Des Normes à Un Système d'accès Aux Données Basé Sur Une Ontologie Pour Les Réseaux d'eaux Usées et Pluviales

HAYDAR Batoul(1), CHAHINIAN Nanée(1), PASQUIER Claude(2)
(1)HSM , IRD, Univ Montpellier – batoul.haydar@etu.umontpellier.fr
(2)I3S, CNRS – claude.pasquier@univ-cotedazur.fr

Alors que les villes du monde entier connaissent une croissance rapide, la demande d'une infrastructure robuste pour répondre aux besoins changeants des résidents urbains devient de plus en plus importante. La gestion des réseaux d'eau urbains nécessite des données précises et normalisées, ce qui est difficile en raison du manque d'interopérabilité et de l'hétérogénéité des données. Notre recherche consiste à tirer parti du web sémantique en développant un système d'accès aux données basé sur une ontologie qui relie un modèle de données conceptuel à des ensembles de données par le biais d'un ensemble de correspondances.

Mots-clés : Réseaux d'eaux usées et pluviales, accès aux données basé sur l'ontologie, couche conceptuelle, couche de données, correspondance, SPARQL

From Standards to an Ontology-Based Data Access system for Sewer Networks

As cities worldwide experience rapid growth, the demand for a robust infrastructure to support the evolving needs of urban residents becomes increasingly important. Managing urban water networks requires accurate and standardized data which is challenging because of the lack of interoperability and the heterogeneity of data. Our research involves leveraging the semantic web by developing an Ontology-Based Data Access system that links conceptual data models with datasets through a set of mappings.

Keywords : Sewer networks, ontology-based data access, conceptual layer, data layer, mapping, SPARQL query

I INTRODUCTION

With the rapid growth in cities all over the world, the construction and expansion of infrastructure is inevitable. The subsequent operations such as repairs and expansions are made to accommodate the changing needs of the city's residents. The city's sewer networks are critical assets in maintaining essential infrastructure and public health. This constant challenge makes it difficult to accurately map the underground wastewater and stormwater networks, especially in rapidly growing cities.

Multiple institutional and operational actors are involved, from ministries to contracting agents, in implementing, monitoring, and controlling the infrastructure [Bel-Ghaddar et al., 2020]. With each employing their individual documentation system resulting in different representations and formats of data [Wang, 2021], having accurate and standardized data becomes difficult due to the heterogeneity of the data in semantics and structures hindering efficient management. Improving knowledge on the state of these networks is a priority to help managers make informed decisions, optimize resources, and ensure the reliability and safety of the network operations.

Ontology-Based Data Access (OBDA) frameworks offer a potentially effective solution facilitating the integration and data access by leveraging ontologies to provide a unified semantic view of the data [Xiao et al., 2018]. An ontology can serve as a standard wastewater and stormwater domain knowledge model that both humans and machines can interpret [Noy & McGuinness, 2001].

Several attempts have been made to model the domain of wastewater and stormwater [Ceccaroni et al., 2000; Sun et al., 2019; Zeb, 2019] focusing either on utility infrastructures or wastewater treatment approaches. Among the many ontologies that describe infrastructures is the Utility Product Ontology (UPO) [Xu & Cai, 2020] with significant gaps in its coverage particularly regarding important concepts such as the connections between network elements. Similarly, a model structured around city infrastructure ontologies is proposed for sustainable urban planning and maintenance excluding sewer pipes [Du et al., 2023]. Wastewater Ontology (WAWO) integrates the knowledge about wastewater treatment plants (WWTP) with knowledge about aerobic processes and aerobic microorganisms [Ceccaroni et al., 2000] overlooking the network components.

Besides ontologies, INSPIRE Directive[1] aims to create a European Union Spatial Data Infrastructure for EU environmental policies that enable cross-border data sharing and provide interoperable services with its utility and governmental services theme that tackles sewer networks. Géostandard Réseaux d'Adduction d'Eau Potable et d'Assainissement (RAEPA) [2] are French geostandards validated by COVADIS and intended to facilitate data exchanges between stakeholders of public drinking water distribution, collective sanitation and stormwater drainage services.

