



HAL
open science

Combining model-driven engineering and sewerage networks : towards a generic representation

Yassine Belghaddar, Abderrahmane Seriai, Ahlame Begdouri, Carole Delenne,
Nanée Chahinian, Mustapha Derras

► To cite this version:

Yassine Belghaddar, Abderrahmane Seriai, Ahlame Begdouri, Carole Delenne, Nanée Chahinian, et al.. Combining model-driven engineering and sewerage networks : towards a generic representation. CiST'20 - 6th IEEE Congress on Information Science and Technology, Dec 2020, Agadir - Essaouira (postponed to June 2021), Morocco. hal-03108996

HAL Id: hal-03108996

<https://hal.science/hal-03108996v1>

Submitted on 28 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining model-driven engineering and sewerage networks: towards a generic representation

Yassine Bel-ghaddar^{1,2,3,4} Abderrahmane Seriai¹
Ahleme Begdouri³ Carole Delenne^{2,4} Nanee Chahinian²
Mustapha Derras¹

¹*Berger-Levrault*, Prols, France

²*HSM, Univ. Montpellier, CNRS, IRD*, Montpellier, France

³*LSIA, Univ. Sidi Mohamed Ben Abdellah*, Fez, Morocco

⁴*Inria Lemon, CRISAM Inria*, Sophia Antipolis Mditerrane, France

Presented to the 6th IEEE 2020 Congress on Information Science and Technology

Abstract

Representing and processing digital data related to underground networks, particularly sewerage networks, is increasingly becoming a priority for the managers of these networks. Indeed better representation would allow them, among others, to improve knowledge and to take the best decisions regarding these generally poorly identified infrastructures. The heterogeneity of data and the multiplicity of data models representing sewerage networks, often specific to each operator, as well as the imperfections associated with both the available data and those collected from different sources, generate complexity in terms of on-the-field interventions efficiency. They also highlight the need for aggregation (unification), control and analysis. The main objective of our work is to merge multi-source data to obtain more precise and complete digital maps of sewerage networks. In this paper, we propose a generic data modelling for data fusion purposes taking into consideration the uncertainty aspects related to the collected data by allowing a confidence value for each data source and for each single data provided by a source.

Keywords Meta-modeling, Sewerage Network, Model-driven engineering, Data sources

1 Introduction

By the time a sewerage network is set up, its graph, where the nodes represent the manholes or inlet grates and the edges represent the pipes, is mapped for the

first time. Later on, several operations such as reparations and expansions may occur in the field to respond to the scalable needs of the citizens. These actions are generally internally reported in several formats, representations and through models, which are often owned by the operators working on the network. Consequently, the combination of data from different sources and eras raises problems of consistency (contradictory information) and requires the establishment of a methodological framework for collecting, centralizing, updating and data archiving in order to facilitate information sharing and communication between the managers. In our vision of data fusion, we consider the uncertainties related to each type of collected information in order, for example, to anticipate and react promptly to potential dysfunctions or to quantify their impact on the results of a numerical simulation of flows in the sewerage network. In this context, and as a first step of our work, we propose in this paper a meta-model for aggregation, control and analysis of data sources related to sewerage networks, before elaborating adapted algorithms to merge heterogeneous multi-sources data. The paper is structured as follows, section 2 introduces the context of our work. We explain the motivation behind this work in section 3 and we present the state of the art and related works in section 4. Section 5 describes our meta-model and its specific viewpoints. Section 6 concludes this paper.

2 Context

2.1 Sewerage networks and management challenges

Sewerage system is a network for collecting and transporting wastewater and storm water to a treatment plant, also called combined sewer system. When a network collects these two types separately, it is called separated sewer system. To set up a sewerage system and make its progress in a territory, different institutional and operational actors are involved. The ministry in charge of sanitation or the ministry in charge of local communities, define sanitation policies and strategies as well as the regulatory framework at national level. The local authorities ensure the respect of regulations related to the quality of sanitation services. Finally, the contracting authorities (municipalities or state agencies) are responsible for the development of services, their quality and sustainability. For the implementation, monitoring and control of these services, they call for actors such as service operators, local associations, and development partners (founders, NGOs, and design offices). The sanitation supply is not only limited to infrastructure installation. Maintenance tasks such as reparation, expansion, damage anticipation and scheduling of interventions are all necessary actions to ensure a permanent and transparent services. Planning is an important task for decision making. It allows to develop a vision of needs in space and time, to quantify and prioritize them in order, among others, to direct funding towards the most necessary investments and at reasonable costs.

