Kalpa Publications in Computing
Volume 22, 2025, Pages 1043–1057
Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

A Distributed Twin System for Managing Asset-process Interactions in Highway Infrastructure Systems

Mengtian Yin[1], Tom Kelly[1], Pieter Pauwels[2], Liming Xu[1], Brian Sheil[1], Junxiang Zhu[1], Yue Xie[1], and Ioannis Brilakis[1]
[1] Department of Engineering, University of Cambridge, United Kingdom
[2] Department of the Built Environment, Eindhoven University of Technology, Netherlands
`my424@cam.ac.uk`, `twk22@cam.ac.uk`, `p.pauwels@tue.nl`, `lx249@cam.ac.uk`, `bbs24@cam.ac.uk`, `jz652@cam.ac.uk`, `yx388@cam.ac.uk`, `ib340@cam.ac.uk`

Abstract

High agencies face challenges managing scattered asset data across maintenance processes and information systems, obstructing efficient retrieval of dynamic cross-system road information for timely interventions. This paper presents a Digital Twin (DT)-based data federation framework to effectively manage fragmented systems and dispersed data for highway infrastructure operation and maintenance. The framework middleware can decompose users' queries and requests from different subsystems based on a metadata database and a distributed system architecture. The connected data ecosystem enables dynamic communication between different asset systems, ensuring efficient maintenance planning and process coordination between different users and teams. The presented framework is demonstrated based on datasets and synthesised systems conforming to asset management practices adopted by United Kingdom (UK) National Highways.

1 Introduction

Highway infrastructure plays a vital role in a nation's transportation system. However, road networks are deteriorating rapidly with increasing traffic flows (Department for Transport, 2022). Degraded highway assets, such as pavement potholes and drainage blockages, pose significant safety issues for road users. As a result, managing effective maintenance of ageing road infrastructure is becoming increasingly vital yet challenging, necessitating crucial decision-making and timely maintenance interventions to prevent asset failures and traffic incidents.

I. Hkust, J. Cheng and Y. Yantao (eds.), ICCBEI 2025 (Kalpa Publications in Computing, vol. 22), pp. 1043–1057

High-quality data and functioning information systems are critical determinants for maintenance planning and execution. Unfortunately, poor information management (IM) has emerged as a barrier for transportation authorities in effectively executing maintenance operations. In the context of highway maintenance, IM refers to a systematic process of data definition, collection, storage, sharing, and usage of data related to the maintenance and operation of highways (Highways England, 2016). In the current practice, transportation agencies collect highway asset data for various business processes and store the data in different information systems. These information systems were developed by different software vendors and procured in different historical periods, so they often use incompatible data schemas and system architectures (National Academies of Sciences and Medicine, 2021). Due to the lack of interoperability and communication between these systems (see Figure 1(a)), asset data is inaccurate, incomplete, conflicting, and less reliable for use in analytics. As introduced in our previous survey (Yin *et al.*, 2024), National Highways (NH) manages most motorways and A roads in the United Kingdom (UK); it currently employs over 50 information systems and data repositories for various asset systems and processes. For example, NH uses the CONFIRM system for routine maintenance management, the Pavement Asset Management System (P-AMS) for pavement systems, and Highways Agency Drainage Data Management System for managing drainage system. These systems manage the data of the same asset in different business logic; each uses different software data models that operate independently from each other. For example, a change of asset status in one system either cannot be synchronized to other systems or entails substantial manual efforts to update other database systems.

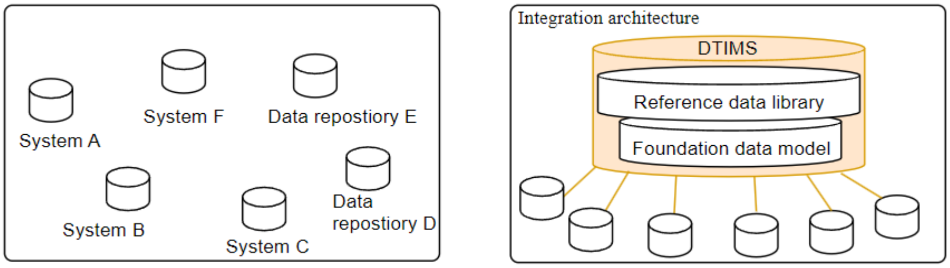
Poor quality of asset data also hinders effective maintenance activities. As illustrated in Figure 1(c), if a change of asset information from a business process cannot be synchronised from one system to other systems, other processes scheduled for the same asset in other systems may be interrupted. For example, a traffic sign is replaced in an emergent repair task (Process A) after a traffic accident, but this information is not sent to another system for routine maintenance management (System B). This causes other engineering teams to receive instructions to inspect or clean the traffic sign that has already been removed (data redundancy effect). The uncoordinated asset data management and maintenance operations lead to economic loss and safety risks to both road users and site workers. It would be beneficial to (a) have a service for end users to retrieve information of assets and processes produced in other systems for planning, design, and coordination; and (c) establish linkages between systems for information sharing.

To address the problems of system fragmentation and data redundancy, this paper presents a novel Digital Twin-based Data Federation (DTDF) framework for managing asset-process interactions across heterogeneous information systems in highway infrastructure maintenance. Grieves first describes the concept of Digital Twin (DT) in manufacturing domain: "*a set of virtual constructs that fully describes a potential or actual physical product from the micro atomic level to the macro geometric level*". A twin system consists of a physical entity, a DT, and there are bi-directional data flow between physical and digital spaces. According to the UK National Digital Twin Programme (2020) of the UK, a DT in the built environment field refers to a digital representation of assets, processes, or systems. Our proposed DTDF framework offers a method to connect the fragmented information systems within an integration architecture, as presented in Figure 1(b).

