Data-Centric Solutions – A Case Study for Road Data Based on Knowledge Graphs

Per-Ola Olsson[1], Johan Wahlberg[1], Justin Hellsten[1], Felix Hansson[1], Lars Wikström[2], Matthew Calvert[3], Peter Axelsson[4], Lars Harrie[1]

[1] Department of Earth and Environmental Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden – per-ola.olsson@mgeo.lu.se; johanjwahlberg@gmail.com; justin.hellsten@hotmail.com; hanssonfelix97@gmail.com; lars.harrie@mgeo.lu.se
[2] Triona AB, Borganäsvägen 26, 784 33 Borlänge, Sweden – lars.wikstrom@triona.se
[3] City Planning Department, Fleminggatan 4, 104 20 Stockholm, Sweden – matthew.calvert@stockholm.se
[4] BIM Alliance Drottninggatan 33, 111 51 Stockholm, Sweden – peter.axelsson@bimalliance.se

Keywords: data-centric, knowledge graphs, CityGML, road data.

Abstract

This study demonstrates how a data-centric architecture can be implemented with a knowledge graph based on generic datasets to enable a road maintenance tools. Two datasets were used: a 3D city model (CityGML/3CIM) and a national road database. Road polygons (CityGML/3CIM) were generated from a municipality basemap and knowledge graphs were created for both datasets. The data-centric tool was implemented in the open-source GIS software QGIS and evaluated to find how well the generic datasets fulfill the data requirement of a road maintenance tool. Much of the data required for road maintenance could be obtained from the two generic datasets; however, some information is missing. To overcome this, the information models for the generic datasets need to be extended or a dedicated information model could be created. For the latter, the information model should only contain the required road maintenance attributes and links on object level to the CityGML/3CIM objects to avoid separate descriptions of the road geometries; such an approach would avoid the data-silo structure that is common within municipalities today.

1. Introduction

Urban planning and maintenance are complex processes involving several actors from the public as well as the private sector. The process requires large amounts of data such as geospatial data (e.g. 3D city models) used by the municipalities, BIM data used by architects and construction companies, transport networks for traffic planners, as well as data included in operational systems. In this study we focus on data for road maintenance which are activities required to preserve, repair and operate transportation infrastructure. Typical information required in road maintenance is the transport network geometry (both in link-node and polygon representations), attributes describing the physical properties of the roads, administrative information (ownership, organisations responsible for maintenance etc.), status of the roads (including activities done and planned for e.g. reparation).

Today, data for road maintenance are commonly locked into a proprietary operational system with its own internal information models, i.e. an application-centric approach to handle the information (Figure 1). This leads to a data-siloed situation where actors, including different departments within a municipality, have problems to share the information. One solution to increase interoperability is to take a data-centric approach. In a data-centric approach, the data is the core (Figure 1); the focus is on the semantics in the information model and data are independent of applications (McComb, 2019). This implies that data can be used by many applications and therefore we denote the datasets used as *generic datasets*. In a data-centric approach the data can be shared among actors, and applications can easily be replaced without influencing the information model. Utilizing a data-centric approach is a requirement to realize an urban digital twin (Batty, 2018; Ketzler et al., 2020).

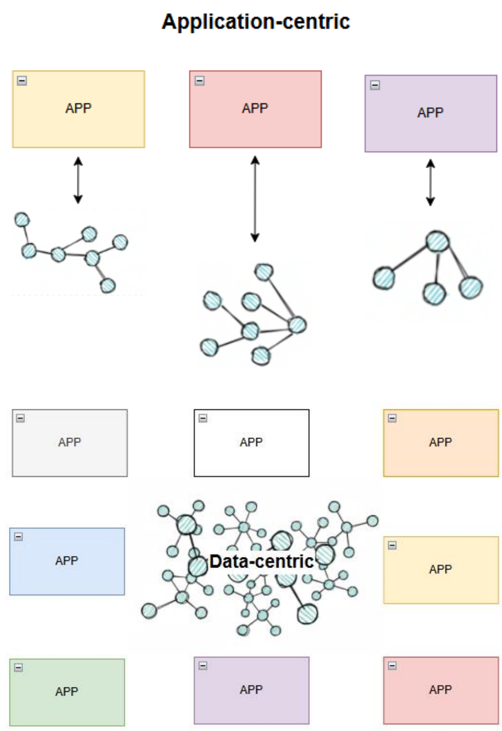

Figure 1. Application-centric (top) and data-centric (bottom) approaches.

A data-centric solution can be implemented using semantic web technologies, and especially knowledge graphs (KGs). A KG is a structured, graph-based data model that represents information as a network of interconnected entities (nodes) and their relationships (edges) (Hogan et al., 2021). Which nodes and edges that can be used is defined in one or several ontologies, where the ontologies must be aligned.