With the absence of a conceptual data model that enables machines to access and integrate wastewater and stormwater networks data, we propose an ontology-based data access system [Poggi et al., 2008] that comprehensively describes the physical structures of sewer networks leveraging the structured information provided by RAEPA and INSPIRE standard and links the ontology and datasets terms through a set of mappings.

Section II provides an overview of INSPIRE and RAEPA data models. In section III, the proposed framework of the OBDA is presented. Section IV describes that data layer of the OBDA system. The mapping between the ontology and datasets terms is defined in section V. In section VI, the queries are presented. The last section provides the conclusion of the paper.

II EXISTING DATA MODELS

The aim of the INSPIRE Directive is to create a European Union spatial data infrastructure for EU environmental policies, enabling cross border data sharing and providing interoperable services. The Directive covers 34 areas of Spatial Data that are necessary for environmental applications, which include utilities and government services concerned with sewer networks. Utility networks, or node-link-node structured networks for the collection, transmission, and distribution of sewer and other networks, is one of the subthemes under the main theme of Utility and Government Services. UtilityLinkSet, UtilityNode, and UtilityNetwork are among the feature types included in the Utility Networks Profile that are comparable to RAEPA. UtilityNodeContainer is used to describe concrete structures and Appurtenance as subtype. The standards provide attributes, constraints and association role to describe the different networks. Examples include pipeDiameter, utilityDeliveryType, authorityRole and geometry with definition and value type.

Furthermore, validated by COVADIS, the French geostandard Réseaux d'Adduction d'Eau Potable et d'Assainissement (RAEPA) aims to simplify data transfers between stakeholders of collective sanitation, public drinking water distribution, and stormwater drainage services. The Utility and Government Services theme requirements in INSPIRE contain the thematic category that the data processed herein belong to. Similarly, it offers descriptions of four primary classes—pipes, structures, appurtenances, and repairs—along with several additional properties pertaining to network managers, data sources, coordinates, and other details.

III ONTOLOGY-BASED DATA ACCESS FRAMEWORK (OBDA)

The OBDA system consists of three parts: the conceptual layer, the data layer and the mappings. Ontology Web Language (OWL) conceptual data model represents the conceptual layer describing the domain of wastewater and stormwater networks. OWL2 QL (Ontology Web Language Query Logic), standardized by the W3C (World Wide Web Consortium), is tailored for OBDA systems and is based on the DL-Lite family of description logics designed to enable the querying of data stored in a typical relational database by query rewriting [Hitzler et al., 2012]. The data layer includes the different sources stored in relational databases independent of the conceptual layer. The terms of the conceptual layer are linked to the data sources enabling querying using the ontology's semantics. The conceptual data model is developed using Protégé [3] (5.6.1), an ontology editor guided by the Simple Knowledge-

[1] https://knowledge-base.inspire.ec.europa.eu
[2] https://www.geoinformations.developpement-durable.gouv.fr/covadis-r425.html
[3] https://protege.stanford.edu/

Engineering Methodology developed by the Medicine School of Stanford University [Noy & McGuinness, 2001]. The domain ontology is represented using OWL 2 QL.

Figure 1. Simple knowledge engineering methodology

III.1 Simple Knowledge-Engineering Methodology

The Simple Knowledge-Engineering Methodology is an iterative process where the conceptual data model will be revised and refined during the different development phases. Figure 1 lists the 7 steps to build an ontology using the simple knowledge engineering methodology. The first step includes listing competency questions (CQs) reviewed by domain experts to determine the boundaries and depth of coverage of the application ontology. In the second step, RAEPA (v1.2) geostandard and INSPIRE directive are used to describe the physical structures of the network. The third step provides a list of terminology from the data models and domain experts to include in the application ontology. In the fourth step, the hierarchical taxonomy is created based on the terms defined before. The object and data properties of the classes are defined in the fifth step where *geo:asWKT* data property is imported from Open Geospatial Consortium (OGC) GeoSPARQL ontology [Battle & Kolas, 2011] and linked to the class *NetworkElement* with data type *wktLiteral* (well-known text representation) literal. In sixth step a limited set of restrictions on classes is described to comply with OWL2 QL such as class expressions disjointness and domain and range of a slot. The last step of the development process is to populate the application ontology. Instead of creating an RDF graph, we linked the developed conceptual data model to a database and created a mapping between the its terms and the dataset creating a virtual knowledge graph.