On a territory, infrastructures are large and must be managed in a collaborative way by the diverse involved actors. Urbanization and the concentration

of populations in cities engender the increase of the dimensions of sewerage networks. For example, in France, this heritage consists of approximately 337,000 km of collectors [ASTEE2015]. The 2012 reform "DT-DICT" as part of the network detection process indicates that France is covered by more than four million kilometres of networks, one third of which are aerial and two thirds are buried or underwater.

The improvement of underground networks, particularly for sewerage networks, has several advantages, especially the impact on public health and environment preservation through the protection of water resources against pollution. The administrative management and the techniques of interventions play a key role in these challenges since they help to reduce damage costs induced from services interruptions and floods.

The expenses associated with the management of these infrastructures are high, particularly repairs, since the components of these networks are subject to degradation and damage caused by several factors: age, environment, etc. Furthermore, the costs of urgent and unexpected operations are far higher than the ones anticipated [ASTEE2015].

In this context, improving knowledge on the state of these networks, mostly underknown, becomes a priority. Indeed, digital technologies such as Geographic Information Systems (GIS) and Computerised maintenance management systems (CMMS) bring great added value and are officially increasingly adopted. For example, the regulations associated to the environment code in France require stakeholders to have digital and precise cartography for sensitive underground networks since January 1st, 2019 in urban units and from January 1st, 2026 in other cases.

These solutions are particularly useful for making spatial and geographic data available to the various actors, to facilitate their communication for optimal decision-making as well as to improve the administrative, economic, and financial management of this type of networks and of interventions to take place near these underground networks.

2.2 Sewerage network representation

A sewerage network is represented by a graph composed of nodes and edges. Nodes represent manholes, equipment, repairs, etc. the edges represent pipes. Each of the nodes and edges has a set of properties in the form of attributes such as, diameters of the pipes, types of materials and positions of the inspection areas for the objects. In recent years, storage and data management solutions for sewerage networks have evolved. Currently, most managers use Geographic Information Systems to create, edit, view, and analyze these data. The data structures and formats supported by these systems are diverse: relational databases, Shapefiles, GeoJSON and CSV files, etc. Moreover, to study the impact of some parameters, such as the discharge rate of consumers into the networks, specialists use hydraulic simulation software.

Although the applications are various, the digital representation of the data remains almost identical in the different solutions:

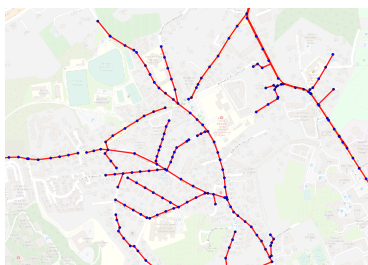


Figure 1: Example of spatial representation of sewerage networks.

	NOM	CODE ROSEE	OBSERV	DATE MAJ	EST KEY	RUE	ENJE ROSEE
1		10113		4/1/2006		51120	GESPER
2		10113		4/1/2006		51130	GESPER
3		10175		11/8/2005		50995	
4	Rapport_EU	10109				51112	
5	R172	10106				51000	
6		10113		28/7/2009		51007	GESPER
7		10111		28/9/2005		51001	SAT
8		10109				51005 AV DE LA MAIRIE	

Figure 2: Example of attributes table.

- Spatial data that are represented by geometric shapes and their relationships: points and lines (figure 1).
- Attributes that are listed in attribute tables where each record is associated with a network object (figure 2).

3 Problematic

When using GIS solutions, data processing includes acquisition, digitization, import, export, and visualization of geographical data. The acquisition can be carried out directly in the field allowing real time data collection and mapping. In addition, advanced data processing may allow considering multi-source data, spatial analysis through interactive queries and maps overlays.

Although the use of digital maps is increasingly adopted, there are still large communities in the world where maps and geographic data are still analog, making their use and update difficult. The detection and digital mapping of buried networks by semi-automatic or automatic approaches is a real scientific and technological challenge. Therefore, there is a large conceptual, technical and semantic gap between the analog and digital mapping models.