The objective of this study is to investigate to what extent it is possible to create a data-centric solution for road maintenance with the same information content as current application-centric solutions. In the study we work with generic datasets for transport networks and 3D city models. We formulate the following research questions (RQs):
RQ1 – How can the generic datasets for transport networks and 3D city models be linked?
RQ2 – To what extent are the generic datasets for transport networks and 3D city models sufficient for road maintenance applications?
RQ3 – How to construct a knowledge graph for transport networks and 3D city models?

2. Background

This section provides a background to data-centric solutions. It starts by describing some general concepts of data-centric solutions followed by a description of some generic datasets that are of interest for such solutions in a road maintenance context. Finally, a section about knowledge graphs is provided.

2.1 Data-centric Solutions

Data is valuable and is often a key asset in any organization. To ensure that data is not locked into silos and application-centric solutions both researchers and practitioner argue for data-centric solutions (see e.g. The Data-Centric Manifesto[1]). Important principles behind these solutions are that data must follow open information models and must be technically accessible (but not necessarily open for everyone to use). In this way the data-centric approach has much in common with the FAIR principles (Wilkinson et al., 2016).

Data centric solutions are also important in the context of artificial intelligence (AI). Some researchers argue that focusing on the development of AI models is not sufficient; we also need to focus on how to structure the data for these models, i.e., a data-centric AI (DCAI) development (e.g. Jakubik et al., 2024; Jarrahi et al., 2023). The DCAI concept is much broader than how we use the term data-centric in this study; to use data-centric solutions is one of many prerequisites to DCAI as well as for e.g. AI models based on Retrieval-Augmented Generation (RAG) approaches (Lewis et al., 2020).

2.2 Generic Datasets – Transport Networks

The INSPIRE Transport Networks schema (TN) is one of 34 INSPIRE data themes defined under the European INSPIRE Directive (2007/2/EC). It provides a harmonized data model for representing transportation infrastructures across Europe, enabling interoperability between national, regional and local datasets. The purpose of TN is to support applications such as mulitimodel routing, transport planning, digital twins, environmental assessment etc. The TN data consists mainly of transport links and nodes where other objects are connected through linear referencing (according to ISO 19148).

The Swedish national road database NVDB is developed and maintained by the Swedish transport administration in cooperation with the Swedish association of local authorities and regions, the Swedish forest industries federation, the Swedish mapping, cadastral and land registration authority, and the Swedish transport agency. The database is complete for vehicle roads and includes a great extent of the pedestrian and bicycle paths. The NVDB has an open information model that describes how the road network is geometrically and topologically represented using links and nodes, as well as which attributes that are used to describe the roads including e.g. functional road class, road width, surface material, permitted speed, road name and load-bearing capacity. NVDB facilitates export of data according to the INSPIRE-TN specifications.

2.3 Generic Datasets – 3D city models

Cities are creating 3D city models to plan and maintain the urban environment. To ensure that these 3D city models work as generic datasets they should follow on open standard. The most used open standard for 3D city models is CityGML, which is used by cities worldwide. Today two version of CityGML are used: CityGML 2.0 (Gröger et al., 2012) and CityGML 3.0 (Kolbe et al., 2021) which are both Open Geospatial Consortium (OGC) standards. CityGML supports city models with several levels of detail (LOD). Several countries have created national profiles of CityGML to adapt it to national requirements. These profiles are often implemented technically as a CityGML Application Domain Extension (ADE) (Gröger et al., 2012; Biljecki et al., 2018).

CityGML is organised into thematic modules. In the context of this study the most interesting is the *Transportation* module, which enables representation of sematic and geometric transportation features, for example roads, tracks, railways and squares (Kolbe et al., 2021). At LOD1 and higher, transportation objects are described using polygonal objects (multisurface geometries). CityGML 3.0 introduces the object types *section* and *intersection* (Kolbe et al., 2021). A section is defined as a segment of a road object and an intersection is defined as a feature type to model complex areas where two or more transportation objects meet and where traffic flow changes (Figure 2). See Beil and Kolbe (2020) for modelling and application perspectives of the CityGML 3.0 transportation module.

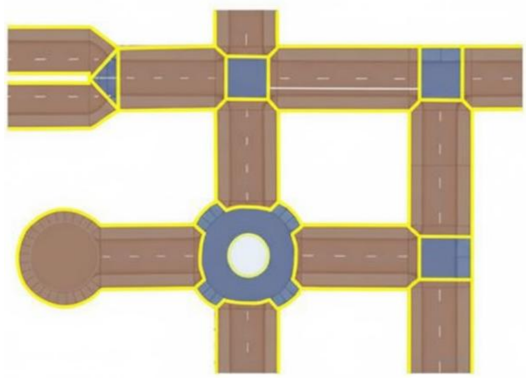

Figure 2. Sections (brown) and different types of intersections (blue) according to CityGML 3.0 (Beil and Kolbe, 2020).

