Visualization Techniques for Electrical Grid Smart Metering Data: A Survey

Maria Stefan, Jose G. Lopez, Morten H. Andreasen and Rasmus L. Olsen
Department of Electronic Systems. Aalborg University, Denmark
{marst, jgl, mha, rlo}@es.aau.dk

Abstract—One of the considerable initiatives towards creating a smart society could be the guarantee of a smart, resilient and reliable power grid. As an attempt to improve the electricity supply service, it would be meaningful for the distributed system operators (DSOs) to be able to monitor the current status of the grid. The prediction of future possible critical situations would then be feasible using the available information, whereas, based on historical data, further grid expansion and reinforcement may be planned. A proper presentation and visualization of the near-real time metering data may constitute the baseline for bringing improvements to the power grid. This paper presents an approach to build an efficient visualization system so that the extracted smart meters information can be used in a meaningful manner. An overview of the use cases related to the visualization features is first presented, as a motivation for the choice of the relevant state of the art research. In relation to the knowledge provided by the metering data, a definition of the big data concept will be further introduced, according to the requirements established by the project definition. Geographic Information System (GIS) tools are useful to help visualize the collected big data in near-real time. For this reason, a survey of existing GIS software will be made so that the choice of the most suitable tool can be justified. Also, the integration of GIS technologies into the Common Information Model (CIM) aims to improve the visualization efficiency. As a consequence, investigating methods for adapting CIM standards to the GIS platform are also important.

I. INTRODUCTION

In the process of the development of a more efficient electrical grid, the concept of the so-called "smart grids" has emerged. The purpose is to create an affordable, reliable and sustainable electricity supply. As a consequence of the development of smart grids, the distributed system operators (DSOs) in the Danish electricity distribution grid are facing operational challenges due to a large number of new smart electronic devices. These devices load the producer utilities with a high amount of data, reporting issues such as cable and converter faults, voltage magnitude outside standard limits and network congestion [1]. In order to address these challenges, intelligent features are required so that the DSOs can obtain an overview of their low voltage grid. This would allow the execution of near real-time daily operations in the grid, as well as long term grid management and planning. The efficiency of the electrical grid could be improved through the collection, processing and analysis of data and the outcome would have people as the main beneficiary. Eventually, the grid’s efficiency would be characterized by user satisfaction, economic implications, population reach etc. In the long run, the pursuit of progress in public and private sectors constitutes an initiative to create a smart society. In the current paper different methods are investigated towards building such a visualization system (GIS tools, CIM modeling, implementation languages). The uniqueness of this research consists of experiments that correlate different methods to achieve the desired visualization.

Current electricity grids are the baseline for future grids, which could account for the changes in innovative technologies, customer needs, environmental issues and increasing network congestion [2] [3]. The future distribution grids’ architecture is evolving from a one-direction power flow towards a bi-directional flow between suppliers and consumers, according to the European Commission’s view on electricity systems. The aim is to create a customer-oriented electricity system that will be flexible, accessible, reliable and economic [2].

Traditional power systems are developing towards digitalization, with the emerging ICT (Information and Communication Technology). In [4] digitalization is defined as a key element in the further development of the power system, being able to provide an efficient time-critical monitoring of the grid state, where issues would be signaled through the display of alarms.

In the energy sector, large amounts of data are accumulated daily. The main source of data in a smart grid is the adaptive metering infrastructure (AMI), where a large number of smart meters are deployed at the user end side [5]. Mechanisms for collecting data from the smart meters are addressed in [6] for the real-time state estimation of the grid. They are advantageous because they can facilitate the power flow control and identify exceeding limits of current and voltages. In other words, data collection mechanisms offer some degree of certainty of data quality. Correctness, completeness and timely data are some of the main attributes that can describe the data quality and it is important that they are ensured prior to its processing and viewing.

Three main steps need to be accomplished in order to obtain the data visualization platform, as shown in Figure 1: from storing data in a database to the "human eyes". Establishing efficient ways of transforming smart meters raw data into meaningful information can contribute to the operation, management and planning of the power grid.

A systematic storage of smart grid data is possible using database structures. As it is presented in [7], database archi-
tectures and methods also allow the processing and analysis of large amounts of data, continuously aggregated by time flow.

