A CIM-Based Graph Data Model for Reliability Studies of Medium-Voltage Cables

Lunodzo Mwinuka
University of Salento/ IST
lunodzo.mwinuka@unisalento.it

Konrad Sundsgaard
Green Power Denmark
ksu@greenpowerdenmark.dk

Massimo Cafaro
University of Salento
massimo.cafaro@unisalento.it

Guangya Yang
Danmarks Tekniske Universitet (DTU)
gyy@elektro.dtu.dk

Lucas Pereira
ITI/LARSyS & Instituto Superior Técnico
lucas.pereira@tecnico.ulisboa.pt

Hugo Morais
INESC-ID & Instituto Superior Técnico
hugo.morais@tecnico.ulisboa.pt

***Abstract*—Managing dynamic data sources using relational databases has proven to be inefficient in terms of both modelling flexibility and query performance in power systems. While graph databases are better suited than relational databases for modeling power grids, they lack standardised models for specific applications like reliability analysis. This paper addresses that gap by proposing a graph data model based on the IEC Common Information Model (CIM). The model is validated using a case study on medium-voltage cable failures. Building on our previous work, this research sets the groundwork for a scalable, standardised approach to support decentralised data integration, advanced analytics and machine learning in power systems.**

***Index Terms*—MV distribution grid, Assets Management, Graph Data Model, Failure Detection, Databases.**

I. Introduction

Power systems are rapidly transforming, leading to an increase in data volume and variety. As information systems become central to grid operations, the continuous evolution of data size and structure is now the norm. However, providing quality data for applications like Machine Learning (ML) is a significant challenge due to two main issues. First, data management is not standardised, forcing stakeholders to rely on their own methods. Second, traditional tools like Relational Database Management Systems (RDBMS) have proven inadequate. RDBMS are too rigid to easily accommodate the rapid structural changes required by modern grids, demanding significant effort to adapt [1].

The distribution grid is one of the key sources of data in power systems as it spans over the majority of consumers, producers, and technical equipment. Furthermore, to operate the system reliably, distribution system operators (DSOs) must understand how the entire network is connected, both for planning and real-time operation. Indeed, a historical insight is important for various analytical tasks, including post-incident analysis, to understand the root cause of component failures; for example, it is highly relevant in the context of long-term investment planning [2].

With the current data management practice, realising these requirements demands substantial efforts for data collection, cleaning, and integration of datasets due to data heterogeneity [3]. Furthermore, existing data models require refinements to enable standardised modelling and the easy integration of power system components (in support of scaling) and other external data sources, such as weather conditions and infrastructure developments (in support of failure reporting). Among the existing models, graph data models can help DSOs adopt standard protocols and implement best-performing, scalable data models that align directly with distribution grid components. Even though there exist previously recommended graph models [4]–[6], the majority are aiming at facilitating data analytics without a look into standard data modelling, which is highly recommended by several initiatives such as data (energy) exchange reference architecture (DERA) [7]. Furthermore, a specific investigation of critical components of the distribution grids, such as medium-voltage cables, was not given adequate consideration.

Fortunately, to support data standards, the International Electrotechnical Commission (IEC) 61968 series, bundled in the Common Information Model (CIM), already exists, describing standards for grid representation, modelling and information exchange. The standard has been recognised as a key element to connect different DSO systems and applications using a standardised data model, enabling interoperability for smart grid applications [8]. In particular, the IEC 61968 parts 4 (interfaces for records and asset management), 6 (interfaces for maintenance and construction), and 11 (data model for distribution-specific components) are well-detailed enough to guide addressing the existing data modelling challenges. However, compliance in both research and industry remains limited. Some works provide a global scope of power systems, which may lead to missing the small context of power systems [1], [4]. Additionally, the CIM profile for DSOs does not cover all the requirements of the DSO's domain, as it follows a high-level design. It thus may not facilitate all activities within

This work was funded by the EU Horizon 2020 programme (Marie Sklodowska Curie grant No 956433, project InnoCyPES) and the Portuguese Foundation for Science and Technology (FCT) under projects LA/P/0083/2020; UIDP/50009/2020; UIDB/50009/2020; CEECIND/01179/2017 (L.P.); and UIDB/50021/2020 (H.M.).

