CODEC: CONNECTED DATA FOR ROAD INFRASTRUCTURE ASSET MANAGEMENT

Sukalpa BISWAS[1], John PROUST[1], Tadas ANDRIEJAUSKAS[1], Alex WRIGHT[1], Carl VAN GEEM[2], Darko KOKOT[3], António ANTUNES[4], Vânia MARECOS[4], José BARATEIRO[4], Shubham BHUSARI[5], Uros JOVANOVIC[6]

[1]*TRL Ltd, Wokingham, UK*
[2]*BRRC, Brussels, Belgium*
[3]*ZAG, Ljubljana, Slovenia*
[4]*LNEC, Lisbon, Portugal*
[5]*RHDHV, Amersfoort, Netherlands*
[6]*BEXEL, Ljubljana, Slovenia*
Corresponding author: Sukalpa Biswas, sbiswas@trl.co.uk

Received 15 March 2021; accepted date

Abstract. Road infrastructure asset management is rapidly transforming into a digital environment where data accessibility, effective integration and collaboration and accessibility from different sources and assets are key. However, current asset management processes are not yet fully integrated or linked, and there are incompatibilities between various systems and platforms that limit the ability to integrate asset management with BIM. The CoDEC project has sought to understand the current status of information management for assets, including inventory, condition and new data sources such as sensors and scanning systems, to identify the challenges and needs for linking and integrating different data sets to support effective asset management. As a result, CoDEC has developed a data dictionary framework to help link/integrate static and dynamic data for the "key" infrastructure assets (road pavements, bridges, tunnels). This will enable BIM and Asset Management Systems (AMS) to exchange data and help optimise and integrate data management across systems and throughout the different asset lifecycle phases, from build to operation. This work will be followed up with three pilot projects to demonstrate the feasibility of integrating asset data from various sources through linked data / semantic web technology to build the connection between AMS and BIM platforms.

Keywords: Asset management, BIM, Data dictionary, Linked data, Ontology, Asset data

Introduction

Building Information Modelling (BIM) is an information management process that has the potential to support asset management from concept to the end of life. The process is designed so that asset information can be generated, captured, maintained and used efficiently and effectively to optimise asset management. However, to date BIM processes have focused on the information gathered during the construction phase of the asset and typically do not address the data requirements required during the operational phase. There is a gap in defining the information requirements for the operational phase, and in how to define/accommodate these across BIM systems. The CEDR Transnational Research Programme funded CoDEC project (Connected Data for Effective Collaboration) had aimed to address this gap by establishing a better understanding of how BIM principles could be applied within the European highways industry to manage asset data during the operational phase. In particular, CoDEC has developed a specification to support the establishment of connections between asset management systems and BIM platforms - to make best use of legacy data and sensor/scanner data provided by new technologies. CoDEC has developed a "Data Dictionary" for infrastructure assets that could form the basis for data structures to support integration between different data management systems, and thereby improve the flow of asset data. Ultimately, the project aims to assist Road Authorities make efficient and effective use of data to support asset management.

1. Current state of using data dictionaries for asset management

A review of existing asset management and BIM practice within European National Road Administrations (NRA) was carried out by initially circulating a survey to a wide group of stakeholders. This was followed up with interviews with experts, which included detailed discussion on legacy and new data types (sensor data), data dictionaries, integration between management systems and the use of BIM platforms for highway assets (roads, bridges and tunnels). The review and consultation aimed to provide insight into the current situation within NRAs regarding data within asset management and their roadmaps for integrating asset management data with BIM, to establish the foundations for the development of the data dictionary.

1.1. Legacy data

Within the context of CoDEC we consider legacy data to be data that provides key information on the assets that fall under the responsibility of the NRA, and is the primary dataset applied for asset management. The data can be applied for the purposes of inventory (what is there), condition (how is it performing) and operation (how is it being used). The condition and operational aspect are commonly more challenging with respect to data collection and management.

The review found that the condition of assets is commonly monitored, but management of the condition data itself is disparate. The data are often stored in separate databases managed by individual NRA departments responsible for particular assets. These specialist data are often only accessible to members of the department and are shared with colleagues in other departments on request. Condition and maintenance management information is shared in various formats such as reports and maps. For pavements many NRAs use dedicated commercial Pavement Management Systems (PMS). For bridges, Bridge Management Systems (BMS) may be developed in-house or commercially sourced. Technology equipment installed on the network is again commonly managed by separate departments. Major tunnels are typically managed individually.