The application ontology is then evaluated to verify its correctness and assess its coverage. During the construction phase, HermiT, a freely available and open-source reasoner engine, is used to identify conflicts. To check if the ontology meets the proposed requirements, we employ the competence questions defined in the first step of the development process of the ontology. The process includes translating these CQs validated by domain experts, into SPARQL queries to retrieve relevant information from the ontology and data sources. The developed ontology is accurate, comprehensive, and inclusive in representing the entire domain.

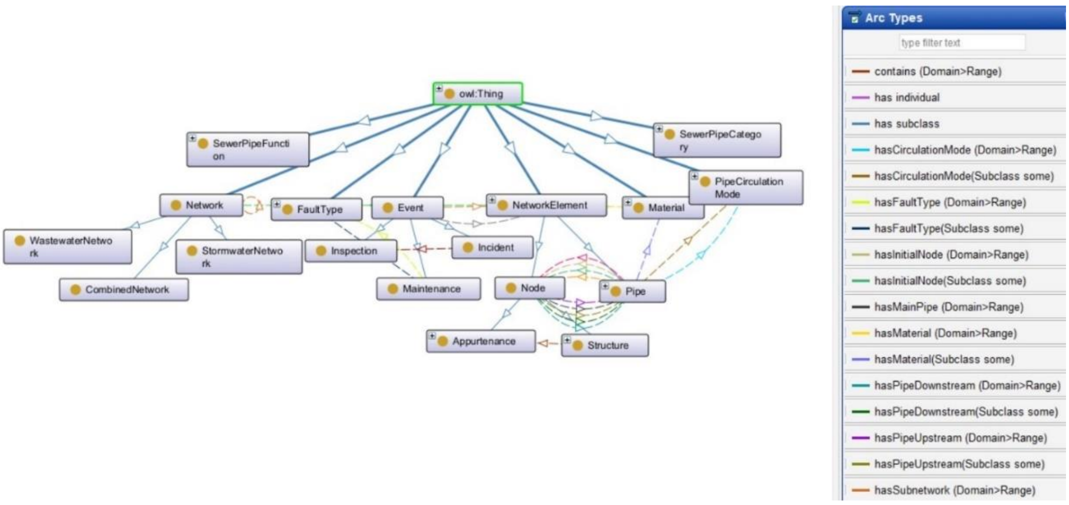

Figure 2. Ontology hierarchical structure with the arcs representing the propoerties

Figure 2 shows the developed ontology hierarchy with the arc types between different classes. The solid blue lines represent the subsumption relation of classes. The bold lines are the direct subclasses of the owl:Thing. Other arcs represent the object properties linking a domain class with a range class. For example, the yellow arc *hasInitialNode* directed toward *Node* class has a Domain: *Pipe* and Range: *Node*. The arcs defined with (Subclass some) represent the existential quantification where there must exist at least one instance of a particular class that has a specific property. For example, the green arc *hasInitialNode* signifies every pipe in the ontology must be connected to at least one initial node.

IV DATA LAYER

Due to the spatial nature of the wastewater and stormwater network data, leveraging PostGIS within PostgreSQL for storing the data sources provides a powerful tool to utilize its spatial capabilities for organizing and querying geographic features. Open Data of Montpellier Méditerranée Métropole provides information on the sanitation networks of the Montpellier Metropole [4] and is divided into several data layers pipe, manhole, appurtenance, storm overflow, structures, pumping stations and treatment stations. We set up the spatial database and import the data into the database. We load the data using QGIS DbManager which allows the connection to various kinds of databases, including PostGIS. After the configuration of the PostGIS Database connection, we import the vector layers which are the different data layers of the networks. Then we set up the database connection in Protégé.

