A Reference Use Case, Data Space Architecture, and Prototype for Smart Truck Parking

Jean Paul Sebastian Piest [1], Stefani Slavova [2] and Wouter van Heeswijk [1]

[1] University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands
[2] Delft University of Technology, Mekelweg 5, Delft, 2628 CD, The Netherlands

Abstract

The logistics sector experiences an increasing shortage of secure truck parkings, in part due to lacking insight into occupancy rates. Extending earlier work – in which a variety of machine learning approaches are evaluated to predict truck parking occupancy – this paper presents a reference use case, data space architecture, and prototype for smart truck parking. The research builds upon case study research and 1.5 years' worth of real-world data of a truck parking in Deventer, the Netherlands. Deploying the conceptual design from prior work, the main contributions of this paper are: (1) a reference use case, (2) a data space architecture, and (3) a prototype for smart truck parking. The reference use case is developed following the use case methodology of Fraunhofer. The architecture is rooted in the international data spaces reference architecture model, modelled using the ArchiMate language and specification, and validated by means of expert opinion. The cross industry standard process for data mining is utilized to develop and deploy a prototype, based on a single case mechanism experiment, for a smart truck parking application. Future research can focus on evaluating and developing ontologies to extend the current conceptual representation of the data objects and the experimental development of federative machine learning approaches among truck parkings.

Keywords

Reference Use Case, Data Space Architecture, Smart Truck Parking, International Data Spaces, IDS, Machine Learning, Prototype.

1. Introduction

Earlier work revealed the shortage of 400.000 secure trucks parkings in Europe and evaluated a variety of machine learning approaches to predict the occupancy of truck parkings [1]. Reviewing literature showed that most existing works focus on car parking occupancy prediction. The case study at A1 Truck Parking in Deventer, the Netherlands, leveraged the literature regarding car parking occupancy prediction to provide insight into predictive features for truck parking occupancy prediction. The CRoss Industry Standard Process for Data Mining (CRISP-DM) was utilized to examine the applicability of machine learning algorithms, using 1.5 years of historical data to evaluate various sets of predictive features. Additionally, an architecture was proposed for a dynamic prediction tool, which can be used by planners, truck drivers, truck parking managers, and road authorities to improve truck parking utilization. This paper builds upon this earlier work.

Deploying machine learning projects involves several challenges, which are not always documented, during each stage of development that are valuable to assess how difficult it is to realize applications [2]. A recent review regarding Internet of Things and machine learning in Intelligent Transportation Systems (ITS) indicated a lack of machine learning coverage in the application area of smart parking and calls for contributions [3]. Truck parking differentiates from car parking, because (1) there are more stakeholders involved, (2) locations are geographically dispersed, and (3) truck drivers must comply

[1]Proceedings of the EEWC 2022, 02-11-2022, Leusden, the Netherlands
EMAIL: j.p.s.piest@utwente.nl (J.P.S. Piest); w.j.a.vanheeswijk@utwente.nl (W. van Heeswijk);
ORCID: 0000-0002-0995-6813 (J.P.S. Piest); 0000-0002-5413-9660 (W. van Heeswijk);

with EU regulations. The specific challenges of the case study at A1 Truck Parking and lack of smart truck parking applications are addressed in this paper.

In this paper, which is an extension of earlier work [1], the conceptual design from earlier work is used to develop a reference use case and create a data space architecture for smart truck parking using the International Data Spaces (IDS) Reference Architecture Model (RAM) [4] and ArchiMate language [5]. Additionally, the CRISP-DM is utilized to develop and deploy a prototype for a smart truck parking application based on the case study at A1 Truck Parking and 1.5 years of real-world data.

The main contributions of this paper are: (1) a reference use case, (2) a data space architecture, and (3) a prototype for a smart truck parking. The reference use case and data space architecture for smart truck parking contribute to the scientific knowledge base regarding the use of machine learning in ITS and present a novel use case for data sharing based on the IDS RAM. The prototype contributes to the validation of the ideas and machine learning models from earlier work. It provides inspiration for scholars that are interested in smart parking applications as well as guidance for system developers to construct data spaces using the IDS RAM.

The remainder of this paper is structured as follows. **Section 2** introduces the use case methodology and reference use case for smart truck parking. **Section 3** presents the data space architecture and describes its main components and interfaces. **Section 4** summarizes the case study, using the steps of CRISP-DM, and the development of the prototype. **Section 5** concludes and discusses future work.

2. Reference use case

This section describes the use case methodology that was used to create a reference use case for smart truck parking.

2.1. Use case methodology

The use case methodology of Fraunhofer was adopted to develop a reference use case for smart truck parking, as shown in **Figure 1**.

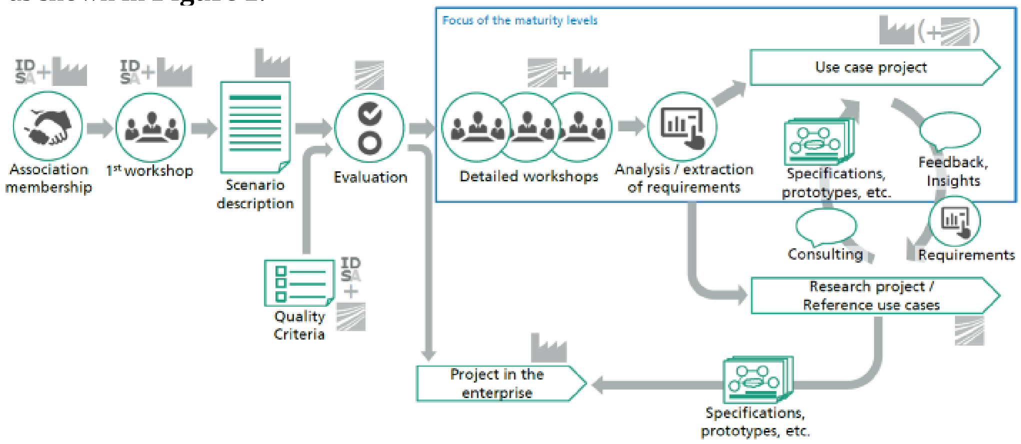

Figure 1: Use case methodology (image source: Fraunhofer [6]).