3CIM[2] is a Swedish profile of CityGML 2.0, which was created by the cities of Stockholm, Gothenburg and Malmö in cooperation with Lund University (Uggla et al., 2023). Technically, 3CIM is based on an ADE following the OGC best practices (Van den Brink et al., 2014). For the transportation module the section and intersection object classes are added to the ADE following CityGML 3.0 conceptual model.

[1] https://datacentricmanifesto.org/

[2] https://github.com/3CIM/3CIM2.0

https://doi.org/10.5194/isprs-archives-L-4-W2-2026-147-2026 | © Author(s) 2026. CC BY 4.0 License.

One aspect of 3CIM is that the standard is made relatively thin; 3CIM mainly contains geometric information where the objects are linked to other databases that includes the semantic information. The links to external information is made through the CityGML *Core* module and the feature *CityObject* using the functions *ExternalReference* and *ExternalObjectReference*

2.4 Knowledge Graphs for Transport Networks

A number of initiatives that develop ontologies and knowledge graphs in which transport networks play a central role exist; a selection of these is outlined below.

One prominent example is LinkedGeoData[3], an effort to add a spatial dimension to the Web of Data and the Semantic Web. LinkedGeoData reuses information from the OpenStreetMap project and publishes it as an Resource Description Framework (RDF) knowledge base in accordance with Linked Data principles. It also interlinks this data with other datasets within the Linking Open Data cloud, thereby enabling cross-domain integration.

The INSPIRE initiative is primarily based on a model-driven approach, in which application schemas are specified in UML and implemented as XML/GML schemas. However, the document "Guidelines for the RDF encoding of spatial data[4]" defines principles for representing INSPIRE application schemas in RDF and Web Ontology Language (OWL). However, presently full RDF representations are limited to pilot activities such as ARE3NA/ISA[5], along with selected national experimental conversions and research prototypes. Consequently, INSPIRE-based Linked Data remains largely in a prototypical or transitional stage rather than a widely deployed production ecosystem.

In the railway domain, the European Union Agency for Railways (ERA[6]) is increasingly adopting Linked Data and Semantic Web technologies. In particular, the Register of Infrastructure (RINF[7]) is a centralized database and web application established under the EU railway interoperability directives to describe the static characteristics and capabilities of the European rail network. ERA's adoption of Linked Data principles aims to reduce fragmentation across national railway information systems by enabling a unified, machine-readable knowledge graph. This transition supports improved data sharing, interoperability, and cross-border coordination in European rail transport.

2.5 Knowledge Graphs in 3D city models

So far there is no official ontology for CityGML, but there are some ontologies used. The Knowledge Engineering group at the University of Geneva has developed a direct OWL translation of the CityGML 2.0 XML Schema; this ontology[8] is frequently used in research. Chadzynski et al. (2021) proposes several improvements to this ontology to make it better align with semantic web standards; their updated ontology is coined *OntoCityGML*. More recently, Vinasco-Alvarez et al. (2024) developed the *CityOWL[9]* ontology based on CityGML 3.0 where they also propose alignments to semantic web standards such as GeoSPARQL[10], Simple Knowledge Organization System (SKOS[11]), and OWL-Time[12]. OGC has an ongoing work to define the CityRDF[13] ontology as a serialization of CityGML/CityJSON. One challenge when developing an ontology for CityGML is that the standard ontology for geospatial data GeoSPARQL does not support 3D yet (see Car et al., 2022 for details), even though work in this direction has been proposed[14].

There have been some studies of presenting CityGML in knowledge graphs. Hor et al. (2016) use KG for integration of IFC (BIM) and CityGML data, where they applied the Graph Matching for Ontologies (GMO) algorithm to measure similarities between the two graphs, with the aim of establishing links between corresponding concepts in the graphs. To integrate BIM and GIS data on a query level, Huang et al. (2020) developed an integrated query interface to KGs of IFC and CityGML data. Recent research have studied how CityGML as KGs can provide answers to questions linked to sustainable development (e.g Chadzynski et al., 2023) and act as an external knowledge base for chatbot systems (Liu and Wang, 2025).

To create a KG of CityGML data, linked RDF data must be created from e.g. files. For this purpose, Akin and Cömert (2024) created a converter from CityJSON files (a JSON encoding of CityGML; see Ledoux et al., 2019) to linked data. Another option is to create a virtual knowledge graph (VKG) on top of the native storage. CityGML data are often stored in relational databases (RDB), e.g. using the 3DCityDB schema (Yao et al., 2018). There have been some studies in presenting the RDB data in form of a VKG. One example is Ding et al. (2023) who created a VKG by mapping classes and attributes in the ontology to the RDB schema and then made SPARQL queries to the VKG to showcase the approach. That study was later extended by incorporating other datasets (e.g. OpenStreetMap) into the KG to enable more query types (Ding et al., 2025). Hansson (2024) created a VKG for the CityGML building theme which was evaluated with SPARQL queries in urban planning applications.