TABLE I
AN OVERVIEW OF CURRENT GRAPH DATA MODELS IN POWER SYSTEMS.

Reference	Year	CIM	IEC	MV Cables
[4]	2016	✓	✗	✗
[5]	2018	✓	✗	✗
[1]	2021	✓	✗	✗
[10]	2021	✓	✗	✗
[11]	2025	✓	✗	✗
This work	2025	✓	✓	✓

DSOs. This makes it impractical to implement it directly. The DSO's domain requires a higher level of detail to support smart grid applications sufficiently [5]. Narrowing down, this work presents an important use case aimed at enhancing reliability studies of medium-voltage (MV) cables, as justified in [9]. A summary of the graph models for power systems following the CIM standard is presented in Table I. Current literature on CIM-based graph models overlooks inter-profile relations and the modeling of MV cable network activities that require extending the standard.

Therefore, this work introduces a scalable, graph-based data model to enhance the reliability studies of MV cables by integrating failure data with external features, such as weather and digging activities. The model complies with DERA and CIM/IEC 61968 standards to demonstrate scaling. We focus on MV cables because they are critical to power distribution, yet vulnerable to failures from ageing, environmental stress, and human activity. In Denmark, they are a major cause of outages. Therefore, accurately assessing their reliability is essential for effective maintenance, investment planning, and regulatory compliance [9]. Hence, we contribute as follows:

- Transform the relational data model presented in [9] into a graph-based data model;
- Integrate the CIM/IEC standards into the model and demonstrate its transformation and scaling;
- Benchmark relational and graph data querying against requirements for failure data management.

II. GRAPH DATA MODELLING FOR MV CABLE SYSTEM

We provide a graph-based data model to facilitate easy data retrieval relevant to reliability studies of MV cables. The model demonstrates a few abstract (parent) and specific (child) classes, wherein the representations allow for a demonstration of how graph-based data modelling could be easier than relational modelling. The model presented in this work has been designed to respond to the requirements listed in the GitHub repository[1].

A. Description of Graph Nodes and Properties in the MV Cable Network

Nodes represent the key entities in the system. Each node type is assigned a label and properties based on the scenario. By convention, dominant nouns become entities, which are then labelled as nodes in the graph data model. Referring to the requirements of our use case, the model must include details about "cables", "cable systems", "DSO", "substations", "events", "cable sections", and "joints". Most of these are directly mapped to nodes.

[1]https://github.com/H2020-InnoCyPES-ITN/Graph-Data-Model-for-Asset-Management-in-a-Distribution-Grid - containing the entire graph-based model, Cypher queries, descriptions, and their equivalent CIM/IEC model

TABLE II
A SUMMARY OF NODES AND LABELS.

Node	Sample name	Labels
DSO	Cop Dis	Dso
MV Cable System	System1	MVCSystem
Substation	station1	mainSubstation, secondarySubstation
Cable section	section	MVCSubsection
Cable joint	Joint	CableJoint
External event	External Event	ExternalEvent, DiggingActivity
Driver	Driver	LocationDriver
Event	Event1	CableRepair, CableFailure

When deciding properties to include, it is essential to ensure that a node can be uniquely identified, meets the model requirements, and returns a data item. For instance, given a Cypher statement `MATCH (d:DSO name: "Operator 1"-[:OPERATES]-(s:Substation) RETURN s.name`, property `name` can be used to uniquely identify a DSO and to answer specific details. A property `s.name` in this case has been used to return the data from the query. Depending on the scenario, a `name` property may not be a good choice for a unique identifier. In such scenarios, other properties can be used, or randomly generated characters can be used. Each property is also associated with the type to which the data will be stored. An overview of the proposed graph-based model is presented in Figure 1. At the centre of the model, there is a DSO node, named "Cop Dis", which expands to other nodes (the Name property in DSO serves as a unique string identifier for each DSO). Table II presents the nodes and their labels as visualised in Figure 1.