Reference use cases are abstractions and simplifications of real-life use cases in enterprises and are suitable for research, feasibility studies, and experimental development [6]. Reference use cases are developed in a controlled environment, but can be used for the implementation of industry testbeds as use case project or enterprise use cases in real-life environments. The use case methodology defines activities to achieve maturity levels: (1) introduction workshop, (2) functional workshop with key decision makers, (3) collaborative workshop with involved stakeholders, (4) technical workshop with developers, (5) formulation of reference use case and requirements, (6) project planning, (7) realization of prototypes (pilots) and/or software, (8) validation, and (9) productive operation (if desired). The use case methodology is supported by the scenario description template and quality control checklist.

The University of Twente developed this reference use case together with the Provincie Overijssel with the support of TNO as regional IDS ecosystem partner and involvement of experts. For the purpose of this study a reference use case is created to (1) assess the feasibility of a data space approach, (2) evaluate the use of the IDS RAM, and (3) demonstrate the main functionality by means of a prototype application. The main deliverables are the architecture specification and prototype.

2.2. Smart truck parking

Earlier work revealed a shortage of 400.000 secure truck parkings in the EU [1]. More specific, there is a lack of visibility concerning truck parking occupancy, causing a suboptimal spatiotemporal utilization that aggravates the shortage

Road transportation companies must comply with EU regulations regarding driving and rest times. Therefore, trucks are equipped with a tachograph. The EU regulations constrains the planning of required stops for individual routes to specific areas. Transport planners can be tasked to find a secure truck parking, but this can also be delegated to the truck driver. Furthermore, traffic dynamics and congestion can affect the planning and may require dynamic rerouting to alternative truck parkings.

Truck parkings are owned and managed by private stakeholders that offer facilities and services. Some truck parkings require a subscription and offer secure parkings while others are open and free of charge. Currently, the involved parties do not (extensively) share or exchange data. Important transport lanes experience capacity problems while other truck parkings experience underutilization. The government is not a direct stakeholder (like it can be in urban car parking), nor are truck parkings a responsibility of road infrastructure authorities. Despite EU regulations in place, the parking situation differs per country.

The aforementioned scenarios result, among other effects, in exceeding driving time limits, illegal truck parking, unsafe driving, and subpar working conditions. This reference use case for smart truck parking aims to establish a trusted data space and evaluate the IDS RAM to (1) establish an ecosystem of (linked) data to address the lack of visibility, (2) conceptualize smart services to optimize planning for trucking companies and increase utilization degrees for truck parking owners, and (3) demonstrate its use by means of prototypes to involved stakeholders.

The use case aims require establishing trust among parties to share data, combining data from several sources (e.g., transport management systems, tollgate systems, open (linked) data), integration of different data types (e.g., static location data, planned routes and stops, real-time traffic data), supporting infrastructure, and governance.

3. Data space architecture

This section presents the data space architecture for smart truck parking, describes the use of the IDS RAM, and summarizes the main feedback from expert opinion interviews.

3.1. Layered viewpoint

This subsection first provides an overview of the layered viewpoint and then subsequently describes its three layers.

3.1.1. Overview and objective

Building on the conceptual design from our earlier work, the ArchiMate language and specification and IDS RAM are utilized to model a layered architecture viewpoint for the trusted data space and smart truck parking application. The overall objective is to create a trusted data space in which truck parking owners can share the occupancy of truck parkings using the trust security profile of the IDS RAM to facilitate transport planners to optimize their planning and dynamically reroute to nearby truck parkings while ensuring data sovereignty. **Figure 2** depicts the layered architecture viewpoint for the trusted data space and smart truck parking application. The layered architecture viewpoint consists of a business layer, application layer, and technology layer. In the remainder of this section each layer will be described.