A more accurate representation of real-time data is closely related to the delays involved in data transmission from the smart meters to the visualization platform. The delays involved in the delimitation of the real-time definition can be defined as the transmission and the processing delays, represented in Figure 2. The transmission delay (smart meter - database) is bound to the adaptive data collection mechanisms presented in [6]. The data processing delay (database - visualization) is the aspect to be analyzed in this research and it will be used for defining the use case related to the real-time display of information.

The dynamics of information impacts the visualization in the sense that information that changes rarely over time can be treated differently. More processing time can be spent on such information compared to very volatile information elements that often change values, meaning that it is highly dependent on the information granularity.

The outlined challenges lead therefore to investigating scalable and secure IT infrastructures for real-time events management in the smart grid. Extracting significant information is important for the decision-making process and this is intended to be made through big data analytics. The challenge of the big data research is to include a set of technologies that would ensure users’ privacy, but still extract the valuable information needed for real-time grid operations and long term scale planning.

A possible implementation of data analytics can be done via the open standard Common Information Model (CIM). This model facilitates exchange of power system network data between companies and allows data exchange between applications within a company; thus, easier implementation of data analytics. Its purpose is to increase reliability and reduce expenses in smart grid infrastructures, as shown in [8].

The paper is organized as follows: Section II introduces the use cases concerning the real-time and historical features of the visualization platform. In Section III the motivation for choosing the CIM standard is presented, along with its newest uses and challenges both in research and industry. Section IV describes the related work regarding: big data use cases, visualization using GIS software and their relation to big data analytics. Furthermore, a definition of big data will be explained here in relation to the previously-mentioned use cases. In Section V different GIS tools are evaluated to motivate the choice of the most suitable one for carrying out the project.

II. DATA VISUALIZATION USE CASES

Data collected from the smart meters is a raw set of facts, without any particular meaning or usability. The process of how knowledge is obtained out of raw data is presented in the DIKW diagram shown in Figure 3.

In order to obtain valuable information out of raw data, some management tools have to be established. First of all, database systems are a fundamental tool for storing any kind of data, for later processing and analysis. The processing of data aims to provide the proper and timely input to the future visualization platform, moving the geospatial data to near-real-time access [9].

Information is the outcome of data processing, whose patterns and facts can be analysed to determine what is needed in the development of the visualization system. It becomes knowledge when it is applied in a particular situation to answer questions such as "why" and "how".

In order to establish the relevant state of the art survey, some use cases have to be defined. The visualization features can be divided into: time-critical (grid monitoring) and non-time critical (grid history).

A. Grid Monitoring

The real-time visualization features of the system are highly dependent on how often data is received and stored in the database and the events involved in the real-time data processing. Hence, one has to define what data needs to be processed for each visualization feature, to achieve the minimum processing delays in this context. At the same time, the visualization features need to be scalable according to the grid size and the volume of data stored in the database.

The delay in data processing is time-critical for the monitoring system. It has to be minimized taking into account:
the amount of data required to make qualified decisions, performance cost of data monitoring and the triggered events, and the amount of time required for the necessary data to be received.

**Example: Voltage drop/rise**

With the current technology it is possible to display the voltage magnitude of the system. The voltage drop is an issue that has to be adjusted and it can be done by monitoring the voltage stability margin, as explained in [3].

The overall quality of the distribution power grid can also be affected by voltage unbalance (phase imbalance) [10]. Furthermore, voltage rise (due to large amount of distributed generation power output) is presented in [11] as one of the main challenges for the DSOs to regulate the voltage levels in the grid.

Another useful feature that enables operators to quickly identify the fault location is be the ability to display real-time alarms in the visualization system [3]. This could be achieved using dedicated visualization features of GIS tools, as presented in Table I. For example, the ArcGIS GeoEvent Processor makes it possible to track dynamic data, which changes location frequently. Likewise, MapInfo’s Animation Layer add in is utilized for applications where data features update constantly.

The difference between the time stamp where the first suspicious event is received and how big is the delay to the actual alarm visualization should be investigated as well.

**B. Grid History**

The non-time-critical features are related to the electrical grid planning, to create models for future state estimates. Therefore, it should be established what parameters to keep track of, which data should be passed on to the visualization system and how to process data so that the database visualization system can be optimized. Data analytics and data mining techniques can help in discovering data patterns, that can be later useful in future grid planning in a quicker and more efficient manner.