The attributes and characteristics associated with the various objects constituting a network are not all available at a given time. This is partly explained by the fact that the networks undergo expansions and repairs but not properly tracked and reported, or through the interventions at different stages by actors, other than the operators who ensure the continuous functioning of the supply

services. However, these attributes may be reported elsewhere, for example, in public databases, calls of tender, repair reports or even in press articles reporting damages.

In addition, since information and communications technology are easy to reach and use, operators currently have access to several sources from which they can collect useful data before interventions in the field, such as images, analogue maps, reports of interventions, sensors, etc. The heterogeneity of the sources makes the extraction of relevant information and its combination a complex and time-consuming task.

On the other hand, imperfections may be found in these data sets and sources, namely inconsistency (abandoned pipelines which still appear on the maps), missing attribute values for some objects, uncertain and sometimes contradictory values. All of these aspects represent various obstacles to operators when merging the data.

Combining multi-source data also requires a unified data model to allow the centralization, updating, archiving, and monitoring of these data. Indeed, we have analyzed digital databases related to sewerage networks to understand the semantics of their data, their relationships and determine their differences. Since the associated data models, when they exist, are rarely available to the public, we proceeded by inferring them from data. In our study, we have used the data provided by reliable sources, particularly, the French open data repository [ope2015]. Among the suppliers are the urban community of the South-East of Toulouse Sicoval, Data Angers, and the region of Pays de la Loire.

As a result of studying these different sources, we identified the following constraints:

- The data models adopted by operators are different. Therefore, exchanging and reusing data is difficult.
- The models do not comply with computer design and modelling rules and standards.
- The attributes provided by the stakeholders are related to their fields of activity. For example, a company specialized in hydraulic modelling provides precise information on the flow of water in a pipeline, while this same information is generally missing in the data coming from another entity expert in the field of structures repairing.
- The history of interventions, necessary for anticipating repairs, is rarely considered in these models.

Thus, a generic model for business data and data sources is needed to overcome the conceptual and semantic gap between existing digital mapping models and to implement an optimal data fusion approach.

4 State of art

4.1 Meta-models

Several definitions have been proposed. According to [Object Management Group2006], a meta-model is a model used to model modelling itself. A Meta-model is a model that defines the structure of a modelling language [Da Silva2015]. The basic idea of a meta-model is to identify the general concepts in a given problem domain and the relations used to describe models [Gascueña et al.2012]. This generality, which we also seek to satisfy in our proposal, is one of the most important axis that have made the meta-modelling, i.e. the creation of meta-models, one of the most important approaches for modelling. Instances of a meta-model are models that must satisfy the meta-model specifications. They enable target systems to be modelled in a consistent and homogeneous manner.

Monitoring activities is one of the concepts where meta-models are used. For example, a meta-model for properties associated with software during execution is presented in [Bertolino et al.2011] to ensure the quality of software and its dynamic adaptation after deployment. Indeed, these properties provide a means to assess and improve the resilience of software through the adaptation and anticipation of abnormal situations. The instances of this meta-model are in this case a model with monitoring properties adapted to the target software. In [La Fosse et al.2020], the authors propose a meta-model for the monitoring of cyber-physical systems (CPS), particularly sensor and actuator networks which require valid data and good coordination between sensors and actuators for their operation.

4.2 Sewerage networks business modelling: related works

To help decision makers in collecting the data necessary for interventions and to diversify their data sources, some solutions have been published. For example: in [Commandre et al.2017], the authors apply deep neural networks to detect the position of manholes, visible on the ground, from a high-resolution image. In [Chen and Cohn2011], to create the cartography of underground networks, researchers use Bayesian fusion techniques to combine hypotheses extracted from the Ground Penetration Radar (GPR), the spatial location of surveyed manholes, as well as the expectations from the statutory records. In [Chahinian et al.2016], to assess whether big data can be used to reduce uncertainty in storm water modelling, the authors propose a process based on text analysis for extracting attributes about the objects of sewerage networks from data available on the web. However, none of these approaches have been submitted along with a data model.

The work in [Abdelbaki and Zerouali2012] is an attempt to design a business data model for sewerage networks to build the digital map of the sanitation network for a municipality in Algeria and to contribute to its efficient management. The authors propose, from the inventory of the various available data sources, a conceptual data model on which the necessary objects for the management

and their relationships are listed.