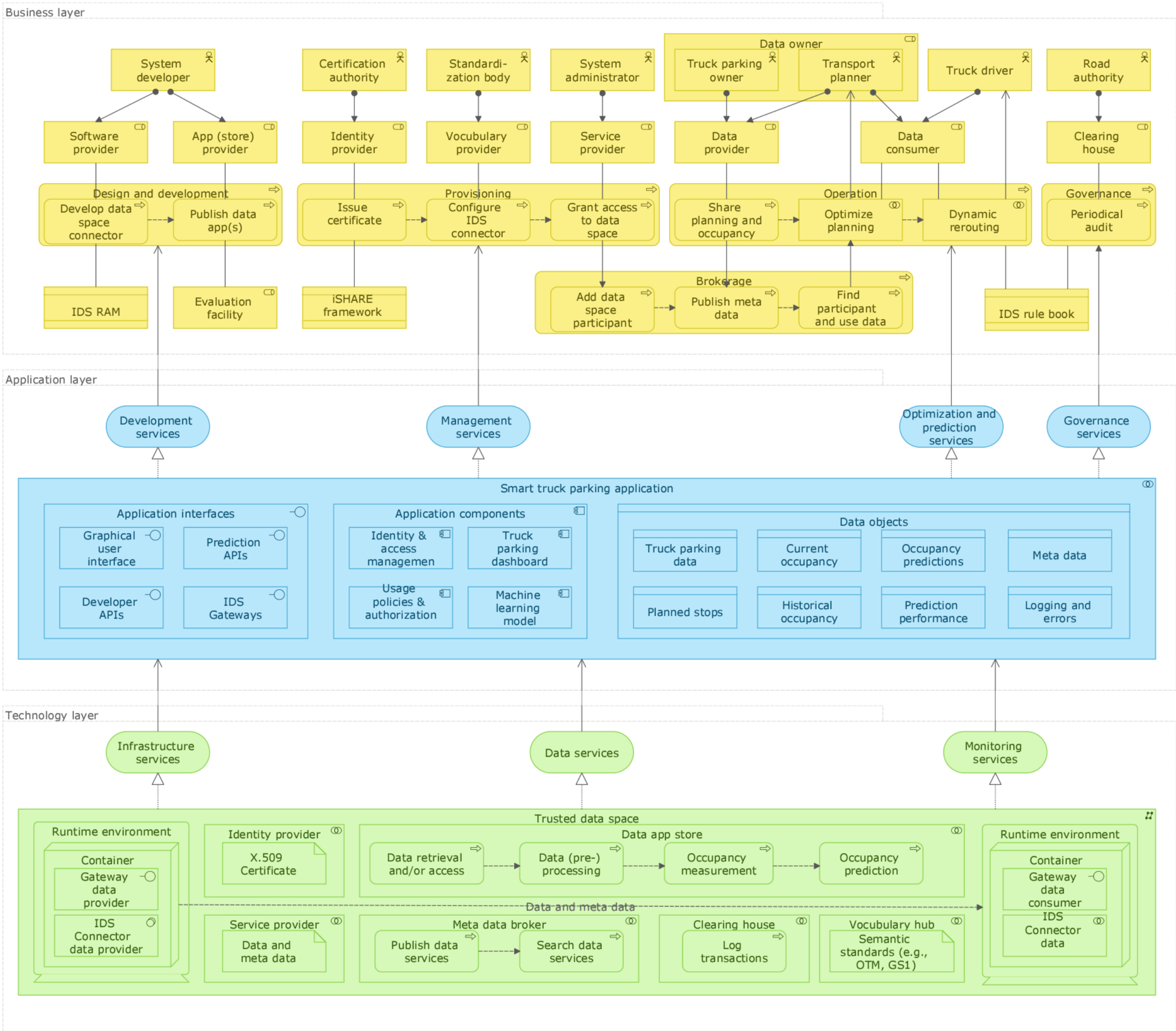

Figure 2: Layered architecture viewpoint for the data space architecture for smart truck parking.

3.1.2. Business layer

The business layer presents the actors, roles, processes, and supporting contracts. Following the stakeholder analysis, the main actors are assigned to the roles to design, develop, provision, operate, and govern the trusted data space and smart truck parking application. All processes are decentralized, and participants must be(come) IDS member. The system developer fulfills the role of software provider and app (store) provider. Using the IDS RAM, the system developer can establish the data space and publish the data app(s). The supporting role of the evaluation facilities is responsible for the certification process (further discussed in **Section 3.3**). Upon certification of the software developer, established data space, and data app(s), the provisioning process can start. Data space participants are assumed to follow the standard onboarding process. A certification authority acts as an identity provider and issues the X.509 certificates for participants. Here, the iSHARE trust framework can be utilized, providing a common approach to Identification, Authentication, and Authorization (IAA). Next, the standardization body provides semantic standards (e.g., the Open Trip Model (OTM), GS1 standards) as a vocabulary provider to support configuration of the IDS connector. The system administrator can subsequently grant participants access to the data space and act as service provider, including decentralized meta data brokerage. This process can be supported by an interoperability contract. The truck parking owner and transport planner are both data owner and data provider. The truck parking owner provides static data regarding the truck parking and dynamic data containing the current occupancy. The transport planner provides insight in planned stops. Based on the shared planning, the transport planner can act as data consumer to optimize the planning. The truck driver acts as data consumer for dynamically rerouting.

The IDS rule book provides guidance to operate and govern the data space. Here, the road authority could act as clearing house and periodically audit the data space to govern trust.

3.1.3. Application layer

The application layer presents the smart truck parking application. The application layer is a technology collaboration between application interfaces, components, and data objects to provide application services to support the processes in the business layer. The smart truck parking application has separate interfaces for end users and connected systems. End users are provided a graphical user interface to access the application components. Separate APIs are provided for developers and connected systems for predictions. The communication with IDS participants takes place via IDS gateways. The main components of the smart truck parking application are the truck parking dashboard and machine learning model. Following the trust security profile, the IAA is decentralized. The smart truck parking application provides remote access to shared data objects. The meta data is published via the meta data broker. Centered around the IAA and usage policies, participants can search data providers, start sharing data and utilize truck parking and planning data.

3.1.4. Technology layer

The technology layer presents the trusted data space and provides infrastructure, data, and monitoring services for the smart truck parking application. The main component in the technology layer is the IDS connector. Each participant needs to configure and install the IDS connector with a valid X.509 certificate. The IDS connector is defined as system software as it is provided based on the IDS RAM. The data provider publishes meta data to the meta data broker. The data consumer can search for participants in the meta data broker. The data app needs to be installed in both connectors. Specific vocabularies are shared via the vocabulary hub. The data provider and data consumer utilize the IDS connector to share data and meta data. Transactions are logged in the clearing house.

3.2. Use of IDS

This subsection describes how the five layers of the IDS RAM are incorporated in the trusted data space and smart truck parking application.

3.2.1. Requirements

Data sovereignty is the core of the IDS RAM and defined as the capability of a natural person or organizational entity to be entirely self-determined with regard to its data [4]. Data sovereignty is about finding the balance between benefits realization and protecting data as strategic asset. This is reflected in six strategic requirements: (1) trust, (2) security and data sovereignty, (3) ecosystem of data, (4) standard interoperability, (5) value adding apps, and (6) data markets. To foster trust, each component is evaluated and participants are certified before being granted access to the data space. Security profiles are shaped according to state-of-the-art security measures and data sovereignty is implemented for data ownership based on usage policies and active monitoring. An ecosystem of data is rooted in decentralized data sharing and using broker services to search participants and data sources. The IDS connector facilitates standard interoperability and provides the infrastructure for value adding apps, including data processing and analytics services. The data market is implemented in the form of domain specific brokerage and a clearing house to log transactions.