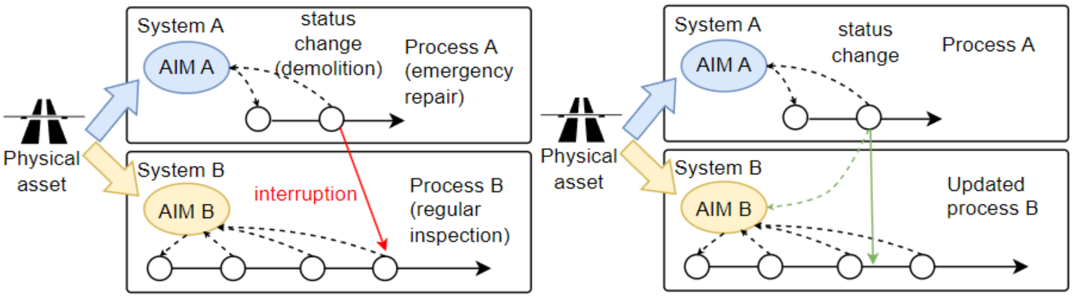
The DT-based connected data ecosystem facilitates seamless data integration and sharing in federated Asset Information Models (AIMs), therefore resolving challenges in cross-system process coordination, as shown in Figure 1(d). This study is among one of the first studies to tackle the system fragmentation problem in highway asset management domain. The contributions of this research can be outlined as follows:

1. This study presents a new data federation method for highway IM based on ontology-based metadata modelling. The DTDF framework generalises the main layers and components of the distributed system architecture to handle data queries towards heterogeneous highway asset systems.

2. This study demonstrates how the proposed system integration approach can effectively check the conflicts between asset interventions at different levels of maintenance processes.



(a) Fragmented information systems (b) DT-based connected data



(c) Interruption on a process due to status change from another process (d) DT-based cross-system process coordination

Figure 1: Problem statement (left) and the proposed DT-based solution (right)

The rest of this paper is structured as follows. Section 2 reviews the relevant studies about the existing IM practices in highway maintenance domain and the DT concept. Section 3 presents the proposed DTDF framework. Section 4 presents a case study in a real-world highway road. Section 5 discusses and concludes this research.

2 Information Management in Highway Infrastructure O&M

Highway agencies started to implement asset management principles (AMPs) in the 1990s (Haas and Hudson, 2015), which represented a systematic coordination of investment, planning, design, construction, operation, maintenance, and demolition of infrastructure assets to maximise their lifecycle values (Uddin, Hudson and Haas, 2013). Asset management principles emphasise asset lifecycle data integration based on Asset Management Systems (AMSs), which refers to a set of interrelated and interacting elements of an organization and the processes needed to achieve Asset management objectives (Hastings and Hastings, 2021). IM takes place within the AMS, or a project management framework (Davidson *et al.*, 2022), where Project Information Models (PIMs) and AIMs are produced and exchanged during the project delivery phase and operational phase, as specified in ISO-19650 (Davidson *et al.*, 2022).

Modern transportation AMSs are becoming increasingly complex, involving many different subsystems (Asghari and Hsu, 2021). In general, there are two kinds of information systems that manage AIMs in the highway domain. The first kind of system looks after a specific asset class and supports professional activities regarding an individual asset system, such as condition investigation and rehabilitation. Typical asset systems include pavement management systems (PMS), bridge management systems (BMS), drainage management systems (DMS), geotechnics management systems, and electricity monitoring systems (Haas and Hudson, 2015). Each asset system integrates relevant contextual information to support its unique workflows. For example, DMSs often incorporate regional flooding data to assist in the decision-making of drainage system maintenance (National Highways, 2023). The second kind of systems concentrate on specific maintenance processes. There are different levels of maintenance processes managed by various agencies, departments, and contractors, such as project-level maintenance, routine maintenance, emergent maintenance, and natural hazard management. Developing a software system that handles all relevant data and operations is common practice for process management in highway maintenance.

Although AMS frameworks aim for seamless interconnection between all subsystems, the integration of existing asset systems and maintenance systems is inadequate and challenging. This causes asset data located in different systems to be inconsistent. According to a report by the Danish Road Directorate (Ebbesen, 2021), the existing AMSs are out-of-date and fragmented, obstructing a systematic coordination across different assets. Data warehousing and data distribution are two primary means of system integration (Li, 2018). Data warehousing is the process of gathering all information from different sources into a centralised data repository (Oti and Gharaibeh, 2020). A centralised database simplifies data management and data retrieval, but it necessitates significant development costs and induces problems with system complexity and data storage. Therefore, real-world AMSs only reach a certain level of data warehousing for specific asset types or processes. On the other hand, data distribution refers to an architectural approach that enables the access and retrieval of data distributed in multiple, disparate sources (Haas, Lin and Roth, 2002). Even though distributed systems allow flexible data sharing, synchronisation, and cross-database queries to support various asset management use cases, it is difficult to handle disparate data formats, referencing systems, and the heterogeneity of various software components.

Digital Twins are emerging as promising solutions for tackling such IM problems. The DT-driven Information Management Framework (IMF) proposed by the UK National DT programme (Hetherington and West, 2020) consists of three essential components: Foundation Data Model, Reference Data Library, and Integration Architecture. Foundation Data Model is an upper-level ontology that addresses problems like space-time, actuality, and granularity. So far, National DT programme has already published a 4D ontology to represent the temporal properties of data (NDT, 2023). Reference Data Library offers controlled vocabularies with a set of classes and attributes to describe domain-specific knowledge. Integration Architecture stands for a digital system that manages multiple asset information systems and offers protocols for data authorisation, sharing, discovery and retrieval across distributed users and providers.

3 Proposed Data Federation Solution

3.1 Overview of the proposed DTDF framework

The proposed DTDF framework provides a systematic approach to integrating dispersed asset data and the associated processes, which includes two parts: (a) a distributed system architecture to integrate fragmented asset systems; and (b) workflows for highway organisations and practitioners to operate twin systems.