Operational asset management data typically includes data from real-time monitoring (cameras), traffic counters, traffic load sensors (via Weigh in Motion, WIM), bridge monitor sensors, monitoring of ventilation systems in tunnels etc. These sensors and devices are all assets that form part of the larger road, bridge or tunnel asset, and hence of the asset network as a whole. This data is typically used for local or network operational management by particular NRA departments, although some of the data is made available to the public (e.g., traffic counts).

1.2. Sensor and scanning data

Emerging technologies are opening up opportunities for NRAs to improve asset data collection, analysis and management. The review identified many technologies that NRAs are starting to use to support management of the network. In particular, a variety of different sensor/scanning technologies and data processing techniques are now available to NRAs, which CoDEC has structured into seven technology families:

- Embeddable and Fixed Sensors (e.g. Weather, WIM, loops, fire detectors, visibility meters, etc.).
- Airborne, terrestrial, and mobile type LiDAR technologies for monitoring highways and structures.
- Satellite Data Monitoring (satellite and aerial imagery, InSAR) and Unmanned Aerial Vehicles (UAVs) for inventory and monitoring of highways infrastructure.
- Internet of Things (IoT) and Connected Sensor Networks.
- Probe Vehicles and systems within Autonomous Vehicles that provide vehicle sensor data that can be used for highways asset management (crowdsourcing).
- Smartphones (crowdsourcing).
- Advanced Data Processing techniques (artificial intelligence, machine learning algorithms, etc.) to extract information on asset condition from images, video footage.

It was found that the integration of data from these new sensors and scanning technologies is not yet as mature as the technologies themselves. Although most of the technologies reviewed have a Technology Readiness Level (TRL) above 6, use cases are typically in the pilot/demonstration stage and are not implemented network wide. However, the extent and maturity of the sensor and scanning technology families differ depending on the asset type. For example, embeddable and fixed sensors, including IoT sensors and sensor networks, are relatively common in structures (bridges and tunnels), while vehicle-based sensing technologies and laser scanning are more common for carriageways.

1.3. Asset management and BIM

Many NRAs use BIM during the design and construction phases of large projects. After construction, the BIM model is sometimes delivered as a detailed "as-built" model, but little further use is made during the remainder of the operational life cycle of the asset (this is more commonly the case for pavement assets than structures such as bridges or tunnels). However, long-term asset management is still typically carried out in a traditional way using dedicated asset management systems:

- Pavement Management Systems (PMS) contain data about the road network, structural information about the layers of the road, traffic density and the current condition (e.g. using condition parameters that can be converted into indicators) of individual lengths of pavement
- Bridge Management Systems (BMS) contain bridge inspection data on each component (usually standardized, e.g. following PIARC recommendations) and data provided by sensors and advanced inspection techniques.
- Tunnel management systems contain data on the tunnel structure itself and the electronic equipment used to support tunnel operation.

Although a few initiatives across Europe have attempted to define standard formats for the data (e.g. AM4INFRA (Marcovaldi, Biccellari, 2018)) contained in these systems in general, the system of data management is very much tailored to the individual AMS within the NRA. Little work has been done to create a standardized "Data Dictionary" for each type of asset. Our review showed that attempts have been made by Highways England through its Asset Data Management Manual (ADMM) (Highways England, 2020), and the Lithuanian NRA has a Lithuanian State Road Information System (LAKIS). However, few other publicly-available data dictionaries were identified in the states contained in the review (Norway, Sweden and Germany), or is other sources providing information (France, Australia/New Zealand). It was found that whilst other countries do not have data dictionaries as such, some NRAs have developed Object Type Libraries (OTL), including Netherlands, Flanders and Finland. Most OTLs/data dictionaries include roads, bridges and tunnels. As noted above, the gap in the availability, extent and content of data dictionaries for highway assets presents a barrier to the exploitation of this data, in particular within the BIM environment.

2. Developing a Data dictionary

The purpose of a data dictionary is to provide a registry of detailed information about the contents of a dataset or database, such as the names of objects, their properties, data types or formats, and text descriptions. A data dictionary thus provides a common schema for anyone needing to use the data it contains and is therefore critical in aligning multiple stakeholders across different organisations, disciplines, and use cases.

One of the key objectives of the CoDEC project is to develop a data dictionary for major highway assets which is both general enough to be widely applicable across various NRAs, while also being specific enough to be operationally useful in practical terms.