3.2.2. Business layer

The business layer specifies the main roles data space participants may assume, their categories, activities, and interactions with other roles. The security profiles and roles function as a blueprint for the underlying functional and technical layers. The IDS RAM describes four categories of roles: (1) core participants, (2) intermediary, (3) software / service provider, and (4) governance body. Core participants are organizations that own, provide, and/or use data and include the following roles: data owner, data provider, data consumer, data user, and app provider. The system developer, truck parking owner, transport planner, and truck driver can be categorized as the core participants of the trusted data space and smart truck parking application. Intermediates are organizations that act as trusted entities in a data space and include the following roles: broker service provider, clearing house, identity provider, app store provider, and vocabulary provider. The road authority and standardization body can be categorized as intermediates. Software and service providers are organizations that provide software and the required technical infrastructure to establish a data space and data space services. The governing body takes care of the certification process and issuing certificates. This role is fulfilled by the International Data Space Association (IDSA) and selected evaluation facilities. The data owner is the main stakeholder which creates data, including meta data, sets the usage policies to make data accessible, and determines the business model (for third parties). The data owner and data provider can be the same organization, but the data provider can also be a software- or service provider. Authorizations are documented in a legal contract, enforced by technical usage contracts, and set(s) of usage policies and restrictions following the IDS RAM. The IDS RAM prescribes mandatory and optional interaction between the roles. Based on the security profile trust, most roles are mandatory and all communication is decentralized. The roles required to design and develop the trusted data space are: software provider, app provider, and evaluation facility. The roles that are required to provision the trusted data space are: identity provider, vocabulary provider, and service provider. The roles that are required to operate the trusted data space are: data owner, data provider, (broker) service provider, and data consumer. The clearing house role is required for the governance of the trusted data space. Digital identities play a pivotal role to ensure trust in data spaces. Static trust is ensured via certification and use of the IDS components. Dynamic trust is governed based on active monitoring and use of IDS components.

3.2.3. Functional layer

The functional layer defines the functional requirements and concrete features. **Table 1** shows how the functional requirements are satisfied.

Table 1

Requirements satisfaction.

Requirements	Features
Trust	Roles, identity management, user certification
Security and data sovereignty	Authentication and authorization, usage policies and enforcement, trustworthy communication, security by design, technical certification
Ecosystem of data	Data source description, brokering, vocabularies
Standardized interoperability	Operation, data exchange
Value adding apps	Data app implementation, data processing and transformation, installing and supporting data apps
Data markets	Clearing, usage restrictions, governance

Trust is implemented in the roles, IAA, and certification. Security and sovereignty are translated in usage policies and restrictions. The X.509 certificate ensures static security. Trustworthy communication measures are in place, including encryption and security domain checks. Usage enforcement is attached to outbound data. The IDS RAM contributes to security by design by means of security profiles and the security architecture that prescribes the isolation of data apps for separation of concerns and remote execution guarantee. The technical certification of the IDS connector and components contributes to a trusted platform. The ecosystem of data is established via data source description, including the syntax and semantics of both the data source and metadata. Meta data brokering realizes an interface for data and meta data search, access, and use for registered data space participants. Furthermore, existing and domain specific vocabularies are provided. Standardized interoperability is realized by running connectors in separate environments, user management, and logging of transactions. The connector handles the data exchange with data sources and enterprise system integration. Value adding apps are implemented as a data app using the subtype for data processing and transformation. The software provider and/or service provider publish the data app in an app store for authorized users, provide access for system administrators via strong 2-factor authentication, and offer support for installing data apps and lifecycle management (e.g., updates, removal). A data market is created for logging of transactions in a clearing house. This clearance encompasses data ownership, data sovereignty, data quality, data provenance, and legal aspects.

3.2.4. Process layer

The process layer specifies the interactions and provides a dynamic view. The IDS RAM contains standard processes, described using the business process management notation, for onboarding, exchange of data, publishing and using data apps. The process layer is incorporated in the business layer. Publishing data apps is captured in the interaction between the app provider and evaluation facility. Onboarding incorporates both legal and technical aspects to obtain a digital identity and issue the X.509 certificate. The connector configuration and provisioning include security setup, availability setup, and granting data consumers access to data sources. The exchange of data enables transport planners to search data providers (e.g., truck parking owners) in the broker service and start sharing data by invoking data operations. Subsequently, the data apps are used to optimize planning and for dynamically rerouting. This requires a specialized data app for data transformation, data analytics, and using the machine learning model for occupancy predictions.

3.2.5. Information layer

The information layer provides a conceptual model, utilizing linked data principles, to describe the static and dynamic aspects of data sharing. More specifically, digital resources are described, published, and made exchangeable while preserving data sovereignty. The IDS RAM describes three levels of model representation: conceptual, declarative, and programmatic. The conceptual representation contains the main concepts and is descriptive. The declarative representation provides a normative view built on semantic web technology standards and standard modelling vocabularies. This includes formal, machine-interpretable specification of concepts. The programmatic representation targets software providers. The information model is domain-agnostic. Domain modelling is delegated to vocabularies. More specifically, the information model provides vocabularies using the RDF ontology, data app interfaces for interaction between technical components and governing intended usage of data app based on its security level and licensing model. The information model, shown in **Figure 3**, is based on a concerned hexagon and used to describe each digital resource and meta data.

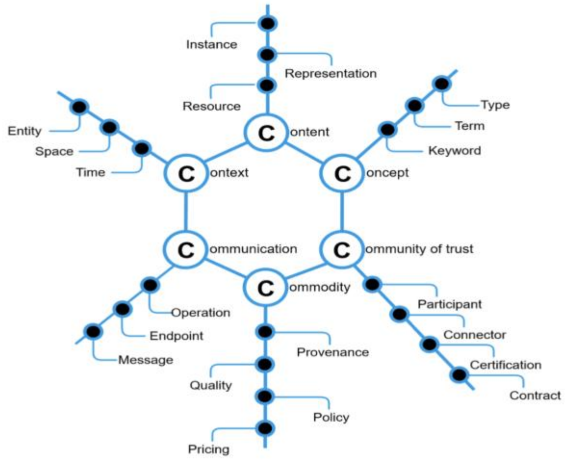

Figure 3: IDS information model, image from [4].