3. Methods

In this study we showcase a data-centric solution for road maintenance using KGs. First (Section 3.1) 3CIM/CityGML data for the transportation module is created and linked to the Swedish national road database (NVDB). Then ontologies (Section 3.2) and KGs (Section 3.3) are created. Finally, a data-centric solution is implemented and evaluated (Section 3.4).

3.1 Creating and Linking Generic Datasets

3CIM data for the transportation module was created from the municipality base map and NVDB data with the Extract Transform and Load (ETL) tool Feature Manipulation Engine (FME) from Safe software[15]. The general workflow (Figure 3) was to create polygons around the NVDB road links with a Voronoi diagram approach and clipping the polygons with the

[3] https://linkedgeodata.org/
[4] https://inspire-eu-rdf.github.io/inspire-rdf-guidelines/
[5] https://interoperable-europe.ec.europa.eu/collection/are3na
[6] https://www.era.europa.eu/
[7] https://www.era.europa.eu/domains/registers/rinf_en
[8] https://cui.unige.ch/isi/onto/citygml2.0.owl
[9] https://dataset-dl.liris.cnrs.fr/rdf-owl-urban-data-ontologies/Ontologies/CityOWL/
[10] https://www.ogc.org/standards/geosparql/
[11] https://www.w3.org/2004/02/skos/
[12] https://www.ogc.org/standards/time-ontology-in-owl/
[13] https://github.com/ogcincubator/cityrdf/
[14] https://github.com/opengeospatial/geosparql3dwhitepaper
[15] https://fme.safe.com/

road edges from the municipality base map that consists of line features (Figure 4, top).

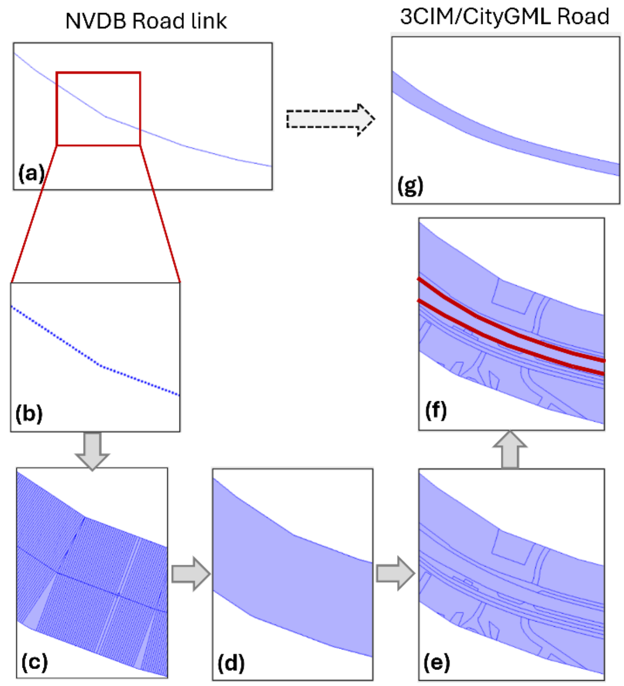

Figure 3. Workflow to create 3CIM/CityGML transportation polygons from a municipality basemap and NDVB road links.

Line features of the object type *road edges* (Swe: *vägutbredning*) in the municipality basemap were selected and a pre-processing step was performed to get topologically connected lines by filling gaps between road edges. The road links in NVDB (Figure 3a) were then converted to point features (one point per vertex; Figure 3b) and Voronoi diagrams created from the NVDB points with edges 30 m perpendicular to the NVDB link (Figure 3c). The Voronoi diagrams were dissolved to create one polygon per NVDB link (Figure 3d) and the polygons were clipped with the road edges from the municipality basemap (Figure 3e). The created road polygons (Figure 3f,g) inherited the attributes from NVDB including the road link id (Swe: *referenslänks-ID*) used as external reference to link the 3CIM data to other datasets.

The procedure to create polygons was performed separately for different types of NVDB links based on the function. In this study polygons with one of the functions *carriageway*, *cycleway*, *walkway* and *access roads* were created. In cases with overlapping classifications, a hierarchy of transportation functions was used to assign a unique class to each surface with carriageway at the highest level and bike/walkway at the lowest level. Areas within the transportation corridor that were not assigned to transportation functions were classified as auxiliary transportation areas. These were identified using geometric characteristics and represent features such as medians, separators, and other non-traffic-bearing components of the road space.