The main purpose of the design and development of the distributed twin system is to integrate fragmented AIMs and associated processes stored in different information systems. Specifically, integration refers to the ability to (a) discover and query data stored in different database and file systems and (b) synchronise the related data elements across different subsystems. The DTDF framework focuses on two important functional features of data and system integration for managing asset-process interactions:

1. Retrieval of information of assets and processes for users to efficiently coordinate, planning, and design maintenance activities. For example, maintenance planners may need to know the scheduled activities of nearby assets, so they can identify conflicts or coordinate with other teams to share working zones.
2. Change propagation management to ensure consistency across systems. In other words, modifying asset data during project or operational phase in one system can influence AIMs and processes managed by other systems. Therefore, it is necessary to compute the impacts of data modification and notify the influenced systems.

These two functional features are closely related to the problems of asset-process interactions discussed in Section 1. Therefore, we devised the system architecture and workflows based on the requirements of these two features, and we abstract the components of the system architecture to ensure the scalability for different application scenarios and use cases.

3.2 System architecture design

The overview of the DTDF system architecture is shown in Figure 2. The distributed twin system contains three interconnected layers: user access layer, federation middleware layer, and data source layer. Additionally, there is a component for security and access control that interacts with all three layers.

User access layer interacts with end users (e.g., maintenance planners) and applications that require retrieval of federated data. End users operate a graphical user interface (GUI) to discover published DT models (i.e., asset systems), query the required data (e.g., nearby assets and recent events). The incoming requests are routed to the federation middleware layer for processing.

Federation middleware layer manages the data flow between the user's request and the various data sources (subsystems). The federation middleware breaks down the queries and interacts with different systems using APIs (Application Programming Interfaces). The ontology and data standardisation layer provides a common ontology to clearly define the domain knowledge (e.g., classification of assets), a metadata schema, and a metadata database. Metadata refers to data that describes data of assets and process stored in subsystems. Metadata includes key description information, such as asset ID, process ID, and system ID of the associated information system. A metadata schema is a structural representation that describes the data schema of metadata. The data aggregation and transformation engine perform ETL (extraction, transform, load) to convert the raw data objects into standardised formats (e.g., Resource Description Framework (RDF)) and send them back to user access layer. Raw data objects may be validated using predefined constraint rules.

The data source layer contains all underlying information systems and data repositories that need to be integrated within the connected data ecosystem. The connected subsystems include different asset systems, such as PMS, BMS, and DMS. Due to the heterogeneous software architecture of different subsystems, a connector is needed for each subsystem to connect with the federation middleware layer.

Figure 3 shows the structure of a metadata schema (ontology) and data examples in the metadata database in form of RDF instance graphs. The minimal required metadata includes (a) traffic elements (e.g., road number, road section); (b) road network assets; (c) processes associated with different assets; and (d) system. The metadata database stores linkage data between AIMs, processes, and systems to allow federation middleware to communicate with correct subsystems for a given data query. For example, the routine maintenance information of pavement assets can be found in a system called Confirm.

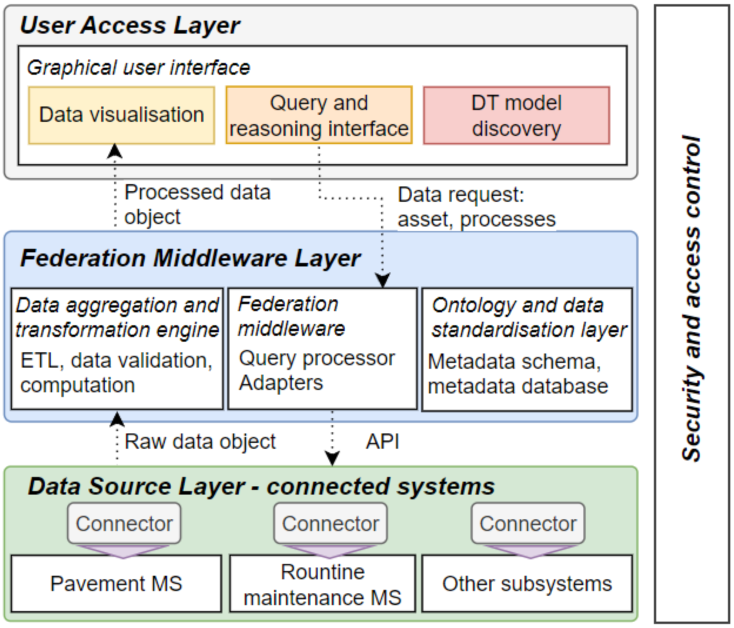


Figure 2: System architecture of the distributed twin systems for integrated highway asset management. Dotted line denotes information flow

3.3 Workflows

The relevant workflows within the DTDF framework are shown in Figure 4. We consider two kinds of workflows: (a) the workflow to facilitate the data federation; and (b) the workflow for end users to access cross-system data and perform their processional maintenance activities.

1. Data federation workflow

Apart from necessary software development according to the system architecture described in Figure 2, Highway organisations should appoint a party (e.g., data management department) to organise the existing datasets and systems to achieve data federation and integration. This IM process is accompanied by the system development process. The workflow for data preparation is as follows.