At the initiative of the Aquitaine region and a public interest group (Planning and risk management), the Commission of Data Validation for Spatialized Information (COVADIS) has published a data standard for drinking water and sewerage networks intended for French municipalities [COVADIS2019]. The committee presented a class diagram describing the minimum and necessary data to be used by the actors participating in the management of these networks (municipalities, PEIC¹, delegates public services, etc.) for the purpose of a simple data exchange between them.

Since this standard describes the minimum necessary, but sufficient, data for the management of the water networks, and since it is also a standard to be adopted at a nationwide level, we adopt it in our work as a business data model. We present in the following the subpart of the COVADIS class diagram related to the sewerage networks (Figure 3).

It is composed of 4 main classes: Node, Pipeline, Reparation and Meta-data whose attributes and related possible values are listed:

- Nodes: represented geometrically by points, they illustrate apparatus (valve, counter etc.) or manholes.
- Pipelines: represented geometrically by lines, they are classified into several categories: wastewater, rainwater etc. Each of the pipes has two end Nodes.
- Repairs: geometrically represented by points, they refer to interventions made in Nodes or Pipes.
- Metadata: are data used to qualify the information of the classes Nodes and Pipes. Namely the name of the source, the date of the last update, the reliability of the year of installation and the quality of the geolocation within respect to the 2012 decree [Leg2012], which defines 3 precision classes: less than 40mm, in the range 40mm and 1.5m and greater than 1.50.

4.3 Sewerage networks and Big Data

Nowadays, we are witnessing an intensive production of data. Every day, a huge amount of data is produced by companies, on social networks, during transactions or through sensors, that conventional computer tools can no longer process and analyse. The research work around these masses of data, also called Big Data, is a response to these obstacles.

On the other hand, and despite the tiny amount of the available data on sewerage networks compared to the Big Data, they share two important characteristics:

- The multitude of data sources.

¹Public Establishment for Inter-municipal Cooperation.

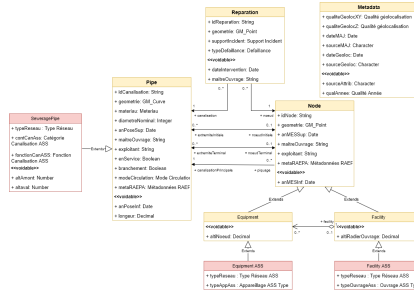


Figure 3: COVADIS sewerage networks business model [COVADIS2019].

- The heterogeneity of the data.

The research works in the domain of the underground networks is mainly confidential [Hafsi et al.2017]. Therefore, the number of publications related to Big Data is more important compared to sewerage networks. To our knowledge, there is no data model for data sources of underground networks. To fill this gap, and since these two characters of the multitude and heterogeneity of the sources have already been examined in Big Data (see for example [Dong and Srivastava2013] or [Boury-Brisset2013]). We have chosen to draw inspiration from the solutions proposed in this field.

The available Big Data systems and platforms are not identical, as they come from multiple providers whose vision is not uniform. In [Erraissi et al.2018] using model-driven engineering, the authors propose a platform-independent meta-model to describe the structures of data sources involved in feeding these large volumes of data, thus allowing programmers to create applications compatible with various products. Figure 4 illustrates this proposition which divides the data sources into 3 broad categories:

- Structured: data whose set of possible values are determined and known in advance, such as relational databases.
- Semi-structured: data that have not been organized into a specialized repository. However, they contain meta-data information, which help their exploitation, for example e-mails.
- Unstructured: data represented or stored without a predefined format.

5 contribution

5.1 Meta-model for sewerage networks data sources

Our target is to propose a generic model for data sources, with the aim of using fusion approaches to combine their data. Therefore, on the one hand,

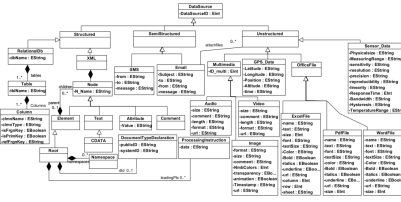


Figure 4: A meta-model for Big data data sources [Erraissi et al.2018]