In the next step, the transportation polygons were divided into *sections* and *intersections* according to 3CIM and CityGML 3.0 (Figure 2). Intersections were first identified as NVDB nodes where at least three NVDB links meet. The geometry of an intersection was then created in a similar fashion as the road polygons. The NVDB links connected to the intersection were converted to point features; all points within 15 m from the NDVB node in the intersection were used to create Voronoi diagrams that were dissolved to create intersection polygons. Since an intersection includes several NVDB links, it inherits the attributes from all links meeting in the intersection. Finally, the intersection polygons were used to clip the road polygons to create road sections that do not overlap with the intersections (Figure 4, bottom). For details see Hellsten (2025). The created polygons were imported to the database 3DCityDB with the Importer/Exporter tool.

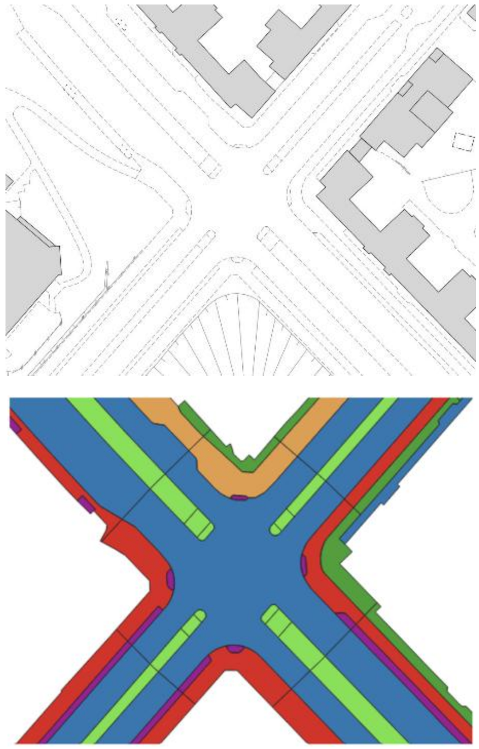

Figure 4. Top is municipality basemap (line features, building footprints as polygons) and below is 3CIM/CityGML data in the transportation module (polygons) with an *intersection* (centre) separated from four connecting *sections* with black lines.

3.2 Creation of Ontologies

The NVDB ontology was automatically generated from the NVDB Feature Catalogue represented in OGC GeoPackage using tools developed by Triona[16]. These tools use C# and the DotNetRDF[17] open-source library to read the NVDB Feature Catalogue and generate the corresponding OWL representation using a set of general conversion rules.

The 3CIM ontology is based on the ontology developed by the University of Geneva[8], which was aligned to GeoSPARQL and extended with 3CIM objects and attributes as well as an alignment to GeoSPARQL.

3.3 Knowledge graphs

For the national road data (NVDB) we created a materialized KG (MKG) using tools developed by Triona and stored it in the triple

[16] https://www.triona.eu/

[17] https://dotnetrdf.org/

store GraphDB[18]. These tools use C# and the DotNetRDF open-source library to read the NVDB dataset and generate the corresponding RDF representation, in accordance to the generated NVDB ontology, using a set of general conversion rules. NVDB uses the TNE[19] product from Triona. TNE uses traditional relational database technology for its native storage so either virtualisation or transformation/materialization is needed to publish the data according to linked data/semantic web principles.

The 3CIM knowledge graph was implemented as a virtual KG. The 3CIM-data was stored in the relational database 3DCityDB and mappings from 3DCityDB to RDF-triplets were created in Protege[20] (ver. 5.6.3) with the “Mapping manager” in the Ontop plugin (ver. 4.2.2). The mappings include an SQL query to retrieve the relevant data (attributes and geometries) from 3DCityDB and the RDF-triplets to populate with the data. The mappings were saved in obda format and Ontop CLI (ver. 5.4.0) was used to create an Ontop SPARQL endpoint (Xiao et al., 2020) that could be accessed online.

3.4 Implementation of an elementary Data-centric Road Maintenance Tool

The Data-centric tool for road maintenance was implemented for the neighbourhood Kista Äng in the city of Stockholm, see Wahlberg (2026) for details. The tool is based on the open-source GIS-program QGIS (ver. 3.34) with the SPARQLing plugin (ver. 0.17.3; Thiery and Homburg, 2026) used to connect QGIS to the KGs (Figure 5). The 3CIM/CityGML data are retrieved from the VKG, the NDVB data from the MKG (Section 3.3) and *ExternalReference* in the 3CIM/CityGML data is used to link the data on object level with a SPARQL query. A Python script was created in QGIS to implement elementary road maintenance functionality; it should be noted that the aim of this study is not to implement the required functionalities but only to investigate that the tool has access to sufficient data for such a tool.

