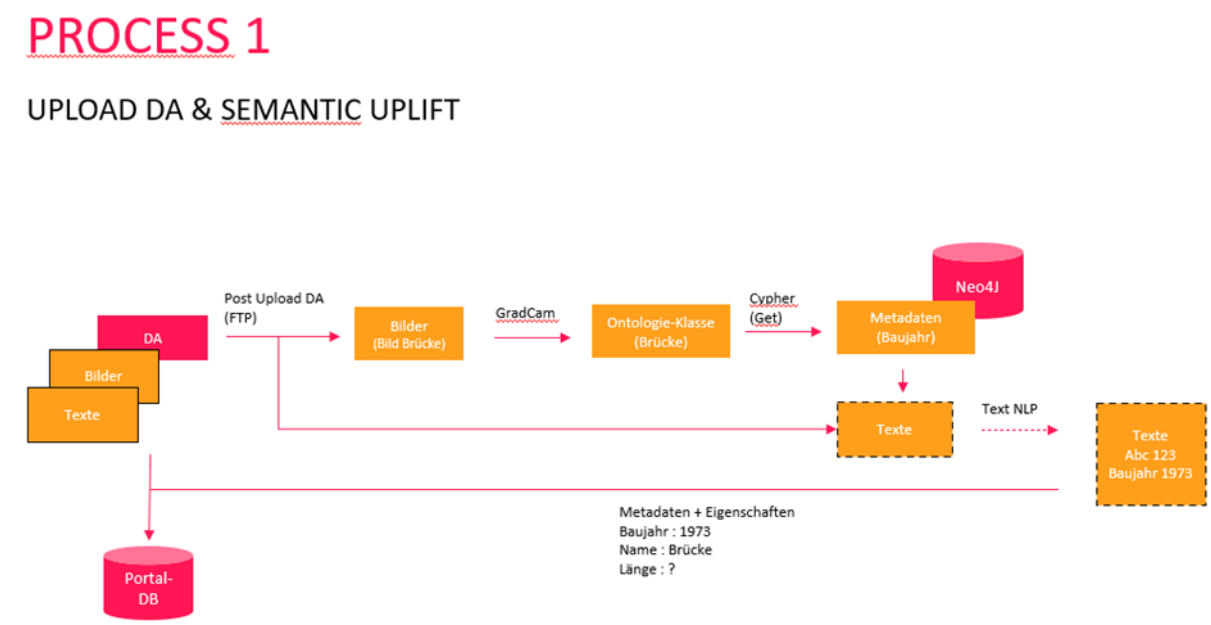


Abb. 10 Detail Prozessbeschreibung zur Integration neuer Daten beim Uploadvorgang

PROCESS 2

CREATE DA

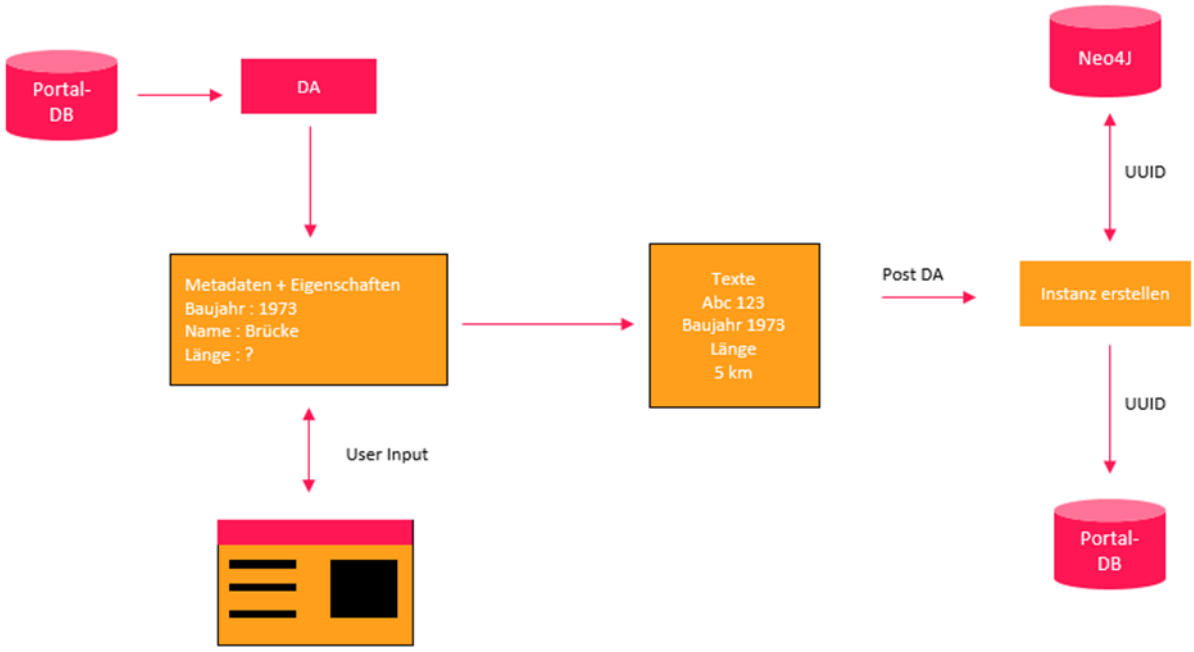


Abb. 11 Detail Prozessbeschreibung beim erstellen neuer Daten über User-Eingabe Dialog in der Talenta Web Anwendung

PROCESS 3

FIND DA

1 Input: Text + Liste Eigenschaften
Output: Eigenschaftenswerte

2 Input: NL-Text
Output: Cipher-Abfrage

Abb. 12 Detail Prozessbeschreibung für die Suche mithilfe von NLP innerhalb der Talenta Platform

5.3.2 Anlage “Darstellungen zur Entwicklung des Prototyps”

Dieses Dokument ist als externe Anlage beigefügt.

5.3.3 Anlange “Verwertungskonzept zur Wirtschaftlichen Nutzung der talenta Plattform”

Dieses Dokument ist als externe Anlage beigefügt.