The supplyArea (Geometry) defines the geographical operational zone of the DSO, which is crucial for queries related to coverage. Other information included is the date registered, reliability index, and others. These properties collectively enable the identification of DSOs, their operational areas, and operational history, aligning with modelling requirements such as determining which DSO operates a specific MV Cable System or Substation.

The Substation node needs a unique identifier. While a name (String) can also be unique, the ID ensures distinctness. For each substation, we also need to know its high and low voltage limits. These can be modelled as the operational voltage parameter of the substation, by presenting high and low voltages as separate properties, or can also be modelled as a single string representing a range (e.g., "10-20 kV"). Substations can be removed, replaced, or repaired. Hence, the corresponding installation needs to be tracked as well. The location is also added. These attributes are essential for identifying substations, determining their voltage characteristics, and locating them.

Also, regarding the MV Cable System, the `name` cannot function as an identifier, and the use of a unique identifier is

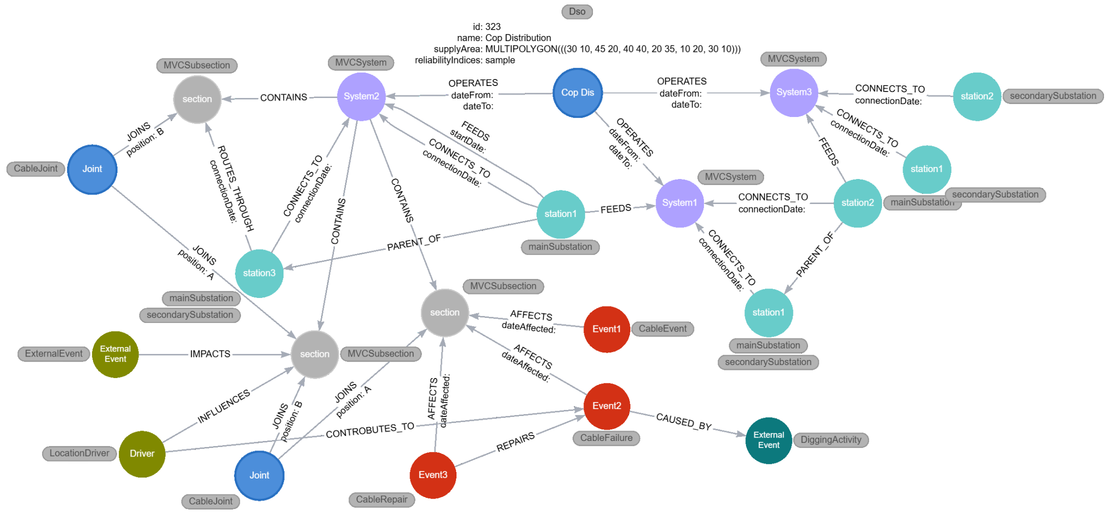

Fig. 1. A graph-based data model: graph nodes and edges for MV cable network and failure data.

required. Properties `averageLoading` and `maxLoading` are intended to reflect the system's operational load. It should be noted that `averageLoading` is a derived property, potentially computed over time, and might not be strictly necessary during the initial design phase. These properties are key for applications related to cable system capacity studies, but might also represent a crucial feature for reliability-related studies.

For the MV Cable Subsection, an ID ensures unique identification. Physical attributes such as `numberOfConductors`, `conductorSize`, `conductorMaterial`, `insulation`, and `conductorType` provide detailed technical specifications necessary for the analysis. The manufacturer allows for tracking of components by origin. The property `inServiceDate` indicates when the subsection became operational, and `length` quantifies its length. It is also essential to track the maintenance service dates of the cable subsection. These detailed properties fulfil several modelling requirements, including failure analysis, maintenance records, and inquiries about specific cable materials or types that may be affected by events such as digging.

The Cable Joint node has an ID that uniquely identifies each joint. In each joint, `jointType` specifies the type of joint, which can be necessary for failure analysis, for example, if certain joint types are more prone to failure. `coordinatesInstalled` gives the precise location of the joint. In this way, the model will facilitate tracking the correlation between the number of joints and cable failures.