V MAPPING

The mappings redefine the semantics of the data based on the ontology [Poggi et al., 2008], where the terms of the conceptual layer are linked to the data sources enabling querying using the ontology's semantics. Each mapping assertion relates an SQL query over the database to a concept or role of the ontology [Botoeva et al., 2015]. They are formalized in R2RML (RDB to RDF mapping language) W3C standard to express the mappings from relational databases to RDF datasets [Das et al., 2011].

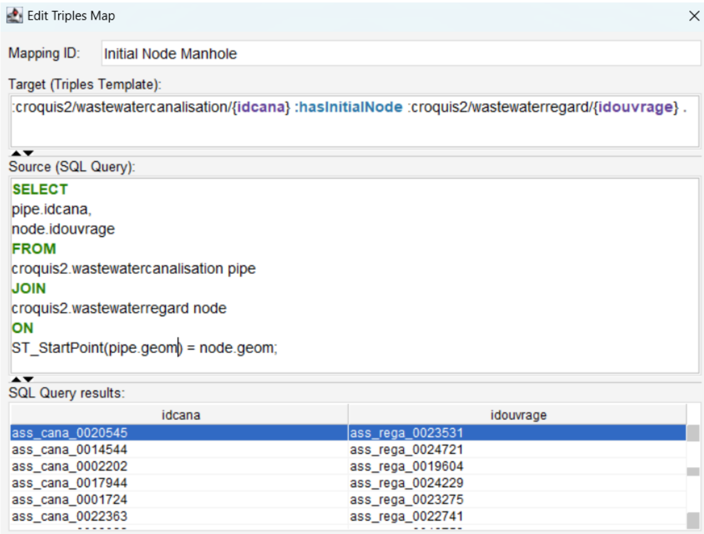

Figure 3. Mapping of the Manhole the initial node of a pipe

Ontop [5] plugin in Protégé is used as an OBDA system allowing the creation of the mappings and the connection to the database as well as SPARQL querying. The mapping manager consists of *Target (Triples Template)* that links the terms in the database to the terms in the ontology and the corresponding *Source (SQL Query)*. An example of the mapping of the object property *hasInitialNode* the links the Pipe with Manhole is shown in Figure 3 where the spatial function from PostGIS ST_StartPoint is used in the query to retrieve pairs of identifiers where the starting point of a

[4] https://data.montpellier3m.fr/dataset/reseaux-dassainissement-de-montpellier-mediterranee-metropole
[5] https://ontop-vkg.org/

wastewater pipe is spatially equal to a manhole. This is true due to the fact that the flow direction in the files is digitized from the initial node (starting point) to the terminal node (ending point) of a pipe.

VI QUERYING

The OBDA system is used for querying to ensure the coverage of the developed conceptual data model by translating the CQs into SPARQL queries. In Table 1 , we provide few examples of CQs and their SPARQL queries translation. Using Ontop SPARQL in Protége, the queries are reformulated into SQL queries and executed by the database. The results returned can be exported to CSV. GeoSPARQL queries can be executed as well to retrieve geospatial data using geospatial functions and properties such as WKT serialization and topological simple features relation.