The CoDEC Data Dictionary covers three key highway civil assets (pavements, bridges and tunnels), as well as preliminary entries for supporting assets/systems such as lighting, fire-fighting, and drainage. CoDEC has considered both legacy data for these assets, and data from new sources such as sensors and scanners that offer the potential to transform NRAs' future ability to manage highway assets.

The CoDEC Data Dictionary has been developed by building on previous work carried out in AM4INFRA to develop a data dictionary on tunnels and bridges, the Highways England UK-ADMM data dictionary (Highways England, 2020), the Data Standard for Road Management and Investment in Australia and New Zealand (DSRMI, for tunnels) (Austroads, 2019) and ifcRoad (buildingSMART, 2020).

2.1. Data Dictionary Structure

Assets vs components

Infrastructure assets are complex 'objects', with many interdependent elements making up what one might call "an asset". There is not always a clear delineation between what can be called an asset, and what is only a part of an asset – in fact, the delineation depends largely on the viewpoint of the person(s) making the judgement (e.g.: is a single kerbstone an asset? Or is it only a component of a whole kerb – and should this whole kerb then be considered an asset? Or, is the kerb a component of the roadway, and the roadway is the asset?). Further, when trying to define an asset based on its constituent parts, it is important understand the level of detail required – an asset like a tunnel could be described down to its individual nuts and bolts, or more simply in broad terms covering only major components.

The issue of defining an asset by its parts is made even more complex when considering that certain 'parts' of an asset are not really physical, but rather are abstractions used in order to describe and manage an asset – a key example of this is a road 'section' (sometimes called a road 'link'), which is simply a way of longitudinally delineating a length of road using defined start and end points. Although the road section of course corresponds to a real, physical road, the actual physical asset is not in reality 'sectioned' in this way – this abstraction is simply used in order to better manage and understand the road, since the road network must often be considered in separate 'sections' for practical reasons.

For CoDEC the guiding principle applied to address the challenge of defining and delineating assets and their components was to consider the ultimate application: *the need to develop a data dictionary that will support the management of that asset* (taking into account the need to accommodate the needs of sensors and sensor data). The information from the review and the experience and knowledge of the team in infrastructure asset management (i.e.: how assets are actually managed, practically), was brought together to draw the necessary judgements on: (1) what constitutes "an asset" vs the components of that asset, and (2) the level of detail needed to adequately describe that asset for the purposes of management.

It was decided to use a two-tier system for describing assets and their constituent parts, and that this system should be the same for all asset types, in order to maintain consistency (especially in terms of level of detail). Two types of objects were defined: Entities, and Elements: 'Entities' correspond to what are considered as assets; 'Elements' correspond to what are considered as components if those assets.

The terminology is based on ISO 12006-2:2001 (International Organization for Standardization, 2001), which defines the terms 'construction entity' and 'element'. The decision to maintain a consistent definition across all assets presented challenges, since different asset types may have very different levels of complexity in terms of their constituent parts. For example, a road pavement can be broken down into constituent parts in a relatively simple way while retaining a reasonable level of detail, whereas a bridge may require a more complex breakdown to adequately describe its components. Essentially, there are pros and cons either to maintaining a consistent approach or to allowing the approach to vary by asset type. CoDEC decided to prioritise consistency in this case.

The 'classification' layer

It is not strictly a requirement of a data dictionary that the objects within it are classified, or organised, in any particular way. Indeed, when considering only the machine-readable applications of a data dictionary classification is not necessary – it primarily functions as an aid to human-readability. Despite the ultimate goal of using the data dictionary in a machine-readable environment, it was decided that the human-readability was important (especially since it is intended to be widely-shared), and so a classification layer was included.

A two-tier classification system for infrastructure asset entities (class and sub-class) based on *UniClass 2015* (NBS, 2015) was used. So, for example, a bridge would be classified as: Entity Class = "Structures", and Entity Sub-Class = "Bridges". For sensors, the only classification which was meaningful within the context of our data dictionary was whether the sensor was a fixed-location or mobile sensor – therefore only one classification tier for sensors was used, with only these two options. Object properties were classified using a single classification tier: for asset entity/element properties we used a classification based on *Omniclass* (Construction Specifications Institute, 2017); and for sensor properties we used a classification based on *SensorML* (Open Geospatial Consortium, 2020).

A special case - sensors

A key aim of CoDEC is to consider sensors and the data they provide, which are increasingly used to support infrastructure asset management. It was decided that the data dictionary should not be simply a 'directory' of different sensor types and their properties, since this would quickly become an impossible task due to the (ever-increasing) number of different sensors available, and the detailed properties of each one. Hence, CoDEC considered it more useful to develop a set of "general" property sets and definitions which would apply to any type of sensor. The data dictionary is hence future-proofed and will remain useful in a wide variety of use cases.