The Cable Event node uses an ID as a unique identifier for all cable events, whether failures or repairs. The dates for the start and end of events should also be modelled, which might differ from when the cable was affected. Similarly, an event location and its cause are equally important. For failure events, the type of failure and repair done to a specific failed component should be modelled.

External events, such as weather conditions and digging activities, may also contribute to cable failures. Hence, necessitating the modelling of the input in a separate node, which holds different kinds, i.e., lightning, cold wave, heatwave, digging, etc. In this context, events should have a date, time, and location. More specific details could also be added as properties in each event, as further detailed in our model.

Location drivers can influence cable events. Hence, drivers, such as road types, railways, soil types, and weather conditions, all with their corresponding coordinates and properties, should be modelled. These properties model static environmental influences on cable subsections, independent of specific failure events, and help identify contributing factors to failures. Properties associated with the nodes described and that further represent the use case are presented in the GitHub repository.

B. *Description of Edges*

Edges represent relationships between nodes. Each edge type is assigned a label, direction, and properties. In graph relationship modelling, it is conventional to look for verbs, which, in most cases, naturally identify relationships among nouns (nodes). For example, a substation's operation under a DSO automatically maps to an `OPERATED_BY` relationship. An MV Cable System is connected to its subsection through a `ROUTES_THROUGH` relationship to track the association between cable systems and their subsections. The DSO is mapped to an `OPERATES` relationship towards the MV Cable System and Substations. The data model also contains a `JOINS` relationship to map the link between cable joints and subsections. In this, each cable joint connects two subsections. The model also includes the `FEEDS` relationship, which assumes that a main substation can feed multiple cable systems. The model also includes the cable event that maps to the subsection through the `AFFECT` relationship. Substations further map to two more relationships: `PARENT_OF` with the secondary substations and `CONNECTS_TO` with the MV Cable System. To better track cable information, repairs, and maintenance are also modelled. In doing so, Cable Events are linked to several other nodes. First, it is related to itself through the REPAIRS relationship, which models repair events that are also associated with failure events. In practice, repair of cables involves splitting the affected section into two parts and adding a new section. A cable event is also linked by a left-to-right relationship, where location drivers can contribute to failures. This relationship has been modelled as CONTRIBUTES_TO. In some cases, external activities such as digging can lead to cable failures. Hence, failure is linked to external events through the CAUSED_BY relationship. Just as the location driver contributes to some failures, it can also influence the MV Cable Subsection through the INFLUENCES relationship. This helps to model the static environment's impacts on cables, independent of failure events.

For each relationship, a direction should be specified between the nodes to model the interaction and actions between them. In the DSO and Substation nodes, with OPERATES relation type, it is conventional for this relationship to be written as $DSO \rightarrow OPERATES \rightarrow Substation$. In some database engines, when directions are not indicated, the left-to-right relationship is picked by default. A representation $DSO \leftarrow OPERATES \leftarrow$ Substation may not reflect the facts of our scenario.

The edges may differ among designs that represent similar scenarios. Hence, it is fair to choose designs that can easily respond to the requirements. For example, modelling voltages can be mapped to a single property, such as operating voltage, or mapped to two different properties: lower-level and higher-level voltages. The latter can be helpful in cases when retrieving cable systems operating at a specific higher or lower voltage. An alternative mapping could involve having voltage as a separate node, which is then linked to Substations via the OPERATES_AT relationship, with lowerLevelVoltage and higherLevelVoltage as the properties of the relationship. Such a design is beneficial for answering questions about which Substations operate at the same voltages in the network. A decision on whether a particular attribute is relevant to implementing the data model will depend on specific stakeholder requirements. It is best practice to control the number of super nodes—nodes having incoming or outgoing relationships. Other relationships, along with their directions and descriptions, are presented in the repository.

C. *Refining the model following CIM-IEC Standard*

This study analysed the presented model and assessed its alignment with the CIM/IEC 61968 parts 4, 6, and 11, while keeping the use case intact. The significant changes involve the naming structure, properties, and level of detail. For instance, all identifying properties of an asset are named as mRID to represent a Master Resource Identifier, aligning with CIM's PowerSystemResource.mRID being used for unique asset identification across systems. Additionally, a node DSO is modelled as an Organisation node, which will then define how it owns or manages Asset records. The node substation retains its original naming, but with updated property names. Voltage levels are now modelled as nominal voltages. Subtypes are used in the CIM standard to differentiate among substations.