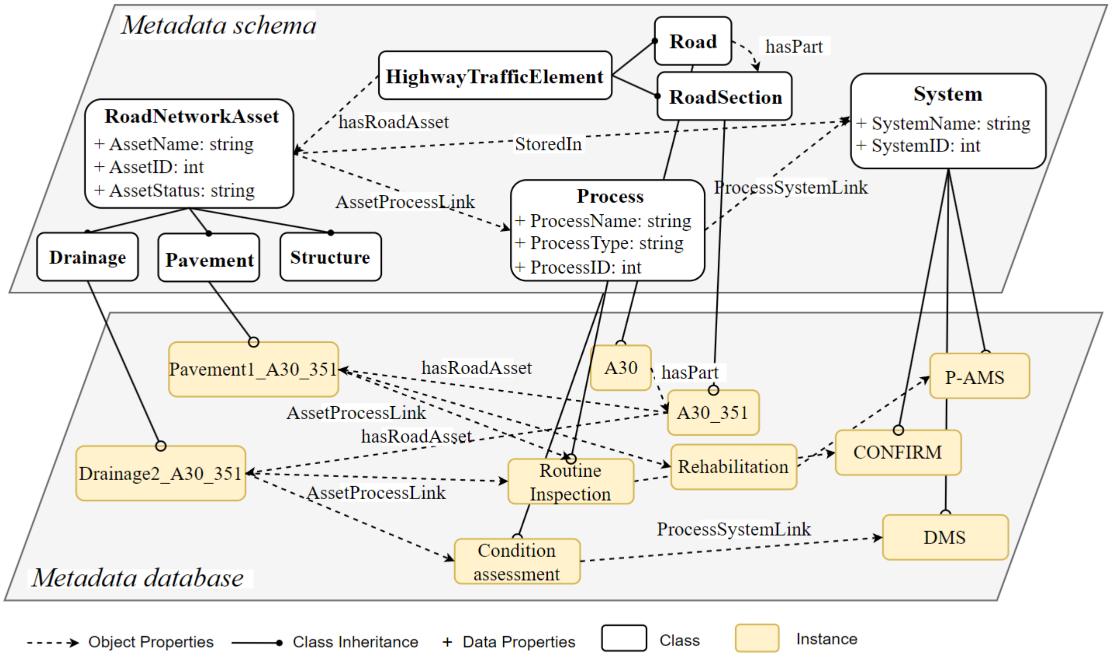


Figure 3: Ontology-based metadata schema and metadata database in the federation middleware of the highway twin system

a) Develop a common ontology to clearly define the classification and hierarchy of assets, processes, and systems. The structure of the ontology is highly dependent on the existing asset management and IM practices in organisations.

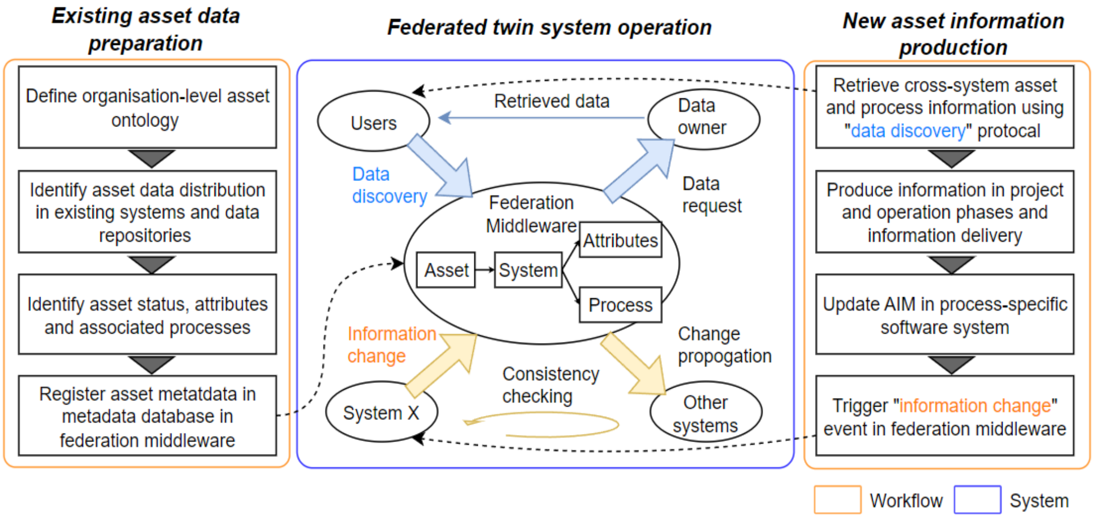


Figure 4: Workflows for facilitating DTDF framework (left) and for end users to access cross-system data (right)

b) Identify the existence of all road assets and the distribution of asset data in different software systems, file systems, and data repositories. For example, pavement asset data may be managed by PMS, Maintenance management systems (MMS), and traffic management systems simultaneously.

c) Identify asset attributes, such as construction and condition data, as well as the associated maintenance processes in different systems. For instance, a routine MMS may schedule regular visual condition surveys for gullies, and a DMS may arrange professional inspection activities for the same drainage system.

d) Collect different subsystems' data models of the identified assets and relevant processes. Software vendors collaborates with highway organisations to develop more specific metadata schema based on the common ontology developed in stage (a), information requirements of use cases, and the existing data models, following the IM framework of ISO 19650 (Davidson *et al.*, 2022). Eventually, metadata is generated and registered in the metadata database of federation middleware. The metadata stores key information to allow federation middleware to be aware of which subsystem the data is stored for processing a query. For example, when a user submits a query: "*retrieve all assets in A30 and the schedules of their maintenance activities in November 2024*", the metadata database would offer information of identifiers of assets, processes and systems, then API requests were sent to respective systems for querying detailed schedules.

2. Workflows for end users

As shown in Figure 2, the DTDF provides a user access layer with GUIs for end users to query cross-system data. The design of GUIs and query interfaces depends on the service provided to end users. The generic workflow can be described as follows:

a) Users who log in to the federated twin system operate the GUI to discover published DT models (e.g., pavement network models from pavement asset systems) in 2D/3D geospatial scenes and request the required data for their professional tasks. The design of the data query protocol is based on metadata schema, showing all available data elements shared by different subsystems. The data request is partitioned into sub-queries in the federation middleware and sent to subsystems. The returned raw data objects are processed into standardised formats in a data aggregation and transformation engine before sending back to end users. Apart from data retrieval, inference and computation are also be operated to check the clash between maintenance activities.

b) Users optimise their work plans after acquiring the asset and process data retrieved from different systems. For example, a maintenance engineer may adjust the lane closure dates to work with other teams concurrently after obtaining the information of recent activities planned by other departments. Afterwards, users perform their professional tasks and produce new information in project and operational phases.

c) Asset data managers handle the new information delivery and update AIMs in the corresponding information system. For example, the information of a repair task in a routine maintenance is updated in a routine MMS.

d) Subsystems periodically communicate with federation middleware to update the change of information. To achieve so, an event system is developed in the federation middleware to control propagation of information change in one system to other systems. The event system first register duplicated attributes between different asset data models and dependencies between AIMs and processes in different systems. For example, the status information of assets can influence many cyclic maintenance activities. When an information change event occurs, the event system identifies the related asset attributes and processes using metadata, and notifies other influenced subsystems.