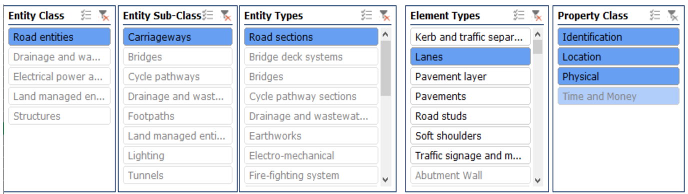
Therefore sensors were not considered as 'assets' in themselves, but rather as separate objects. The data dictionary focused on identifying the property sets which would apply across in general to types of sensor. However, when considering sensors CoDEC considered it necessary to develop different property sets for sensors that have fixed-locations and those that are mobile. A key reason for this is to address differences in the approach taken to referencing the location of fixed and mobile sensors. In addition there can be differences in how sensors are defined - for example, one can consider an array (or network) of multiple fixed-location sensors but this does not apply to mobile sensors. Therefore, sensors CoDEC has placed sensors in their own dedicated section of the data dictionary, separate from asset entities and elements.

2.2. Data Dictionary Content

Once the structure was decided, the 'content' of the data dictionary was developed – i.e. identifying and defining the assets (entities), components (elements), and their properties; the sensor types and their properties; and the properties of the sensor data.

From our review of existing data dictionaries developed for the purpose of highways asset management it became clear that there were very few relevant examples available in the public domain. Ultimately, CoDEC identified two potential sources: the data dictionary developed as part of the AM4INFRA project (Marcovaldi, Biccellari, 2018), and Highways England's (English NRA) Asset Data Management Manual (ADMM) (Highways England, 2020)). However, a detailed review of both of these sources concluded that neither fully met needs of the CoDEC project as the AM4INFRA dictionary tends to a high/theoretical level, whilst the ADMM is set at a lower level with specific focus on Highways England's operating context. However, CoDEC drew on these to design a Data Dictionary that fall within the space between them - i.e. detailed enough for practical application, yet generic enough to be applicable across the many operating contexts found in European NRAs.

Further review, combined with workshops undertaken within the CoDEC consortium (drawing on the practical expertise of members to translate knowledge of asset management and asset data directly into the data dictionary), and assessment of real-world datasets was used to inform the content of the data dictionary. This included considering examples of sensor-produced datasets to identify the general properties shared across sensors, which provided key input to the 'Sensor' and 'Sensor Data' sections of the dictionary. Further workshops were held in which the data dictionary content was presented and discussed with representatives from CEDR NRAs to validate the approach and the content. An extract from the data dictionary is shown in Figure 1 (the figure is truncated to fit, and as such does not show all fields). This shows an example of the content and structure for the 'Lane' element (part of the 'Road section' entity).



Objects				Properties		
Entity Class	Entity Sub-Class	Entity Types	Element Types	Property Class	Property Name	Property Definition
Road entities	Carriageways	Road sections	Lanes	Identification	Lane ID	Unique reference identifier for lane section
Road entities	Carriageways	Road sections	Lanes	Identification	Pavement section ID	Unique reference identifier for pavement section
Road entities	Carriageways	Road sections	Lanes	Identification	Lateral position	The lateral position of the lane (related to the function of the lane) and represented as a code from NRA's classification (e.g. CL1, CR1)
Road entities	Carriageways	Road sections	Lanes	Identification	Lane designation	Additional information on the lane designation (e.g. permanent running lane, hard shoulder, etc.)
Road entities	Carriageways	Road sections	Lanes	Physical	Geometry type	How the geometry of the asset/component is represented - for example: linear, point, polygon
Road entities	Carriageways	Road sections	Lanes	Physical	Width	The width of the lane
Road entities	Carriageways	Road sections	Lanes	Location	Latitude (Start)	Latitude coordinate, at the start of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Longitude (Start)	Longitude coordinate, at the start of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Altitude (Start)	Altitude, at the start of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Latitude (End)	Latitude coordinate, at the end of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Longitude (End)	Longitude coordinate, at the end of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Altitude (End)	Altitude, at the end of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	Start chainage	Start chainage of the lane section
Road entities	Carriageways	Road sections	Lanes	Location	End chainage	End chainage of the lane section

Figure 1: Extract from the CoDEC data dictionary showing the entries under the 'Lane' element

3. Linked Data Methodologies and Tools

Linked Data and Semantic Web methodologies use ontologies to structure and share data. An ontology can be defined as a "formal, explicit specification of a shared conceptualization" (Studer et al., 1998), meaning that concepts, their constraints, and their relationships are encoded in a way that is systematically structured, explicit and machine-readable. This allows ontologies to be used to integrate and retrieve information, obtain semantically enhanced content, and to support knowledge management.