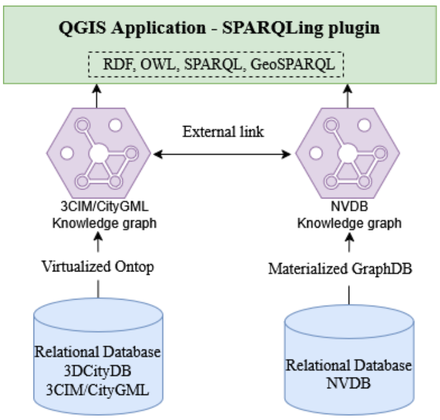

Figure 5. Architecture of the implemented data-centric solution.

Data used for the tool were: (1) 3CIM/CityGML data of the feature types *TrafficArea* and *AuxilliaryTrafficArea* from the 3CIM/CityGML VKG and (2) the attributes *Road authority* (swe: *Väghållare*), *Street type* (swe: *Gatutyp*) and five additional attributes retrieved from the NDVB MKG. These NDVB attributes in combination with the road polygons from 3CIM are sufficient for much of the road maintenance operations. However, for some operations additional data are required, such as attributes for the dry road and winter road classifications; these attributes are stored in the object class *operational maintenance areas* in the current road maintenance tool. Since these attributes are not included in the 3CIM/CityGML or NDVB information models, they were added as attributes to the 3CIM/CityGML polygons in QGIS with overlay operations (intersect). This was performed by the Python script implemented as part of the data-centric tool.

4. Results and Discussion

The implemented data-centric tool successfully combined the two generic datasets 3CIM/CityGML and NDVB from the created KGs with a SPARQL query. This demonstrates that much of the data needed for road maintenance can be obtained from generic dataset. For operations that require data on operational maintenance areas, the data-centric tool does not provide sufficient information even though the classifications for summer and winter roads were added as attributes in the implemented tool (see winter roads in Figure 6). This limitation is partly due to misalignment between the road polygons in the 3CIM/CityGML data and the geometries in the current application-centric road maintenance data, and differences in classifications of road objects in the two datasets; this leads to errors when spatial operations are used to combine the dataset.

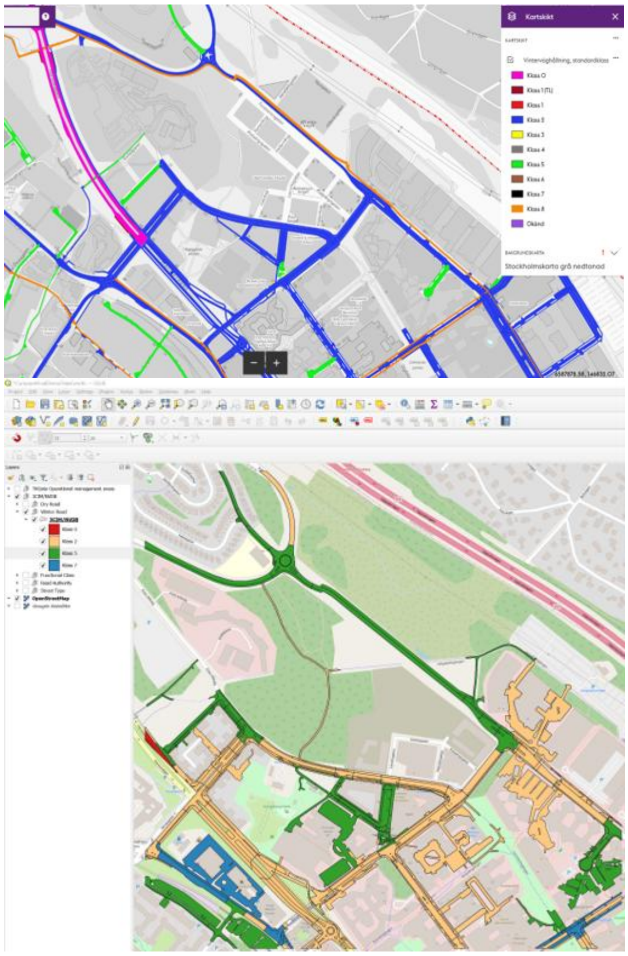

Figure 6. Road maintenance for winter roads. Top is the current application-centric solution and below is the data-centric QGIS solution implemented in this study.

[18] https://graphdb.ontotext.com/
[19] https://tne.triona.se/en/
[20] https://protege.stanford.edu/

To overcome the limitations of the data-centric tool, the information models for the generic data sources need to be extended to include the required attributes or a dedicated information model for road maintenance could be generated. The information model for 3CIM is intentionally kept thin with the aim to keep much of the information near the source of the data (traffic administration office in this case) and should not be extended. Since NVDB is used as the national reference system for road data, while maintenance information for municipal roads is managed locally by the respective road owners (i.e., municipalities), an architecture in which maintenance data is stored locally and linked to NVDB is the most appropriate approach. Hence, we suggest creating a national information model for road maintenance; such information model should only contain the required road maintenance attributes and links to the 3CIM/CityGML on object level to avoid separate descriptions of the road geometries. The development of a national ontology for road maintenance could provide significant value by supporting semantic interoperability and offer a shared conceptual model for municipalities. Figure 7 illustrates a solution based on KGs where the geometries are stored in the 3CIM/CityGML data and road maintenance attributes (here *Winter road* and *Dry road*) are stored in a road maintenance KG in a similar fashion as the NVDB data. This solution would prevent the data-silos in the municipality where different departments to a large extent work with their own datasets, and there would not be any misalignment between road polygons in the 3CIM/CityGML data and the traffic administration data.