The above workflows allow highway participants to effectively retrieve information that is generated in other business processes. This information helps them coordinate with other departments, agencies, and contractors for maintenance planning and execution.

4 Case Demonstration

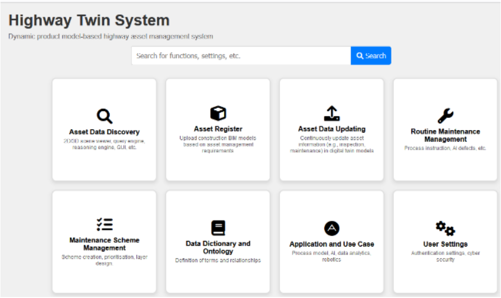
We conducted a case study on a real-world highway to demonstrate the proposed DTDF framework. Specifically, we implemented a prototype called a Highway twin system (HTS) that contains a user access layer, a federation middleware layer, and a data source layer. We developed several synthesised asset information management systems and MMSs in the data source layer, all working independently. We then collected asset data of a long road segment in the UK - Chiverton to Carland Cross, of the A30. The selected road segment includes serval roundabouts, overpasses, underpasses, and intersections with other roads.

In this case study, we consider a lighting facility outage event. In this situation, maintenance managers should arrange an engineering team to inspect the associated underground utilities and repair problematic asset components. This activity involves lane closures and the excavation of pavements. In this scenario, the HTS is used to retrieve assets in the same road section and the associated maintenance activities scheduled in different asset systems. This helps different teams coordinate intervention activities considering the dependencies of different assets.

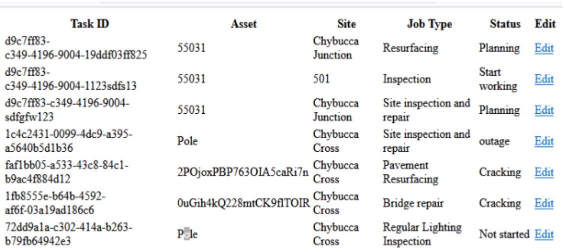
4.1 Implementation details

We implemented the HTS using Django MVC (Model-View-Controller) framework. We created a Django project to manage all components in the user access layer and the federation middleware layer, and several different Django projects that simulate the operation of subsystems. Every project uses a unique port, and they communicate with each other using HTTP protocols. Within the federation middleware, we use the official National Highways ontology (Department for Transport, 2020) as a common organisation-level ontology, and we deployed an ontology-based metadata schema by extending the National Highways ontology with constructs of the existing data models used by different subsystems, including P-AMS (TRL Software, 2022) and CONFIRM (Highway Agency, 2009).

In the user access layer, we use CESIUM (2018), a web-based 3D geospatial development framework, to display data and allow end users to interact with the GUI. The implemented prototype is presented in Figure 5. The HTS offers a main interface (see Figure 5 (a)) for asset data managers to manage the connection with subsystems, metadata registration, and data specifications within the federation middleware. End users can view and retrieve distributed asset data in a GUI in the user access layer (see Figure 5 (c)). Figure 5 (b) shows an interface of a synthesised routine MMS in the data source layer, listing work orders for different assets