The Resource Description Framework (RDF), RDF Schema (RDFS) and the Ontology Web Language (OWL) were developed by the World Wide Web Consortium (W^3C). The RDF provides the basis for the "creation, exchange and use of annotations on the web" (Pan, Horrocks, 2009), using statements in the form of triples (subject, property, object). RDFS introduces class and hierarchy concepts, while OWL provides additional vocabulary and expressiveness (e.g. disjointedness, cardinality, object and data properties).

The CoDEC Data Dictionary draws on these concepts to provide a shared vocabulary and common language to enable the integration and sharing of data between different management systems. Linked data technologies, such as ontologies, are used to encode asset and sensor data in a formal, comprehensible, and explicit way.

3.1. Methodology for developing Linked Data

The Data Dictionary describes asset data using concepts, relationships and properties about highway assets, sensors, and sensor data. As noted above, the CoDEC ontology has been developed to build on the EUROTL framework (INTERLINK project, 2018). As a first step each Data dictionary concept or relationship was mapped to an existing class or property in EUROTL, as shown in Table 1. Properties can be defined either as an object property or data property, meaning a semantic relation between classes (for objects) or between the class and data (e.g. strings or numbers). Where a mapping is not present in the EUROTL CoDEC has created a new class or property (the CoDEC ontology has been developed using Stanford's Protégé (Musen, 2015)). However, this is always a sub-class of an EUROTL concept. This means that new CoDEC classes are an extension of the previous and ensures interoperability between the two ontologies.

As an example, the Bridge concept already exists in the EUROTL Framework (AM4INFRA:Bridge (Marcovaldi, Biccellari, 2018)). However the concept of a Structural Element (or equivalent) is not found in EUROTL. Hence, a new Structural Element class was created in the CoDEC ontology, as a sub-class of the already existing EUROTL concept EurOTL:PhysicalObject.

Table 1. Example of data dictionary to ontology mapping

Data Dictionary			**Ontology**		
Property	Description	Format	Domain	Object/Data Property	Range
Bridge ID	The unique reference identifier for bridge	String	bridgeID	is-a	Bridge
Bridge name	The name of the bridge	String	bridgeID	rdfs:label	xsd:string
Environment	Classification of surrounding environment (e.g. Rural/Urban)	String	bridgeID	inEnvironment	xsd:string
Region/District/Area	Relevant geographical situation	String	bridgeID	prov:atLocation	eurotl:LocationByIdentifier
Owner	Owner of the asset	String	bridgeID	hasOwner	prov:Agent (person or Org.)

3.2. API

An API has been developed by the CoDEC project to support data linking between systems. This is based on the principle of OpenAPI, as shown in Figure 2. On the bottom layer, the data integration environment is composed of a set of aligned ontologies, namely the EUROTL framework, the CoDEC ontology (based on the data dictionary) and a sensor ontology (the Sensor Network Ontology (Open Geospatial Consortium, 2020)). In order to manage the complexity of the linked data environment and create a separate layer that can be used without interfering with other layers, CoDEC exposes a set of REST services. These services are responsible for communicating with the linked data environment (typically through a set of SPARQL queries), and can be used by any application, as long as it has permission to access both services and data.

This layered approach has several advantages, the most critical one being the separation provided by multiple layers, which allows modification of the linked data structures without affecting the normal behaviour of external applications, as they "just need" to know how to call the services (their inputs and outputs).