The upper part of the concern hexagon contains descriptive aspects and the lower part the data sharing aspects. Content is the most important aspect, as it relates to the digital resource. The representation of content is described in terms of structure, format, schema, mapping, and an instance is provided. A digital resource can be data, text, audio, image, video, or other formats. In this case, truck parking data is the main digital resource. Digital resources can also be containers (e.g., an archive with historical occupancy) or software (e.g., machine learning model). Content is linked to shared, formally defined, concepts and a specific context in time, place, and real-world entities. In this case, the content is linked to physical truck parkings and organizations such as A1 Truck Parking. Communication covers the aspects related to data sharing and data exchange. Here, the communication protocols (e.g., HTTPS, MQTT, IDSCP) and interface definitions are specified in terms of endpoints, operations, and messages. The OTM or GS1 standards can be utilized to standardize the data exchange regarding locations. The smart data model for parking can possibly be used for data sharing [7]. Commodity is concerned with provenance, quality, policies, and pricing. Provenance means that all content related activities can be traced back to participants. This also involves quality dimensions (e.g., measurement, meta data, certificates) and policies (e.g., rules, permissions, obligations, constraints, prohibitions). Additionally, the pricing model is described. A chargeable service fee might stimulate truck parking owners to join the trusted data space. The community of trust provides data regarding the participant, connector, certification, and usage contract. Here, participants are registered and the deployment context of the connector is described in term of security profile, service catalog and host. The community of trust is linked to an evaluation facility, certification level, and usage contracts for the digital resource.

3.2.6. System layer

The system layer is concerned with the logical software components and their decomposition, including integration, configuration, deployment, and extensibility of these components. In the application layer, the smart parking application is specified as a technology collaboration of interfaces, components, and data objects. The IDS connector architecture provides standardized interoperability based on the IDS RAM and separates the configuration from execution. The runtime consists of a configuration manager, configuration model, validator, and specific configurations (e.g., network, workflows). All execution is containerized and orchestrated by an execution core container. This core container has a data router and data bus. Each data app is deployed in a separate container. A data app can be self-developed, provided by a third party, or a supporting data app (e.g., adapters, smart data apps for processing, transformation, storage). Data apps are made available via the app store, a secure platform for distributing data apps, and published in a registry with available data apps. The app store provides additional features for provisioning of data apps to connectors and are complemented by additional services such as billing or support.

3.3. Perspectives of the IDS

This subsection discusses the three perspectives that need to be implemented across the five layers of the IDS RAM.

3.3.1. Security

The security perspective is essential to fulfill the requirements and uses existing standards and best practices. The IDS RAM contains seven security concepts: (1) secure communication, (2) identity management, (3) trust management, (4) trusted platform, (5) data access control, (6) data usage control, (7) data provenance tracking. Security levels are determined by the profiles and definition of roles and interaction patterns in the business layer. On the functional layer, the roles are further detailed in data usage control via policies and contracts. The process layer provides a dynamic view on security via certificates, IAA, use of tokens, and active monitoring. Information security is based on the information model and vocabularies. Security is technically realized in the IDS connector. The IDS RAM provides a participant and resource-based differentiation with its security profiles. The decentralized setup can be extended to central clearing house logging for provenance storage and policy management.

3.3.2. Certification

The certification perspective can be subdivided in participant and component certification. The IDS RAM prescribes a certification process for the business layer. Participants must be(come) member of the IDS to join a trusted data space. The compliance of the implementation to the IDS RAM and use of technical components (e.g., connector, clearing house) is evaluated in the functional layer. The compliance to IDS RAM in terms of support of the processes is part of the process layer. The compatibility with the information model is evaluated in the information layer. On the system layer, specific connector implementations and interactions between the components are evaluated. Component certification is based on a standard process.

3.3.3. Governance

The governance perspective defines the duties and rights regarding data management. The business layer facilitates development of new business processes considering data ownership, provision, and consumption. It is important to describe the core service concepts such as brokerage and clearing house for governance. In the functional layer, interoperability and connectivity are rely on trust, security, and data sovereignty. The identity provider, clearing house, and governance entities fulfill key roles. The process layer provides a dynamic view on the onboarding, exchanging data, publishing data, and using data apps. The information layer provides a vocabulary to describe data by meta data and a common framework for standardized collaboration and establishing agreements and/or contracts. The system layer contains the technical implementation of security levels and governance of intended use.

3.4. Expert opinion

This subsection contains the results of the expert opinion interviews to verify the data space architecture.

3.4.1. Interview setup

An expert opinion is conducted based on a structured questionnaire and interview to discuss the results. The draft version of this paper was shared with a questionnaire with open questions regarding (1) the motivation and goal, (2) the feasibility and solution alternatives, (3) benefits, risks, and drawbacks, (4) assessment of the layered architecture, (5) assessment of the use of the IDS RAM, and (6) overall evaluation. Four experts on data sharing with experience in the logistics industry participated in the expert opinion. All participants were interviewed individually and provided consent to incorporate their input in this paper and acknowledgment.

3.4.2. Summary of results

We summarize the most salient insights from the expert interviews. Overall, the experts were positive regarding the motivation of the smart truck parking data space and goal to share truck parking location data and provide smart services. In terms of feasibility, two experts indicated that the feasibility should be evaluated by means of an actual implementation. This will also clarify the benefits, risks, and drawbacks. The experts indicated that the layered architecture is clear and incorporates the main IDS concepts. One expert emphasized that concrete viewpoints for participants can clarify the role and use of the data space. Some questions were raised regarding the public availability of truck parking location data. The experts advised to make the smart truck parking solution available as part of a larger data space (e.g., mobility). Another expert introduced the idea to develop a location-based service that can be deployed on the on-board computer of a truck to provide parking recommendations based on the tachograph and available nearby truck parkings taking into account driving time regulations.

4. Case study and prototype

This section summarizes the results and discusses challenges from earlier work and presents the developed prototype based on the steps of CRISP-DM.

4.1. Case study: A1 Truck Parking

This subsection summarizes the main results and discusses the challenges and solutions in the case study at A1 Truck Parking.

4.1.1. Business understanding

The case study at A1 Truck Parking in Deventer is conducted in collaboration with the Province of Overijssel, which developed a long-term program focusing on improving truck parking facilities. The main challenges during the business understanding phase linked to stakeholder involvement and data ownership. During interviews with responsible policy makers from the Province of Overijssel, a total of 14 stakeholders were identified, as depicted in **Figure 4**.

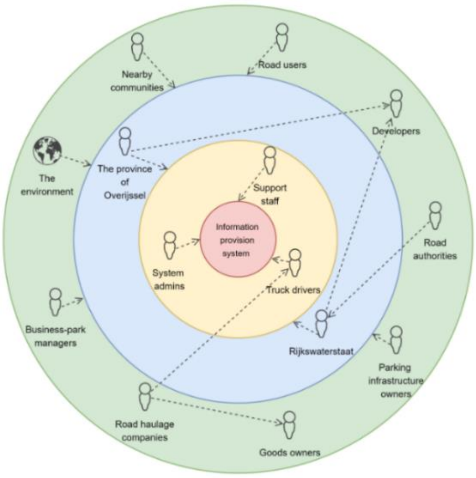

Figure 4: Stakeholder diagram based on [8].