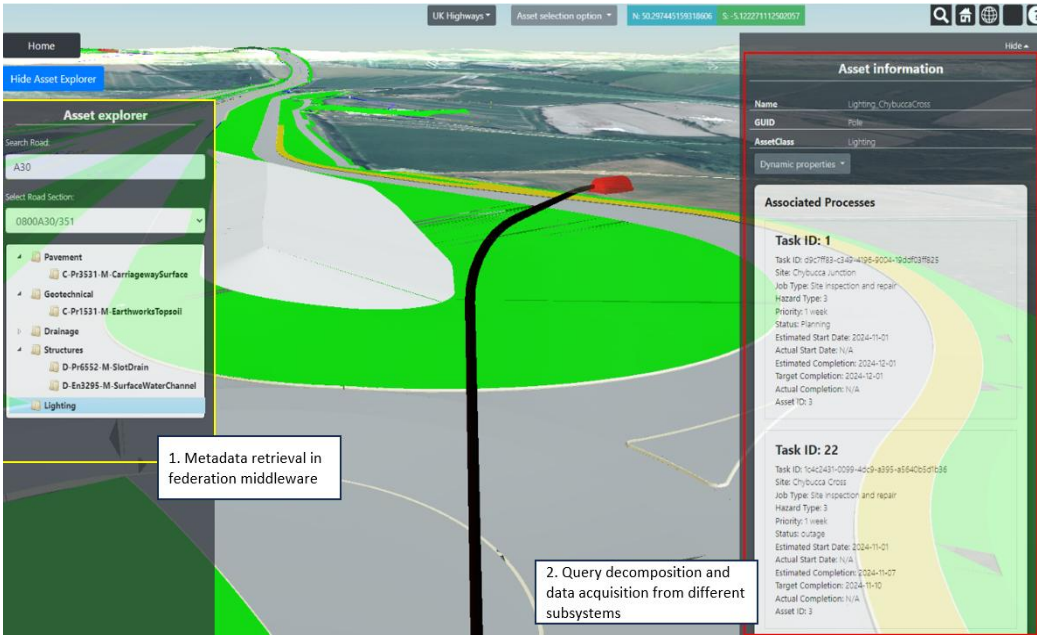


(a) Main interface of HTS



Task ID	Asset	Site	Job Type	Status	Edit
d9c7ff83-c349-4196-9004-19ddf03ff825	55031	Chybucca Junction	Resurfacing	Planning	Edit
d9c7ff83-c349-4196-9004-1123sdfs13	55031	501	Inspection	Start working	Edit
d9c7ff83-c349-4196-9004-sdfgfw123	55031	Chybucca Junction	Site inspection and repair	Planning	Edit
1c4c2431-0099-4dc9-a395-a5640b5d1b36	Pole	Chybucca Cross	Site inspection and repair	outage	Edit
faf1bb05-a533-43c8-84c1-b9ac4f884d12	2POjoxPBP763OIA5caRi7n	Chybucca Cross	Pavement Resurfacing	Cracking	Edit
1fb8555e-b64b-4592-af6f-03a19ad186c6	0uGih4kQ228mtCK9flTOIR	Chybucca Cross	Bridge repair	Cracking	Edit
72dd9a1a-c302-414a-b263-b79fb64942e3	Pole	Chybucca Cross	Regular Lighting Inspection	Not started	Edit

(b) Synthesised routine MMS



Name: Lighting_ChybuccaCross
GUID: Pole
AssetClass: Lighting

Task ID: 1

Task ID: d9c7ff83-c349-4196-9004-19ddf03ff825
Site: Chybucca Junction
Job Type: Site inspection and repair
Hazard Type: 3
Priority: 1 week
Status: Planning
Estimated Start Date: 2024-11-01
Actual Start Date: N/A
Estimated Completion: 2024-12-01
Target Completion: 2024-12-01
Actual Completion: N/A
Asset ID: 3

Task ID: 22

Task ID: 1c4c2431-0099-4dc9-a395-a5640b5d1b36
Site: Chybucca Cross
Job Type: Site inspection and repair
Hazard Type: 3
Priority: 1 week
Status: outage
Estimated Start Date: 2024-11-01
Actual Start Date: N/A
Estimated Completion: 2024-11-07
Target Completion: 2024-11-10
Actual Completion: N/A
Asset ID: 3

1. Metadata retrieval in federation middleware

2. Query decomposition and data acquisition from different subsystems

(c) Data query and visualisation interface in

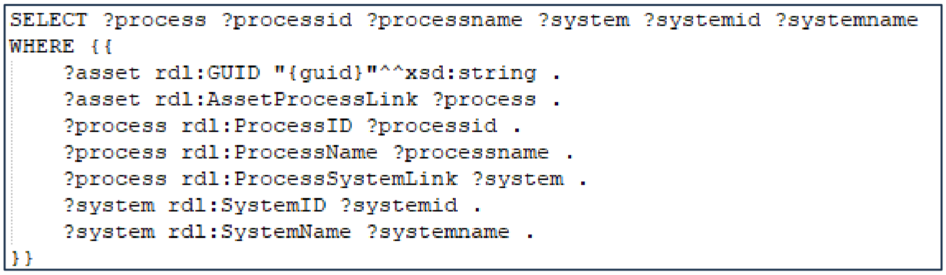
Figure 5: Prototype of the federated twin system for highway maintenance management

4.2 Data collection

We collected asset data of the selected road segment in the forms of Industry Foundation Classes (IFC) specifications (buildingSMART International Ltd., 2019) from industry partners. The collected data involves a diverse range of asset classes, including pavement, earthwork, drainage, utilities, line markings, lightings, traffic signs, and structures. Due to the data confidentiality issues, we generated synthesised maintenance work orders that conform to the data models of the MMSs used by the NH.

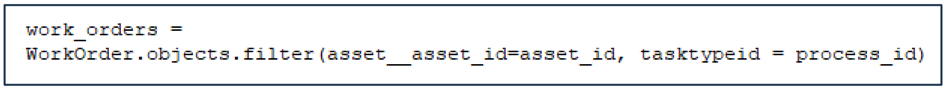
4.3 Results

The resulting RDF instance graph (metadata database) contains 1498 triples, 17 class counts, and 228 instances. We operate the HTS prototype to query asset and process data distributed in different subsystems, as illustrated in Figure 5. Given an example of a lighting outage event, we first retrieve the road number and road section where the lighting facility is placed. As shown in the left-hand side of the interface, different classes of assets located at the selected road section can be returned from the metadata database. Afterward, we can select a specific asset instance or asset group and request its associated process information stored in distributed subsystems, as shown in the right-hand side of the interface in Figure 5. This involves a multi-level data retrieval process as follows:



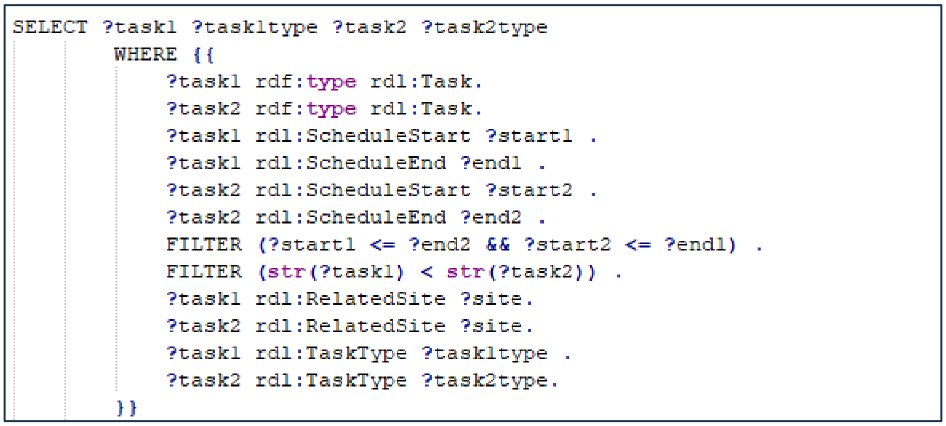
```
SELECT ?process ?processid ?processname ?system ?systemid ?systemname
WHERE {{
    ?asset rdl:GUID "{guid}"^^xsd:string .
    ?asset rdl:AssetProcessLink ?process .
    ?process rdl:ProcessID ?processid .
    ?process rdl:ProcessName ?processname .
    ?process rdl:ProcessSystemLink ?system .
    ?system rdl:SystemID ?systemid .
    ?system rdl:SystemName ?systemname .
}}
```