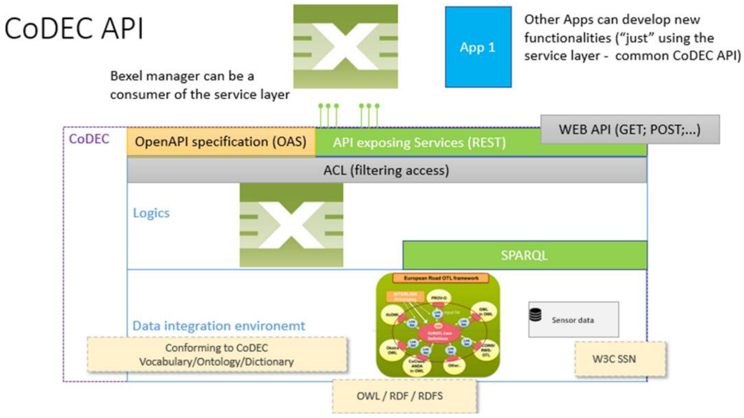


Figure 2. CoDEC API overview

3.3. Linked Data structure

Once the CoDEC ontology is finalized, with all the required classes and relationships, the next step is to integrate real data (namely, the data required for the pilot projects – see the next section for more details). The ontology is populated with data instances and made available in a linked data environment in order to be retrieved by external applications (via the API). The linked data environment that CoDEC uses is GraphDB, a database that follows RDF and SPARQL (query language for ontologies) specifications. Figure 3 presents a sample linked data structure for the fictional example of data for a bridge and a pavement.

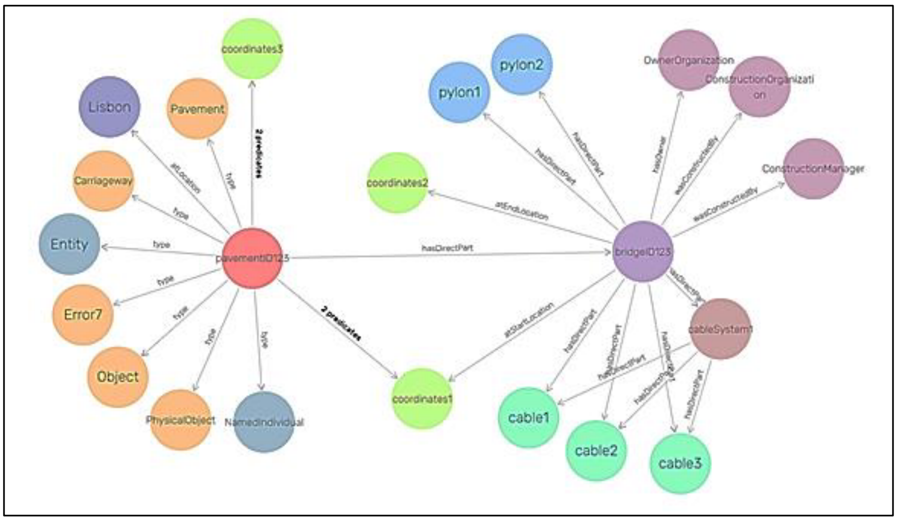


Figure 3. Sample of linked data using CoDEC ontology for a bridge and a pavement.

4. Demonstrating the concept

CoDEC will demonstrate the application of the data dictionary and ontology developed during the project, using a linked data approach, through three pilot projects. The pilot projects aim to:

- Demonstrate the connectivity between the Data Dictionary and EUROTL.
- Demonstrate visualisation of the integrated data (including sensor/scanned data sets) in an asset

management system.

– Demonstrate the ability to link data between BIM models, GIS and Asset Management Systems.

The projects will include the development of tools (including an Application Programming Interface, API via the Bexel Manager and the use of 3D visualization software) to demonstrate linking data between BIM platforms and asset management systems.

The pilot projects are being conducted with the support of three implementation partners: Agentschap Wegen & Verkeer (AWV, Belgian (Flemish) NRA) (Project 1); Rijkswaterstaat (RWS, Dutch NRA) (Project 2) and FTIA (Finnish NRA) (Project 3):

– Pilot project 1 – **Tunnel** - to show how the BIM model of a Tunnel can be enriched. It will link the BIM OTL to tunnel sensor data (such as thermal, visual sensors etc.) and provide an advanced 3D visualisation of this linked model.
– Pilot project 2 – **Bridge** - to integrate data from multiple bridge specific sensors, and demonstrate software tools and guidelines for linking, preparing and visualizing bridge data.
– Pilot project 3 – **Highway** – to link data between BIM models and a road network within a GIS (Geographic Information System) environment, requiring accurate spatial mapping between the two.

At the time of writing this paper all three pilot projects are still in progress. The section below provides a description of pilot project 3 in order to give an example of the way in which the CoDEC ontology is being applied in a practical use case.