The MV cable system, an abstraction in the original model, is restructured as an AssetContainer or EquipmentContainer to group related ACLineSegments (representing MVCSubsections). MVCSubsection aligns with ACLineSegment, with properties like conductorMaterial and lengthInKm linked to CableInfo via Asset.AssetInfo, per IEC 61968-4. The CableJoint node is modelled as a Junction (IEC 61970::Wires) for electrical connectivity and an Asset with JointInfo for physical properties, addressing both network topology and asset management requirements.

To track cable event and maintenance information across the system, CableEvent nodes for failures and repairs are refined to align with IEC 61968-6. Cable failures are mapped to FailureEvent, linked to affected Assets, whilst repairs are represented as WorkOrder with associated WorkTasks, capturing detailed maintenance processes. Also, ExternalEvent nodes are modelled as ActivityRecords with faultCause attributes, whilst LocationDriver attributes (e.g., SoilType) are integrated into Asset.Location or LandProperty, per IEC 61968-4. Dynamic weather data are represented as MeasurementValues tied to MeasurementPoints. In practice, most of these will need to be retrieved from external databases via available public Application Programming Interfaces (APIs). Edges such as OPERATES, FEEDS, and CONNECTS_TO are preserved but mapped to CIM's Terminal and ConnectivityNode structures for precise representation of electrical topology. The REPAIRS edge aligns with WorkOrder associations to FailureEvent. A brief presentation of edge mapping is available in the GitHub repository.

III. Implementation of the Graph Data Model

The designed model was implemented using the Neo4j (v5) database and Cypher for model implementation and data querying. We also used Cypher to generate synthetic data that aligns with the model and the use case, ensuring that all data types, including complex types such as location and relationship mapping, are enforced. Within the scope of reliability studies of MV cable systems, certain requirements can be designated as mandatory. For instance, it is crucial to be able to retrieve information about cable systems and their ownership, substations and their associated sections, areas, components, DSOs, and the substation with the most failures, including the reasons for these failures, the impacts of digging activities, and other relevant details. A detailed list of query requirements is available in the GitHub repository. Similarly, a Cypher code is also presented to fetch graph data. We tested model performance by comparing it with a relational-based model run implemented in PostgreSQL (v15). Both systems were loaded with an identical, synthetically generated dataset, loaded with 5 DSOs, 50 substations, 100 asset containers, 500 AC line segments, 300 junctions, 200 failure events, 150 work orders, and 100 external activity records. We run three queries: Q1 responds to the number of failures that occurred in a given year, Q2 identifies DSOs with the most failures, and

Q3 examines the leading factors for the most failures. Query formulation and its syntax are detailed in GitHub.

IV. Results

The performance evaluation is presented in Table III, where PSQL represents PostgreSQL, BA represents Block Accessed, Records represents the number of returned results, and DH represents Database Hits. For simple aggregation tasks (Q1), the relational database demonstrated superior performance, executing in 0.532 ms with only four block accesses. However, for queries requiring deep traversal of interconnected data (Q2 and Q3), the graph database was significantly faster, with execution times up to 1.7 times quicker (21 ms compared to 35.181 ms). Although the graph model registered a higher number of internal database hits, its native relationship-based processing proved more efficient than the computationally expensive JOIN operations required by the relational model for complex, multi-hop queries. With larger data and more complex queries needed in the industry, the observed performance gap would be even more pronounced.

TABLE III
A comparison of query performance.

Query	PSQL		Records	Neo4j	
	Time	BA		Time	DH
Q1	0.532 ms	4	1	1 ms	601
Q2	35.181 ms	88	5	21 ms	22,522
Q3	16.789 ms	6	7	6 ms	4,875

V. Proposal for Extension of the Model

By leveraging the CIM/IEC framework, the model enables effortless extension to incorporate additional asset details beyond the mini-world. Given that graph-based structures intuitively excel at representing node connections and efficiently retrieving interconnected data, such extensions pose no significant challenges. But cables are just one of many components in the distribution grid. CIM considers all physical utility equipment as Assets. To align with this requirement and enable further scalability, the cable system can now be defined as one of the data items in the `Asset` node, alongside other assets such as switches and transformers. As illustrated in the model, the design introduces an asset node with multiple labels to denote specific types of assets. Each asset is linked to associated maintenance, failure, and repair nodes, consistently with the previously described structure.