Listing 1. Metadata retrieval in the federation middleware to acquire the associated processes and systems of an asset.



```
work_orders =
WorkOrder.objects.filter(asset__asset_id=asset_id, tasktypeid = process_id)
```

Listing 2. Data retrieval in the distributed subsystems using localised query mechanism.



```
SELECT ?task1 ?task1type ?task2 ?task2type
        WHERE {{
            ?task1 rdf:type rdl:Task.
            ?task2 rdf:type rdl:Task.
            ?task1 rdl:ScheduleStart ?start1 .
            ?task1 rdl:ScheduleEnd ?end1 .
            ?task2 rdl:ScheduleStart ?start2 .
            ?task2 rdl:ScheduleEnd ?end2 .
            FILTER (?start1 <= ?end2 && ?start2 <= ?end1) .
            FILTER (str(?task1) < str(?task2)) .
            ?task1 rdl:RelatedSite ?site.
            ?task2 rdl:RelatedSite ?site.
            ?task1 rdl:TaskType ?task1type .
            ?task2 rdl:TaskType ?task2type.
        }}
```

Listing 3. Clash detection of tasks retrieved from different asset systems.

a) An ontology-based metadata retrieval is first operated in the federation middleware to retrieve any processes linked to the selected asset and systems linked to the processes, as shown in Listing 1. The query processor extracts key information, such as ProcessID and SystemID, from SPARQL queries to perform next-level data query in subsystems.

b) Having obtained a list of AssetID, ProcessID, and SystemID, the query processor decomposes the query into sub-queries for different subsystems before an adapter sends these sub-queries to subsystems using HTTP GET protocol.

c) When a subsystem, such as a routine MMS, receives the data request, a connector transforms the data request into queries that comply with the local database settings and schema. For instance, a query for the synthesised MMS based on the Django MVC framework is based on Python and SQL commands, as presented in Listing 2. The retrieved results, such as scheduled tasks, are sent back to the federation middleware.

d) The data aggregation engine gathers the requested JSON objects from different subsystems, populates RDF instance graphs, and conduct data inference before sending the processed data to the front end.

We searched the assets and processes that are spatially close to the broken lighting and perform clash detection, as shown in Listing 3. Table 1 demonstrates an example of searching for assets and processes from other systems that occur at the same time as the lighting utility inspection and repair task. Different engineering teams can coordinate their work based on the interdependencies of assets and processes. For example, a utility repair activity should precede a pavement resurfacing activity because the former may involve excavation of pavements.

Asset class	Associated process	Schedule start	Schedule end	System
Lighting	Utility inspection and repair	2024-11-01	2024-11-07	Routine MMS
Pavement	Pavement resurfacing	2024-10-30	2024-11-03	Pavement MS
Structure	Bridge deck repair	2024-11-01	2024-11-10	Structure MS

Table 1: Conflicted processes of nearby assets in A30 across different systems

Finally, the lighting asset information is updated in the routine MMS after maintenance. This triggers an “information change” event in the federation middleware layer based on HTTP POST protocols. The change propagation can be inferred by predefining the dependency relationships between asset attributes and processes across all systems. In this case, the change of an attribute’s “maintenance date” in the MMS would influence the regular structure integrity inspection activities scheduled in an electricity management system (EMS). Therefore, HTS would be able to send messages to that EMS to notify the change of asset status.

Table 2 shows a qualitative comparison between the HTS and the existing AMSs and data lake method adopted by the UK National Highways. The HTS allows cross-system retrieval of process information, but the existing AMSs fail to provide functions to obtain task data from different asset systems. Based on this, the HTS also offers semantic rules for conflict checking across various maintenance activities, providing opportunities for improved coordination. At last, the existing AMSs do not update or periodically update the latest asset information across information systems. In contrast, our event-driven design of the information updating mechanism allows more timely data synchronization.

Category	Cross-system process retrieval	Automatic asset-process constraint checking	Asset information update
HTS	Yes	Ontology-based conflict checking	Event-driven dynamic update
Existing AMSs	No	No programs for multi-system consistency checking	No update or periodically update

Table 2: Qualitative comparison between HTS and existing AMSs

5 Conclusion

Data and system integration is a longstanding challenge for highway agencies to efficiently plan and perform maintenance activities. The difficulty originates from heterogeneous structures of different software components, incompatible data models, and poor communication protocols. As a result, the road asset data is often inaccurate and inconsistent across different systems and the associated processes are often uncoordinated. There is a lack of clear information management processes and enabling systems to effectively integrate dispersed datasets and fragmented systems in an economic way.

This paper presents a novel Digital twin-based data federation (DTDF) framework based on the UK NDT Information management framework. The DTDF framework includes a distributed system architecture and two workflows. The distributed system architecture offers a federation middleware to decompose queries from user access layer into sub-queries and communicate with subsystems. A distributed data query is enabled by (a) metadata schema and metadata database to retrieve the linkage between assets, processes, and systems; and (b) communication protocols and connectors to activate data retrieval and processing in subsystems. On the other hand, the DTDF framework also illustrates IM workflows for (a) preparing and registering the existing asset data in the federation middleware; and (b) end users to discover published DT models, retrieve cross-system data, and produce new information. Compared with the existing studies, such as (Le and Jeong, 2016), that purely focuses on data modelling for decision-making, this study provides a holistic solution to integrate fragmented asset systems, including system architecture, metadata modelling, and communication protocols.