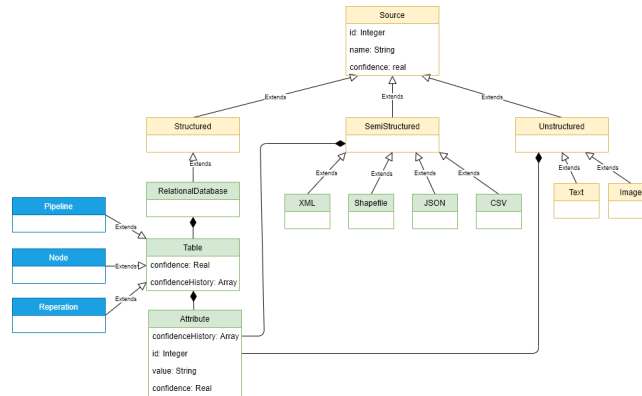


Figure 5: Meta-model for sewerage networks data sources.

our solution should encompass the available approaches for collecting data. On the other hand, the data sources are diverse and may change over time. An exhaustive modelling of data sources and their possible relationships is not a generic solution. Figure 5 illustrates our meta-model.

For a better understanding, four viewpoints of this meta-model are presented in the following paragraphs.

5.2 Data sources viewpoint

We summarise in Figure 6 the data sources viewpoint where the main entity source characterises any entity capable of providing data, information or knowledge about sewerage networks. Interpretation of the structuring aspect of data sources is as follows:

- Unstructured sources: whose formats require significant pre-processing before extracting business data about a network. For example, the location of sewer manholes from images.
- SemiStructured sources: whose formats require simple pre-processing before extracting business data about a network. For example, parsing CSV

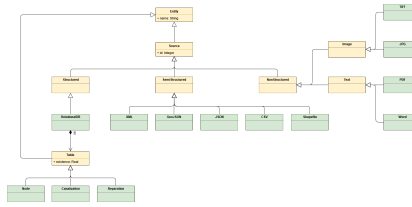


Figure 6: Data sources viewpoint.

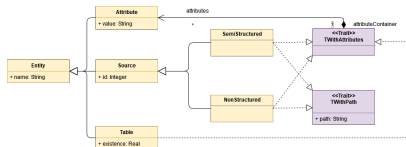


Figure 7: Attributes viewpoint.

or XML files.

- Structured sources: represent relational databases that directly provide business data about a network. Generally, these data are provided by the official managers of sewerage networks.

5.3 Attributes viewpoint

The attributes and their relationships with data sources are highlighted in this viewpoint (Figure 7). Each attribute is an entity identified by a name and possesses a String value representing a semantic data. The aggregation of attributes by data sources is encapsulated within the "TWithAttributes" entity. Since it represents, in the case of structured sources, the different attributes within the table of a relational database. Thus, the entity "Table" is connected to this entity. As for semi-structured and unstructured sources, it includes the pre-processing operations, defined by instances of this meta-model, to extract attribute values about sewerage networks. Moreover, data source path is handled within "TWithPath" entity.

5.4 Confidences viewpoint

We modelled data imperfections in this viewpoint (Figure 8) by allowing confidence attributes to each source, table (representing an object of the sewerage network) and to each attribute characterizing this table (the object). This means that:

- Each source has a confidence value that indicates the reliability or the certainty of the information it provides. This metric can be modelled by

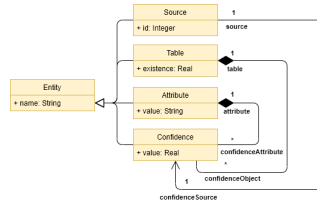


Figure 8: Confidences viewpoint.

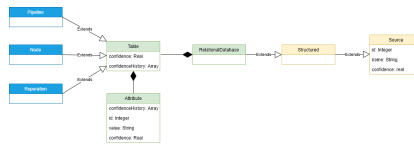


Figure 9: Business model viewpoint

the various available mathematical tools, such as probabilities.

- For the components or objects of sewerage networks, this value represents the uncertainty regarding their existence. Indeed, it is possible, for example, for a pipe or a manhole to be represented on a map by mistake.
- As for attributes, the confidence is related to the confidence of the data sources providing them.

Moreover, objects and attributes default confidence values are those associated with their sources. However, this does not imply not having different confidence values later on. For example, an approach identifying manhole covers from images would define the existence confidence of the detected object as the precision of its detection. Meanwhile, the attribute position of the detected object may have a different value, since the detection and the computation of its position are two separate operations. To keep track of the previous data fusion operations, the confidence history is stored too.

5.5 Business model viewpoint

In our meta-modelling, we distinguish Business data, characterizing the sewerage networks components, from data about the sources, from which business data are extracted, such as images, documents, calls for tenders, etc.