This stakeholder analysis, which is based on the onion model of Alexander and Robertson [8], visualized the complex nature of the problem, (lacking) relations between public and private organizations, and difficulty to formulate a clear business case. However, truck drivers directly experience the effects of the core problem, they are given instructions depending on their planning and are constrained by regulations regarding driving times and mandatory resting. Transport planners make the decisions which truck parking to visit, but have no or limited insight in the occupancy and availability of truck parkings. Unless there is a booking platform in place, the communication is unstructured and data including planned stops is not shared. Parking infrastructure owners have commercial interests and no clear incentive to share data with road authorities and invest in digital infrastructures. Road authorities maintain the public infrastructure and have no direct relation with parking infrastructure owners. Current ITS can measure traffic flow and influence driving behavior via Dynamic Route Information Panels (DRIP). Although road authorities could incorporate truck parking occupancy in DRIPs and Province might be willing to invest in digital infrastructures, they have no access to the data of truck parkings. The case study showed that the willingness and capability to provide data raises specific challenges. Initially, the aim was to conduct case study research at three truck parkings in Overijssel. Eventually, the owner of A1 Truck Parking Deventer was willing to collaborate and capable to provide data in time in a structured format. Detailing the benefits for each individual stakeholder proved helpful to involve stakeholders and convince them to collaborate in the project. This emphasizes the need for data sovereignty to build trust and offer a clear business case with incentives for the involved stakeholders.

4.1.2. Data understanding

Available data were analyzed during the data understanding phase. A1 Truck Parking has a capacity of 100 parking places and uses an electronic toll gate system. The provided Excel file contained records with 1.5 years' worth of historical transactions. The raw data did not contain clear labels, which introduced challenges regarding data augmentation. Based on the instances, most columns were self-explanatory or could easily be verified. Each incoming and outgoing vehicle is registered with a timestamp. Current occupancy is not registered, but could easily be calculated using the timestamps. Measuring and visualizing the hourly, daily, and weekly variations proved to be effective to understand the main predictive features. The provided dataset showed clear temporal patterns and spikes during the night and weekend.

4.1.3. Data preparation

The data preparation phase focused on feature engineering and collecting additional data with localized weather conditions and Dutch and German holidays. Preparing the data revealed that three weeks of data were missing (3.7%) due to the temporary closure for construction work. The missing values were replaced by means of an imputation method using a custom averaging technique. Additional features were created, including time of the day, day of the week, weekend, and holiday. Additionally, a lookback window was created of eight hours. This way, fluctuations in historical occupation can be measured and incorporated as feature.

4.1.4. Modeling

The initial evaluation of machine learning algorithms during the modelling phase was done using RapidMiner [9], a data science platform that allows simultaneous comparison of multiple algorithms, tracking performance and runtime. More specifically, the Root Mean Square Error (RMSE) and runtime in seconds were evaluated. For the given dataset, the RapidMiner platform recommended the following algorithms: generalized linear model, artificial neural network, decision tree, random forest, gradient boosted trees, and support vector machine. After the initial evaluation and results in related works, the decision tree algorithm was selected due to the best runtime performance with 4 seconds of training time and third best RMSE of 0.0029. For the actual model development, the Python library Scikit learn was utilized. The dataset was split into training (80%) and testing (20%) sets to allow model validation. Additionally, a variant of the 5-fold cross validation, called time series split, was used for hyperparameter tuning. The maximum depth and minimum sample split of the decision tree were tuned simultaneously by applying a grid search. Two candidate models were created for feature selection. The first candidate uses preselected features from related work regarding urban car parking occupancy prediction. The second candidate model relies on a feature elimination strategy. In sequential experiments, we excluded a category (or multiple categories) of features to determine their importance for the prediction. The features were clustered based on their data source and nature (e.g., temperature and rainfall; holidays in the Netherlands and in Germany). The experiment yielded a second candidate model, with a simpler configuration, which slightly outperforms the first candidate model when tested on the training set. The experiments suggest that truck parking behavior is not heavily influenced by meteorological conditions and holidays. For accurate predictions, utilizing solely time features and historical occupancy suffices and actually slightly improves performance.

4.1.5. Evaluation

Both candidate models were compared in depth during the evaluation phase. To compare performance using the test set, we selected the RMSE, Mean Absolute Error (MAE), R^2 (coefficient of determination) and adjusted R^2. The results of the model evaluation show that the second candidate model outperforms the first on every metric, with lower prediction errors and higher correlation coefficients. Therefore, the second decision tree was selected as the prediction model for deployment.

This candidate model configuration has important additional advantage regarding its ease of implementation, as all input features can be derived from a singular data source and the model requires fewer resources and development efforts. Additionally, it is less computationally demanding in a runtime environment, which contributes to user experience, and suitable for a (near to) real-time predictive system.

4.1.6. Deployment

During the deployment phase, a design for an integrated dynamic information system was proposed for actionable truck parking occupancy predictions and dynamic rerouting to nearby truck parkings. The high-level system architecture visualizes the main components and interfaces for the access of data sources, data retrieval, (pre-)processing of obtained data, (temporary) storage of data, making truck parking occupancy predictions, communicating predictions (incl. performance and errors), and information provisioning. Additionally, the system architecture addresses deployment issues related to security and proposes incorporation of identity and access management. From a dynamic point of view, the proposed system is updated on an hourly basis by measuring ingoing and outgoing trucks. The time dependent features and historical occupancy are simultaneously generated and stored in the database. Using the most recent input data, the decision tree predicts parking occupancy for the next hour. Based on the differences between observations and predictions, the decision tree is periodically updated in an online fashion. An Application Programming Interface (API) links third parties to the system. The system provides a dashboard for monitoring, informs truck drivers via a mobile application, whereas road authorities can incorporate truck parking occupancy in Dynamic Route Information Panels (DRIPs). Due to time constraints the actual development was not part of the case study. Extending earlier work, a prototype was developed and instantiated. Next, the developed prototype is described.