We demonstrate the DTDF framework by implementing a system prototype called Highway Twin System (HTS), which consists of (a) a web-based software system that supports interactions with end users and processing federated queries; (b) several independent software systems that simulate the operation of fragmented asset information systems and maintenance management systems; and (c) communication protocols to connect different systems. We tested the HTS based on a case study on real-world highway road in United Kingdom. Different kinds of road assets, including pavements, drainage, utilities, and structures, are registered in the HTS. We then operate the HTS to handle a simulated lighting outage event. The results show that HTS can allow end users to effectively search nearby assets that are located at the same road section and retrieve their attributes, as well as the associated processes distributed in different subsystems. These results can help different departments, engineering teams, and agencies in highway organisations coordinate their activities, avoiding clashes and improving productivity by scheduling co-intervention procedures.

This research benefits the highway industry by offering a practical and cost-effective data federation solution to manage fragmented systems and dispersed datasets without the need for re-development, restructuring, and migration of existing asset database systems. The connected data ecosystem can help different practitioners within HOs efficiently retrieve asset data and perform data analytics for different data-driven asset management processes. It may also be used by users from external parties to plan, design, construct, operate, and maintain other infrastructure assets, considering the interdependencies between infrastructure systems. Furthermore, a federated twin system makes it possible to cross-validate asset data and manage the influence of information change in one system on other systems, improving the data quality, data accuracy, and data completeness.

There are several limitations of this research. First, the scope of the distributed data query is small, and it only includes basic asset data attributes and process information, such as schedule start date and schedule end date. Second, it is inevitable to make efforts to develop connectors for different subsystems. It may be laborious to investigate the existing data schemas and construct appropriate data queries within subsystems. Furthermore, mapping data schemas of different subsystems to ontology-based metadata schema may be difficult when handling larger road networks or more complex asset systems. Third, the proposed framework does not provide a complete cybersecurity mechanism to restrict the user access and data usage rights. In the future work, the DTDF framework will be further extended and refined with more real-world use cases implemented and evaluated.

6 Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034337. We wish to thank Dr. George Economides, Dr. Stefano Cavazzi, Dr. Matt Peck, Yogesh Patel, Kieran Mulvey, Simon Hayton, James Hewson, Sameer Kesava, and Alex Gillies for their contributions to this paper.

References

Asghari, V. and Hsu, S.-C. (2021) 'An open-source and extensible platform for general infrastructure asset management system', *Automation in Construction*, 127, p. 103692.

buildingSMART International Ltd. (2019) 'Industry Foundation Classes (IFC)'. Available at: https://technical.buildingsmart.org/standards/ifc.

CDBB (2020) 'The approach to delivering a National Digital Twin for the United Kingdom', p. 20. Available at: https://www.cdbb.cam.ac.uk/files/approach_summaryreport_final.pdf.

Cesium (2018) 'Cesium - an open source Javascript library for world-class 3D globes and maps'. Available at: https://cesiumjs.org/.

Davidson, S. *et al.* (2022) 'ISO 19650 Guidance 1: Concepts, Edition 3'.

Department for Transport (2020) *Highways England and Local Authority : Data Discovery*. Available at: https://assets.publishing.service.gov.uk/media/5fb52dc4d3bf7f63d7075a44/DfT-and-Highways-England-data-discovery-report-accessible.pdf.

Department for Transport (2022) *Road traffic statistics*. Available at: https://roadtraffic.dft.gov.uk/summary#:~:text=Road traffic by road type&text=motorway traffic increased by 13.3,billion vehicle miles of traffic.

Ebbesen, M. (2021) 'Data-driven asset management on the state roads of Denmark', *Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations - Proceedings of the 10th International Conference on Bridge Maintenance, Safety and Management, IABMAS 2020*, pp. 3215–3219. Available at: https://doi.org/10.1201/9780429279119-436.

Haas, L.M., Lin, E.T. and Roth, M.A. (2002) 'Data integration through database federation', *IBM Systems Journal*, 41(4), pp. 578–596. Available at: https://doi.org/10.1147/sj.414.0578.

Haas, R. and Hudson, W.R. (2015) *Pavement asset management*. John Wiley & Sons.

Hastings, N.A.J. and Hastings, N.A.J. (2021) 'ISO 55000 series standards', *Physical Asset Management: With an Introduction to the ISO 55000 Series of Standards*, pp. 595–621.

Hetherington, J. and West, M. (2020) 'The pathway towards an information management framework-A "commons" for digital built britain'.

Highway Agency (2009) *Network Management Manual*. Available at: https://assets.publishing.service.gov.uk/media/5a8000cf40f0b62302690edd/nmm_part_Composit_0_9.pdf.

Highways England (2016) *Information Management and Technology Strategy*. Available at: https://assets.publishing.service.gov.uk/media/5a7fffb840f0b62302690e77/Information_Management_and_Technology_Strategy.pdf.

Le, T. and Jeong, H.D. (2016) 'Interlinking life-cycle data spaces to support decision making in highway asset management', *Automation in construction*, 64, pp. 54–64.

Li, Z. (2018) *Transportation asset management: Methodology and applications*. CRC Press.

National Academies of Sciences and Medicine, E. (2021) ‘Guidebook for Data and Information Systems for Transportation Asset Management’.

National Highways (2023) *GDMS - Geotechnical & Drainage Management Service*. Available at: https://gdms.assetia.cloud/home.

NDT (2023) *Introduction to 4D Ontology*. Available at: https://github.com/dstl/IES4/blob/master/4dOntologyIntro.md.

Oti, I.C. and Gharaibeh, N.G. (2020) ‘A Note on Assessment of Transportation Infrastructure Data Management Practices’, *Public Works Management & Policy*, 25(2), pp. 132–150.

TRL Software (2022) *iROADS:infrastructure asset management system*. Available at: https://trlsoftware.com/products/digital-asset-management/iroads/.

Uddin, W., Hudson, W.R. and Haas, R.C.G. (2013) ‘Public infrastructure asset management’, *(No Title)* [Preprint].

Yin, M. *et al.* (2024) ‘How Can Digital Twins Be Used in Highway Maintenance? A Questionnaire Survey for Industry Practitioners’.