4.1. Pilot project 3 - Linking BIM & GIS for Highway Asset Management

Whilst BIM models are often created for the design/construct phase, highways are primarily managed during the operational phase using GIS-based Asset Management Systems (AMS). Likewise, LiDAR scans – increasingly being used to gather physical data about highways assets – are often translated into BIM models rather than GIS-based information. In all of these cases, BIM models often hold information useful for asset management which is not being made available in the GIS-based AMS where these assets are managed. In other words, BIM models represent a new source of data which can be used to enrich (and/or complement) the data held within legacy systems – this is a particular focus of the CoDEC project. Pilot project 3 aims to address this issue by providing methods to link data between the BIM and GIS-based environments.

This pilot project focuses on the carriageway pavement as the key asset. Although some of the methods developed are specific to the challenges of linking pavement asset data, the general approach is more widely applicable to other non-linear highways assets. The project has access to multiple BIM models of sections of the road network within TRL's Smart Mobility Living Lab, located in the London Borough of Greenwich, UK, as well as a GIS-format network model of the whole of the borough's road network (provided by the local roads authority). Within the BIM models, the pavement is represented as a series of 2D 'slabs' of ~10m length. The challenge for this pilot project is to derive useful information from the geometry of these slabs, and then link that information with the corresponding location (i.e.: 10m subsection) within the GIS network model – an example is given in Figure below. The pilot project team has had to take innovative approaches to tackling both of these challenges and have developed solutions which will be disseminated in future CoDEC communications.

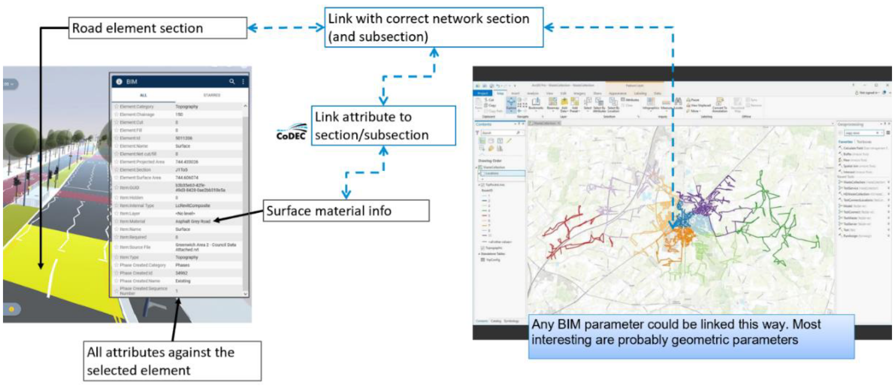
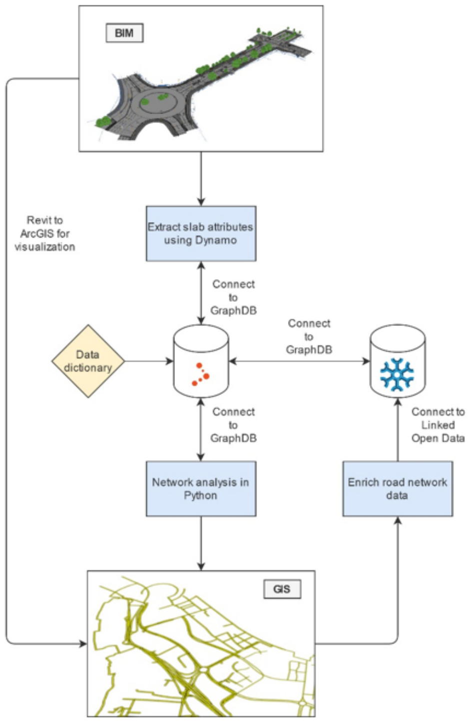


Figure 4: Example illustrating the linking of data for a particular subsection of pavement between the BIM model and GIS

The CoDEC ontology (described above) is being used to provide a structuring framework for organizing and linking the relevant data for this pilot, and so far is proving successful in that task. The linking methodology is summarized in Figure below. The pilot project has provided a practical opportunity to test the content and structure of the ontology (which is based on the data dictionary) – this has directly resulted in improvements and additions to the ontology.



From BIM-csv export to Linked Database

1. Read the csv file in Python.
2. Convert the longitude and latitude values to British National Grid coordinates
3. Assign the coordinates to the corresponding slab object.
4. Create a convex hull around the slab object.
5. Read the vertices of the convex hull and store them in a GeoSPARQL Polygon object in Well Known Text (WKT) format.
6. Build a SPARQL query to insert the slab data to the Linked Database.
7. Execute the query using the GraphDB API.