4.2. Prototype

This subsection presents and describes the prototype for the smart truck parking application.

4.2.1. Prototype development

Based on the case study at A1 Truck Parking, a prototype was developed to demonstrate the solution architecture to the involved stakeholders. **Figure 5** depicts the prototype and its components and interfaces.

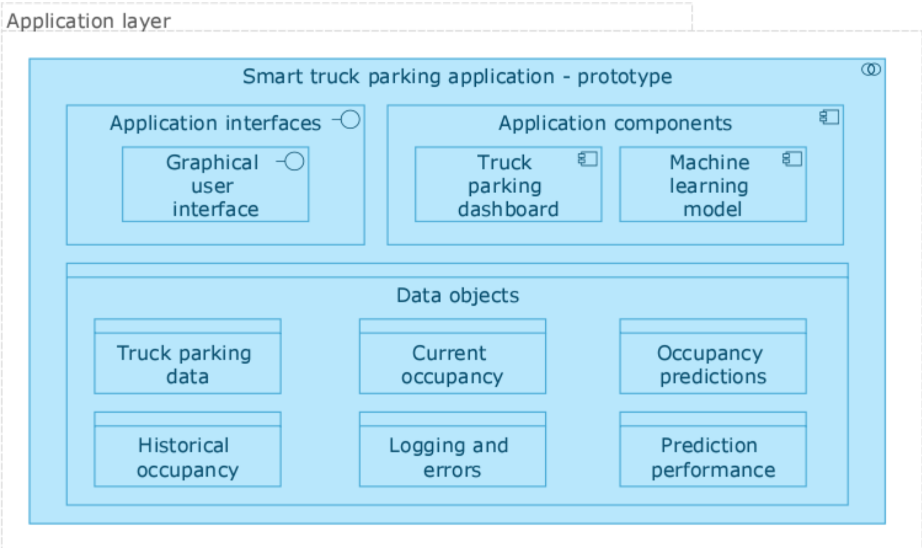

Figure 5: Smart truck parking application prototype.

The instantiated prototype contains a database to store the data objects, the machine learning model for predictions, and the dashboard. The prototype visualizes the current occupancy of the truck parking and provides insight into temporal patterns. Although the prototype demonstrates feasibility, it has some limitations. For one, it runs on a local computer and does not incorporate APIs for dynamic retrieval of data from data sources of truck parkings and IDS gateways. Furthermore, the prototype does not contain the envisioned application components for identity- and access management, nor usage policies and authorizations.

4.2.2. Entity Relationship Diagram

The database stores static data on truck parking, (pre-)processed input data, generated predictions, and errors. **Figure 6** presents the Entity Relationship Diagram (ERD) used for the prototype. The ERD fulfils the requirements of the machine learning model and provided data of A1 Truck Parking. As a result of the model selection, the ERD contains no external features such as weather or holidays.

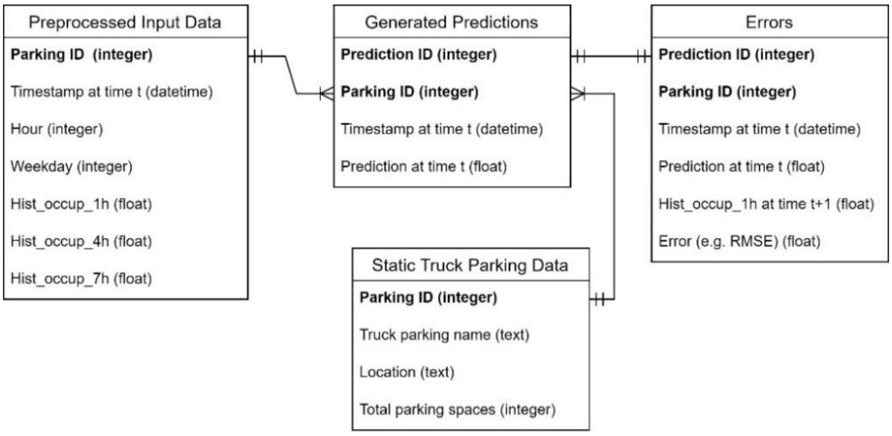

Figure 6: Entity Relationship Diagram for the prototype.

4.2.3. Dashboard

The dashboard visualizes the expected occupancy of the truck parking. **Figure 7** depicts the established prototype dashboard using data of A1 Truck Parking.

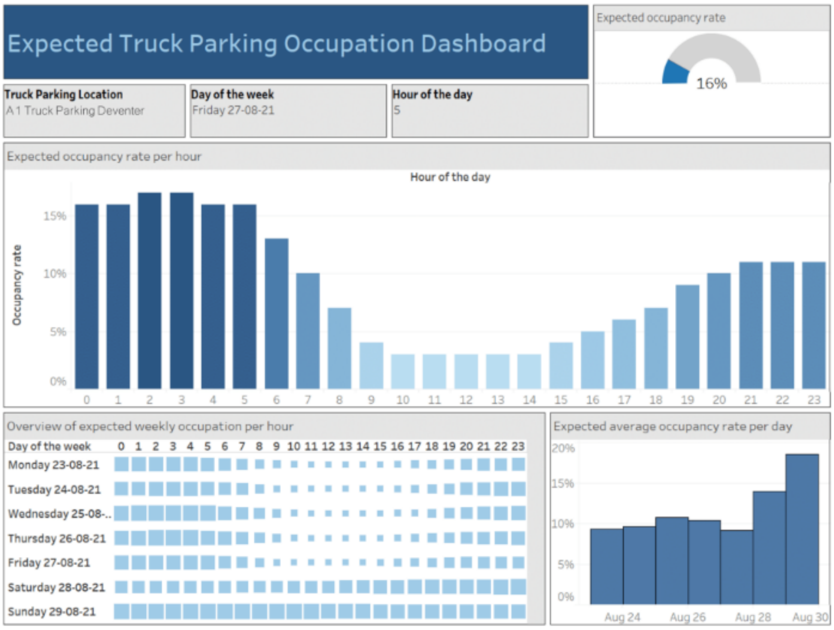

Figure 7: Prototype dashboard based on data of A1 Truck Parking.