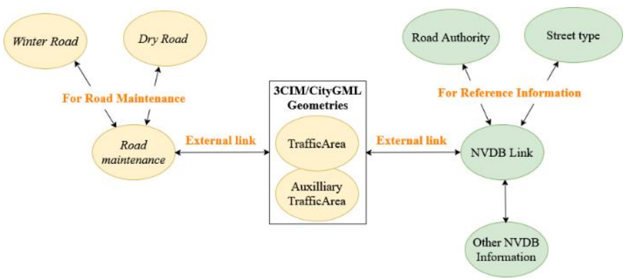

Figure 7. Architecture of a data-centric solution where the road maintenance data are stored in a knowledge graph at the local traffic administration.

5. Conclusions

In this study we show how a knowledge graph approach can be applied to create a data-centric solution for road maintenance applications. The data-centric solution is based on two generic datasets: a national transport network (NVDB) and a 3D city model (3CIM/CityGML). This approach requires that the datasets are linked on object level. To enable that, we utilized Intersection and Section objects in the transportation module of 3CIM/CityGML that were aligned with the road links in NVDB.

Even though the generic datasets used in this study facilitate much of the functionality required for road maintenance some information is missing compared to the current application-centric solution. To overcome this, the information models for the generic data sources need to be extended with additional attribute, or a dedicated information model for road maintenance could be generated. If the latter approach is taken, it is important to keep the information model thin and only include information that is required for road maintenance.

In the study we used knowledge graphs as a technical solution. In general, this approach worked well, even though some of the tools are not completely mature (especially on the client side). To build production system using knowledge graphs would, likely, require an official CityGML ontology.

To conclude. In this study we have showcased an architecture of a data-centric solution for road maintenance applications. There are some additional work required with information missing in the generic datasets and to improve the technical tools. These obstacles are possible to solve which would facilitate solutions to avoid data-silos within the municipality.

Acknowledgements

This study was financed by the project *Interoperabilitet – Digital samverkan för den byggda miljön* (Formas 2023-02575). Data was provided by The Swedish Transport Administration (Trafikverket) and the city of Stockholm. Thanks to Dorothy Furberg and Josefin Isaksson at the Stockholm Traffic Administration Office for helpful discussions.

References

Akın, A.T., Cömert, Ç., 2024. CityJSON2RDF: a converter for producing 3D city knowledge graphs. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII, 15–19.

Batty, M., 2018. Digital twins and smart cities. *Environment and Planning B: Urban Analytics and City Science*, 45(5), 817–820.

Beil, C., Kolbe, T.H., 2020. Combined modelling of multiple transportation infrastructure within 3D city models and its implementation in CityGML 3.0. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VI-4/W1, 29–36.

Biljecki, F., Kumar, K., Nagel, C., 2018. CityGML application domain extension (ADE): overview of developments. *Open Geospatial Data, Software and Standards*, 3(1), 13.

Car, N.J., Homburg, T., 2022. GeoSPARQL 1.1: motivations, details and applications of the decadal update to the most important geospatial LOD standard. *ISPRS International Journal of Geo-Information*, 11(2), 117.

Chadzynski, A., Krdzavac, N., Farazi, F., Lim, M.Q., Li, S., Grisiute, A., Herthogs, P., von Richthofen, A., Cairns, S., Kraft, M., 2021. Semantic 3D city database—an enabler for a dynamic geospatial knowledge graph. *Energy and AI*, 6, 100106.

Chadzynski, A., Li, S., Grišiūtė, A., Chua, J., Hofmeister, M., Yan, J., Tai, H.I., Lloyd, E., Tsai, Y.K., Kraft, M., 2023. Semantic 3D city interfaces—intelligent interactions on dynamic geospatial knowledge graphs. *Data-Centric Engineering*, 4, e20.

Ding, L., Xiao, G., Pano, A., Fan, H., Calvanese, D., & Meng, L., 2023. Querying CityGML Data through Virtual Knowledge Graphs. *Abstr. Int. Cartogr. Assoc.,* 6, 53. https://doi.org/10.5194/ica-abs-6-53-2023

Ding, L., Xiao, G., Pano, A., Fumagalli, M., Chen, D., Feng, Y., Calvanese, D., Fan, H., Meng, L., 2025. Integrating 3D city data through knowledge graphs. *Geo-spatial Information Science*, 28(2), 780–799.