From Linked Database to GIS

1. Build a SPARQL query to read the slab data from the Linked Database.
2. Execute the query using the GraphDB API.
3. Convert the slab geometries to ArcGIS Polygon features.
4. Intersect the slab features with the road network features.
5. Measure the start and end chainage of the intersect result on the network.

From GIS to Linked Database

1. Store the network information acquired through the GIS analysis to the slab object in Python.
2. Build a SPARQL query to insert the slab to the Linked Database.
3. Execute the query using the GraphDB API.

Figure 5: Steps involved in linking data between the BIM models and the GIS using the CoDEC-developed tools and schema

Conclusions

Consultation with European NRAs showed many NRAs use BIM during the design and construction phases of large projects and sometimes stored as a detailed "as-built" model, but little further use is made during the remainder of the operational life cycle of the asset. Long-term asset management is still typically carried out using dedicated asset management systems. Lack of the interoperability of the data types, their standards, and isolated data management/storage principles act as barriers for effective data exchange.

It is clear that new data sources (from new sensor and scanning technologies) hold huge potential for improving the way that we manage highways assets – indeed, it is absolutely necessary that highways asset management adapts to capitalize on the increasing pace of the digital revolution. However, it is important to take a methodical and structured approach to dealing with the underlying data – the increasing volume, variety and velocity of the data demands it.

The CoDEC project (although not yet complete) is developing and demonstrating practical ways to structure asset-related data in such a way as to make them operationally useful. The CoDEC approach relies in large part on having a methodically developed framework for the data (the data dictionary structure), and translating this into a machine-readable framework (the ontology), in order to provide a means of making data more interoperable – a step on the journey to the ultimate goal of making data seamlessly available when and where it is needed.

It is important to note that it has taken significant effort to get this far, and there is still further to go within the scope of the project. However, CoDEC already shown that in ongoing pilot projects, with an initial investment of effort, it is possible to create a data structure which, when applied using the right methods, unlocks great potential for enriching legacy data, systems, and ways of working. When used within a digital-led organizational context, there is huge potential for such an approach to unlock value from (and add value to) asset data, ultimately leading to multiple benefits in terms of increased asset knowledge (and therefore better asset management), more efficient use of data (and therefore lower costs), and more effective use of data (and therefore greater benefits).

Acknowledgements

The CoDEC project would like to thank all the experts who were involved in stakeholder consultation activities by sharing their knowledge, experience and providing suggestions for the project.

Funding

The CoDEC project was carried out as part of the CEDR Transnational Road Research Programme Call 2018: BIM. The funding for the research was provided by the national road administrations of Austria, Belgium – Flanders, Denmark, Finland, Germany, Netherlands, Norway and Sweden.

References

Austroads. (2019. Data Standard for Road Management and Investment in Australia and New Zealand (DSRMI), version 3.0. Retrieved from https://austroads.com.au/publications/asset-management/ap-r597-19

buildingSMART. (2020). Industry Foundation Classes (IFC), IfcROAD. Retrieved from http://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/

Construction Specifications Institute. (2017). OmniClass Construction Classification System. Retrieved from https://www.csiresources.org/standards/omniclass

Highways England. (2020). Asset Data Management Manual (ADMM), version 12.0. Retrieved from https://www.standardsforhighways.co.uk/ha/standards/admm/index.htm

INTERLINK project. (2018). Final Project Report – INTERLINK Approach with European Road Object Type Library. Retrieved from https://roadotl.eu/static/ireport/index.html

International Organization for Standardization. (2001). ISO 12006-2:2001 Building construction — Organization of information about construction works — Part 2: Framework for classification of information

Marcovaldi, E., Biccellari, M. (2018). Asset Data Dictionary, Deliverable D3.1 of the AM4INFRA project. Retrieved from https://am4infra.eu/

Musen, M.A. (2015). The Protégé project: A look back and a look forward. AI Matters. *Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015*. DOI: 10.1145/2557001.25757003.

NBS. (2015). Uniclass 2015. Retrieved from https://www.thenbs.com/our-tools/uniclass-2015

Open Geospatial Consortium. (2020). OGC SensorML: Model and XML Encoding Standard, version 2.1. Retrieved from https://www.ogc.org/standards/sensorml

Pan, J. Z., Horrocks, I. (2007). RDFS(FA): Connecting RDF(S) and OWL DL. *IEEE Transactions on Knowledge and Data Engineering, 19(2), 192-206*. https://doi.org/10.1109/TKDE.2007.37

Studer, R., Benjamins, R., Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering, 25(1–2):161–198,* 1998.