Table 1. CQs and their corresponding SPARQL queries

Competency Question	SPARQL Query
1. What are the material and diameter of each pipe in the network?	SELECT ?pipe ?diameter ?material WHERE { ?pipe a :Pipe ; :hasDiameter ?diameter ; :hasMaterial ?material . }
2. What pipes are Forced?	SELECT ?pipe ?circulationMode WHERE { ?pipe a :Pipe ; :hasCirculationMode ?circulationMode . ?circulationMode a :ForceMain . }
3. How are different components of the wastewater and stormwater networks connected?	SELECT ?pipe ?initialNode ?terminalNode WHERE { ?pipe a :Pipe ; :hasInitialNode ?initialNode ; :hasTerminalNode ?terminalNode . }

VII CONCLUSION

The developed framework provides a unified semantic schema for sewer networks to access data and address issues of data heterogeneity. In this framework, we propose an ontology-based data access system allowing data on the sewer networks to be accessed and queried through the developed conceptual data model in Protégé. It is based on RAEPA and INSPIRE data models and validated by domain experts. A mapping is created between the ontology terms and the data sources of Montpellier Métropole using Ontop.

VIII ACKNOWLEDGEMENT

This research has received support from the European Union's Horizon research and innovation programme under the MSCA-SE (Marie Skłodowska-Curie Actions Staff Exchange) grant agreement 101086252; Call: HORIZON-MSCA-2021-SE-01; Project title: STARWARS (STormwAteR and WastewAteR networkS heterogeneous data AI-driven management).

This research has also received support from the ANR CROQUIS (Collecte, représentation, complétion, fusion et interrogation de données de réseaux d'eau urbains hétérogènes et incertaines) project, grant ANR-21-CE23-0004 of the French research funding agency Agence Nationale de la Recherche (ANR).

IX REFERENCES

Battle, R., & Kolas, D. (2011). Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, *3*(4), 355–370.

Bel-Ghaddar, Y., Seriai, A., Begdouri, A., Delenne, C., Chahinian, N., & Derras, M. (2020). Combining model-driven engineering and sewerage networks: Towards a generic representation. *Colloquium in Information Science and Technology, CIST*, *2020-June*, 48–53. https://doi.org/10.1109/CiSt49399.2021.9357171

Botoeva, E., Calvanese, D., Santarelli, V., Savo, D. F., Solimando, A., & Xiao, G. (2015). Beyond OWL 2 QL in OBDA: Rewritings and Approximations. *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:6374600

Ceccaroni, L., Cortés García, C. U., & Sànchez-Marrè, M. (2000). *WaWO: an ontology embedded into an environmental decision-support system for wastewater treatment plantmanagement*.

Das, S., Sundara, S., & Cyganiak, R. (2011). R2RML: RDB to RDF mapping language. *Http://Www. W3. Org/TR/R2rml/*.

Du, H., Wei, L., Dimitrova, V., Magee, D., Clarke, B., Collins, R., Entwisle, D., Eskandari Torbaghan, M., Curioni, G., Stirling, R., Reeves, H., & Cohn, A. G. (2023). City infrastructure ontologies. *Computers, Environment and Urban Systems*, *104*, 101991. https://doi.org/10.1016/J.COMPENVURBSYS.2023.101991

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012). *OWL 2 Web Ontology Language Primer (Second Edition)*. https://api.semanticscholar.org/CorpusID:57542951

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05 and ….

Poggi, A., Lembo, D., Calvanese, D., Giacomo, G. De, Lenzerini, M., & Rosati, R. (2008). Linking Data to Ontologies. *J. Data Semant.*, *10*, 133–173. https://api.semanticscholar.org/CorpusID:1325494

Sun, Y., Wang, Z., & Wang, K. (2019). Ontology-Based Event Detection for Wastewater Treatment. *ISWC (Satellites)*, 97–100.

Wang, M. (2021). Ontology-based modelling of lifecycle underground utility information to support operation and maintenance. *Automation in Construction*, *132*, 103933. https://doi.org/10.1016/J.AUTCON.2021.103933

Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R., & Zakharyaschev, M. (2018). Ontology-Based Data Access: A Survey. *International Joint Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:44064877

Xu, X., & Cai, H. (2020). Semantic approach to compliance checking of underground utilities. *Automation in Construction*, *109*, 103006. https://doi.org/10.1016/J.AUTCON.2019.103006

Zeb, J. (2019). An ontology of condition assessment technologies for sewer networks. *Infrastructure Asset Management*, *7*(1), 36–45.