VI. Discussion & Future Work

Aligning to CIM requires extending use-case requirements. While the original model supports efficient querying for predefined tasks, its CIM alignment is limited by available information. Full adoption of CIM may require further normalisation, especially for asset catalogues and work management. Future work will assess its value and feasibility. Alternatively, modelling directly according to CIM and extracting relevant data per use case offers flexibility and warrants further investigation. As key differences lie in naming, granularity, and interconnections, refining the model for improved compatibility remains an open area for exploration. Depending on the scenario, full CIM compliance may not be necessary. However, identifying alignment gaps supports integration and future extensions. Hence, developing adaptive models that apply standards selectively is a key direction. For internal analyses, the original model remains sufficient; however, in collaborative and multi-stakeholder contexts, adherence to standards becomes essential. This work supports global efforts in decentralised data management. Future steps include applying the model in peer-to-peer (P2P) overlays and further exploring the impacts of different designs.

VII. Conclusion

This paper presented a graph-based data model to facilitate reliability studies of MV cable systems. Two models are presented: one that has been transformed from a previous relational data model [9], and a refined graph-based data model that follows CIM/IEC standards (parts 4, 6, and 11). By doing so, the model effectively captures relationships among assets, failures, and environmental drivers, ensuring scalability and interoperability. The design's adoption of these standards facilitates data sharing among stakeholders and future extensions to other grid components, promoting standardisation and decentralised data management for smart grid applications.

References

[1] P. Shi, G. Fan, S. Li, and D. Kou, "Big data storage technology for smart distribution grid based on neo4j graph database," in *2021 IEEE 4th International Conference on Electronics Technology (ICET)*, 2021, pp. 441–445.

[2] K. Sundsgaard, "Predictive maintenance of medium voltage cable networks using machine learning," Ph.D. dissertation, DTU, 2025.

[3] K. Sundsgaard, J. Z. Hansen, G. Yang, M. Cafaro, and P. K. Hansen, "Exploring data collection and fusion of medium voltage cable failures - a danish case study," in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2023, pp. 1–5.

[4] G. Ravikumar and S. A. Khaparde, "A common information model oriented graph database framework for power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2560–2569, 2016.

[5] D. Ascher and C. Kondzialka, "Towards model-driven cim-based data exchange for dsos," *Energy Informatics*, vol. 1, no. Suppl 1, p. 23, 2018.

[6] B. Kan, W. Zhu, G. Liu, X. Chen, D. Shi, and W. Yu, "Topology modeling and analysis of a power grid network using a graph database," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 1355–1363, 2017.

[7] E. Commission, D.-G. for Energy, M. J. Couto, K. Kotsalos, and K. Kukk, *European (energy) data exchange reference architecture 3.0*. Publications Office of the European Union, 2023.

[8] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. González, *The Common Information Model CIM: IEC 61968/61970 and 62325-A practical introduction to the CIM*. Springer Science & Business Media, 2012.

[9] K. Sundsgaard, L. J. Mwinuka, M. Cafaro, J. Z. Hansen, and G. Yang, "A decentralised relational data model for reliability studies of medium-voltage cables," in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*. IEEE, 2024, pp. 1–5.

[10] V. M. Balijepalli, F. Sielker, and G. Karmakar, "Evolution of power system cim to digital twins - a comprehensive review and analysis," in *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2021, pp. 1–6.

[11] J. Hwang, G.-H. Kim, S.-A. Seo, J.-U. Song, S.-I. Lim, and Y.-S. Oh, "Common information model-oriented ontology database framework for improving topology processing capability of distribution management systems considering interoperability," *Applied Sciences*, vol. 15, no. 8, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/8/4105