Figure 9, illustrates this business viewpoint where we adopted the COVADIS (4.2) standard classes that inherit the properties of the Table entity. For genericity purposes, other business models may easily replace it, provided that the appropriate connections are respected.

6 Conclusion

In this work, we proposed a meta-model for data sources related to sewerage networks inspired from the field of Big Data. Our proposition considered the three important aspects of i) structuration of the data sources, ii) associated confidences and iii) genericity related to the business domain. Currently, we are implementing our meta-model in Moose, a platform of software and data analysis, in order to identify a use case along with a suitable workflow to instantiate this meta-model and validate its functioning through a series of data queries.

References

- [Leg2012] (2012). Legifrance. <https://www.legifrance.gouv.fr/>.
- [ope2015] (2015). Open platform for french public data. <https://www.data.gouv.fr/>.
- [Abdelbaki and Zerouali2012] Abdelbaki, C. and Zerouali, M. (2012). Modélisation dun réseau dassainissement et contribution a sa gestion a laide dun système dinformation géographique-cas du chef lieu de commune de chetouane-wilaya de tlemcen algérie. *LARHYSS Journal P-ISSN 1112-3680/E-ISSN 2521-9782*, (10).
- [ASTEE2015] ASTEE (2015). Gestion patrimoniale des rseaux dassainissement. <https://www.astee.org/>.
- [Bertolino et al.2011] Bertolino, A., Calabrò, A., Lonetti, F., Di Marco, A., and Sabetta, A. (2011). Towards a model-driven infrastructure for runtime monitoring. In *International Workshop on Software Engineering for Resilient Systems*, pages 130–144. Springer.
- [Boury-Brisset2013] Boury-Brisset, A.-C. (2013). Managing semantic big data for intelligence. In *STIDS*, pages 41–47.
- [Chahinian et al.2016] Chahinian, N., Piat-Marchand, A.-L., Bringay, S., Teisseire, M., Boulogne, E., Deruelle, L., Derras, M., and Delenne, C. (2016). How can big data be used to reduce uncertainty in stormwater modelling?
- [Chen and Cohn2011] Chen, H. and Cohn, A. G. (2011). Buried utility pipeline mapping based on multiple spatial data sources: A bayesian data fusion approach. In *Twenty-Second International Joint Conference on Artificial Intelligence*.
- [Commandre et al.2017] Commandre, B., En-Nejjary, D., Pibre, L., Chaumont, M., Delenne, C., and Chahinian, N. (2017). Manhole cover localization in aerial images with a deep learning approach.

- [COVADIS2019] COVADIS (2019). Standard de données réseaux d’aep & d’assainissement, version 1.2. <http://www.geoinformations.developpement-durable.gouv.fr/>.
- [Da Silva2015] Da Silva, A. R. (2015). Model-driven engineering: A survey supported by the unified conceptual model. *Computer Languages, Systems & Structures*, 43:139–155.
- [Dong and Srivastava2013] Dong, X. L. and Srivastava, D. (2013). Big data integration. In *2013 IEEE 29th international conference on data engineering (ICDE)*, pages 1245–1248. IEEE.
- [Erraissi et al.2018] Erraissi, A., Belangour, A., and Tragha, A. (2018). Meta-modeling of data sources and ingestion big data layers. *Smart Application and Data Analysis for Smart Cities (SADASC’18)*.
- [Gascueña et al.2012] Gascueña, J. M., Navarro, E., and Fernández-Caballero, A. (2012). Model-driven engineering techniques for the development of multi-agent systems. *Engineering Applications of Artificial Intelligence*, 25(1):159–173.
- [Hafsi et al.2017] Hafsi, M., Bolon, P., and Dapoigny, R. (2017). Detection and localization of underground networks by fusion of electromagnetic signal and gpr images. In *Thirteenth International Conference on Quality Control by Artificial Vision 2017*, volume 10338, page 1033803. International Society for Optics and Photonics.
- [La Fosse et al.2020] La Fosse, T. B., Cheng, Z., Rocheteau, J., and Mottu, J.-M. (2020). Model-driven engineering of monitoring application for sensors and actuators networks. In *Software Engineering and Advanced Applications*.
- [Object Management Group2006] Object Management Group, I. (2006). Meta object facility (mof) 2.0 core specification. <https://www.omg.org/spec/MOF/2.0/>.