To build the prototype, we generated predictions for a week by training a model consisting of time variables only (hour of the day, day of the week, and weekend or not). Then, to make the dashboard more interactive, filters were added, allowing the user to specify the location of the truck parking area and the day and time in which they are interested. Finally, we implemented the forecast into four visualizations. First of all, the dashboard includes a gauge chart (in the top right), which gives an instant overview of the expected occupancy rate of the chosen parking location for the specified day and time. Second, to visualize the expected occupancy rates over time, a bar chart (in the middle) shows the expected occupation for every hour of the day specified by the user. Third, a type of heat map (in the bottom left) shows an overview of the expected occupation for the whole week by the hour. The varying size of the squares resembles the variability of the expected occupancy rates. Fourth, the dashboard includes another bar chart (in the bottom right), which displays the expected averages of the occupancy rates for each day of the week. The proposed dashboard design provides valuable insight into the expected parking lot occupation not only to truck drivers but also to parking infrastructure owners. The design is validated with representatives from the A1 Truck Parking Deventer, who regarded it as useful for their planning and analysis of the occupation.

5. Conclusion

This research presents a novel reference use case and data space architecture for smart truck parking using the IDS RAM. The data space architecture is modelled using the ArchiMate language and specification using its layered viewpoint. The data space architecture incorporates the mandatory roles and components based on the trust security profile and is mapped to the five layers and three perspectives of the IDS RAM. Extending earlier work, the current work takes a step forward regarding the actionable use of truck parking occupancy predictions by presenting prototype for a smart truck parking application.

The reference use case and data space architecture provide a starting point to develop the smart truck parking application and discuss emerging mobility data spaces. However, some aspects require additional modelling and stakeholder interaction. The first aspect is the app store, which may be provided with the data app, but can also be part of a larger data space (e.g., mobility data space). The second aspect is the business model for the data space. The current architecture does not include billing. Stakeholder interaction is essential for further design verification and development of the data space for smart truck parking.

Contrasting research regarding of urban car parking occupancy prediction, the literature regarding truck parking occupancy prediction is limited. Furthermore, most related work uses statistical approaches and incorporate external factors to a limited extent. Leveraging features from related work regarding urban car parking occupancy prediction, our earlier work explored and integrated features from multiple data sources to develop a machine learning approach. Based on case study research at A1 Truck Parking and a systematic comparison and evaluation of various machine learning models, the decision tree performed best for one hour ahead predictions with a RMSE of 0.0029 and required 4 seconds training time. The case study implies that prediction of truck parking occupancy is influenced most by historical and temporal features. Freight transport appears stable with respect to weather conditions. This might depend on the location and moderate climate in the Netherlands. The prototype demonstrates the feasibility of the solution architecture. However, the research and prototype rely on a single case mechanism experiment, limiting the model validation. Additional case study research is needed, ideally at locations with more extreme weather conditions, to assess the relevance of weather conditions as predictive feature.

As data spaces are still in an early stage of development, it is important to develop a testbed and adoption strategy. The testbed can demonstrate the potential benefits of the data space architecture to the stakeholders and identify challenges that may arise during development and implementation. It is recommended to involve software providers of truck parking platforms that can fulfill the role of software/service provider to establish an ecosystem of data. This way, the testbed provides the foundation to experiment with federated machine learning among truck parkings while ensuring data sovereignty. Considerable work needs to be done to align the data objects to the information model of the IDS RAM, which is currently limited to conceptual representation and textual description. Existing interoperability standards and smart parking data models can possibly be used and/or extended. It is recommended to involve standardization bodies and evaluate existing standards. When proceeding development of the data space for smart truck parking, the IDSA should be involved to assess compliance with the IDS RAM, technical support, and certification. In line with this development step, additional experts could be brought in to strengthen the analysis and add novel viewpoints.

Future work may evaluate and develop ontologies based on conceptual representation and experimental development of federated machine learning approaches among truck parkings. Together with the involved experts, the reference use case will be further developed and implemented as project use case in an industrial testbed.

6. Acknowledgements

The authors would like to thank the Province of Overijssel and A1 Truck Parking Deventer for their support. Additionally, the authors thank Danniar Firdausy, Harrie Bastiaansen, Thomas Nijman, and Robbert Janssen for their expert opinion.

7. References

[1] Slavova, S., Piest, J.P.S., and van Heeswijk, W.J.A. (2022). Predicting truck parking occupancy using machine learning. Procedia computer science, 201, 40-47. DOI: https://doi.org/10.1016/j.procs.2022.03.008.

[2] Paleyes, A., Urma, R.G., and Lawrence, N.D. Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Comput. Surv. Just Accepted (April 2022). DOI: https://doi.org/10.1145/3533378.

[3] Zantalis F, Koulouras G, Karabetsos S, and Kandris D.A. (2019). Review of Machine Learning and IoT in Smart Transportation. Future Internet. 11(4):94. DOI: https://doi.org/10.3390/fi11040094.

[4] Otto, B., Steinbuß, S., Teuscher, A., and Lohmann, S. (2019). Reference Architecture Model [online]. Version 3.0. Available at: https://internationaldataspaces.org//wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf [last accessed 22-10-2022].

[5] The Open Group (2019). ArchiMate® 3.1 Specification [online]. A Standard of The Open Group. Available at: https://pubs.opengroup.org/architecture/archimate3-doc/ [last accessed 22-10-2022].

[6] Fraunhofer. (2016). Use case methodology [Power-slides]. Available at: https://industrialdataspace.jiveon.com/login.jspa [last accessed 22-10-2022].

[7] Fiware. (2022). Smart data models [online]. Available at: https://www.fiware.org/smart-data-models/ [last accessed 22-10-2022].

[8] Alexander, I., and Robertson, S. (2004). Understanding project sociology by modeling stakeholders. IEEE software, 21(1), 23-27. DOI: https://doi.org/10.1109/MS.2004.1259199.

[9] RapidMiner. (2022). RapidMiner [online]. Available at: https://rapidminer.com/ [last accessed 22-10-2022].