Gröger, G., Kolbe, T.H., Nagel, C., Häfele, K.H. (eds.), 2012. OGC City Geography Markup Language (CityGML) Encoding Standard, Version 2.0. *Open Geospatial Consortium*. http://www.opengeospatial.org/standards/citygml

Hansson, F. (2024). *Linked geodata: CityGML represented as a virtual knowledge graph*. Master thesis, Department of Physical Geography and Ecosystem Science, Lund University, Sweden (in Swedish).

Hellsten, J., 2025. *Automatisk generering av transportytor för 3D-stadsmodeller enligt 3CIM-specifikationen*. Master thesis, Department of Physical Geography and Ecosystem Science, Lund University, Sweden (in Swedish).

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.-C.N., Polleres, A., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A., 2021. Knowledge graphs. *ACM Computing Surveys*, 54(4), 1–37.

Hor, A.H., Jadidi, A., Sohn, G., 2016. BIM–GIS integrated geospatial information model using semantic web and RDF graphs. ISPRS *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-4, 73–79.

Huang, W., Olsson, P.O., Kanters, J., Harrie, L., 2020. Reconciling city models with BIM in knowledge graphs: a feasibility study of data integration for solar energy simulation. ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, VI-4/W1, 93–99.

Jakubik, J., Vössing, M., Kühl, N., Walk, J., Satzger, G., 2024. Data-Centric Artificial Intelligence. *Business & Information Systems Engineering*. https://arxiv.org/abs/2212.11854

Jarrahi, M.H., Memariani, A., Guha, S., 2023. The principles of data-centric AI. *Communications of the ACM*, 66(8), 84–92.

Ketzler, B., Naserentin, V., Latino, F., Zangelidis, C., Thuvander, L., Logg, A., 2020. Digital twins for cities: a state-of-the-art review. *Built Environment*, 46(4), 547–573.

Kolbe, T.H., Kutzner, T., Smyth, C.S., Nagel, C., Roensdorf, C., Heazel, C. (eds.), 2021. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, Version 3.0.0. *Open Geospatial Consortium*. http://www.opengeospatial.org/standards/citygml

Ledoux H., Arroyo Ohori K., Kumar K., Dukai B., Labetski A., Vitalis S., 2019. CityJSON: A compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, 4:4

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D., 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems* (NeurIPS), 33.

Liu, S.Q., Wang, C., 2025. KCityChatBot: a knowledge graph based chatbot system for large-scale CityGML dataset. International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, XLVIII, 99–105.

McComb, D., 2019. *Data-Centric Revolution: Restoring Sanity in Enterprise Information Systems*. 1st ed. Technics Publications, Basking Ridge.

Thiery, F., Homburg, T., 2026. SPARQLing Unicorn QGIS Plugin. Software, Version 0.17.3. https://plugins.qgis.org/plugins/sparqlunicorn/ (15 June 2026).

Uggla, M., Olsson, P., Abdi, B., Axelsson, B., Calvert, M., Christensen, U., Gardevärn, D., Hirsch, G., Jeansson, E., Kadric, Z., Lord, J., Loreman, A., Persson, A., Setterby, O., Sjöberger, M., Stewart, P., Rudenå, A., Ahlström, A., Bauner, M., Hartman, K., Pantazatou, K., Liu, W., Fan, H., Kong, G., Li, H., Harrie, L., 2023. Future Swedish 3D city models—specifications, test data, and evaluation. *ISPRS International Journal of Geo-Information*, 12(2), 47.

Van den Brink, L., Stoter, J., Zlatanova, S., 2014. Modeling an application domain extension of CityGML in UML. *OGC Best Practice*, OGC 12-066.

Vinasco-Alvarez, D., Samuel, J., Servigne, S., Gesquière, G., 2024. Towards an automated transformation of an nD urban data model to a computational ontology network: from UML to OWL, from CityGML 3.0 to CityOWL. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W1, 231–238.

Wahlberg, J., 2026. *Towards Data-Centric Geospatial Solutions: A Knowledge Graph Approach to Road Maintenance in Stockholm*, Sweden. Master thesis, Department of Earth and Environmental Sciences, Lund University, Sweden.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018.

Xiao, G., Lanti, D., Kontchakov, R., Komla-Ebri, S., Güzel-Kalaycı, E., Ding, L., Corman, J., Cogrel, B., Calvanese, D., Botoeva, E., 2020. The Virtual Knowledge Graph System Ontop. In: *The Semantic Web* – ISWC 2020. Springer, Cham.

Yao, Z., Nagel, C., Kunde, F., Hudra, G., Willkomm, P., Donaubauer, A., Adolphi, T., Kolbe, T.H., 2018. 3DCityDB—a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML. *Open Geospatial Data, Software and Standards*, 3(1), 5.