**Example: Power balancing**

An example of how tracking historical data helps providing a better understanding of the grid state is presented in [12], for forecasting electrical loads. Statistical analysis can be created and further used to evaluate grid operating conditions and thus, failure assessment.

The collected historical data contains information about grid events, for example frequent oscillations in the electrical load power. Important information related to power failures can be extracted using historical data, while data mining techniques can provide an in-depth explanation for the failure cause. Parameters such as temperature, weather conditions and electrical load may be considered in the data analysis. Making use of historical data creates the foundation for a more efficient future grid planning, creating the opportunity to avoid power failures.

**III. CIM Modeling**

The initial motivation for the development of the CIM was driven by the increasing requirements for Energy Management Systems (EMS) to enhance upgradability, scalability and interoperability. Ambitious European energy and climate goals dictate the future and will change the very nature of the power system, promoting cooperation between European utility enterprises (DSOs) in order to support the implementation of the EU energy policy. Special focus is on the integration of Renewable Energy Sources (RES), which expands distributed power generation, and the Internal Energy Market (IEM) to meet the EUs energy policy objectives of affordability, sustainability and security of supply. This will invariably increase the data exchange necessities both internal and between the European utility enterprises and hence the relevance and need of a common semantic framework like CIM [13].

The use of ICT and advances in the design of power systems result in an increase and variation of the data. Its analysis and recognition require advanced data mining techniques, which are however heavily restricted by insufficient description of the data models. These issues lead the development from a vendor hard- and software specific API to a focus on common semantic and syntax models for data exchange between the EMS database and applications, which advanced into the CIM architecture comprising several standards under the International Electrotechnical Commission (IEC) Technical Committee 57 (TC57) and associated workgroups (WG) [14].

The integration of CIM has been adopted by Statnett, Norway's national main grid owner and operator. The future Statnett vision for CIM integration is centered on the perceived benefits in conjunction with predicted necessities for next generation smart grid development. CIM is envisioned to provide One common Power System Model (PSM), delivering a complete enterprise service oriented integration that is adaptable to future requirements; introducing a standard for data exchange via a common Enterprise Service Bus (ESB) (CIM EAI message bus - Figure 4), capable of handling all data governance and data management services. As part of the development of the future power system, Statnett is undertaking pilot projects analyzing the potential and performance of smart grid technologies and communication systems. One such project includes Demand Side Response (DSR) load control via AMI, where CIM was utilized as the standard for data exchange between the distribution management system (DMS) and the AMI front end, utilizing CIM XML messaging (DSO - AMI front end; IEC 61968-9) [15].

The European Network of Transmission System Operators (ENTSO-E) is a collaboration between 42 European TSOs representing 35 countries, including Norwegian Statnett and Energinet.dk of Denmark. By the same token, smart meter penetration of European consumers, i.e. households and Small and Medium Enterprises (SMIs), is currently on the rise and is expected as a continuous trend for the near future. ENTSO-E utilize CIM IEC standards to provide common data exchange formats to ensure compatibility for the various information
sharing between transmission system operators (TSOs), third parties and service providers alike. Direct cooperation is made with the IEC workgroups responsible for CIM for transmission (IEC TC57 / WG13) and CIM for energy markets (IEC TC57 / WG16) securing TSO influence and compliance, and supporting the continuous development of CIM [17].

The uses and challenges of CIM in the research area are addressed in [14], which deals with data driven interactive visualization of power systems. CIM based model visualization is done via data manipulation algorithms based on empirical or mathematical derived utility data. The main benefit is to provide smart grid operation and analytics decision support tools, enabling electric system operators and analysts to perpetually monitor big data information and events by images, diagrams, animations promoting communication and interpretation with enhanced pattern recognition. Furthermore, emerging business requirements of the electric power industry gain from the power systems data visualization.

Complete GIS and Supervisory Control and Data Acquisition (SCADA) integrated for monitoring of the power distribution network are presented in [18], including a common graphical user interface and shared network model. CIM is utilized for modeling the distribution network and its standard compliance facilitates data exchange. This allows for the aggregation of power equipment information and spatial GIS data with real time operational status information, enhancing decision support and abnormal alert response. Application development costs are reduced with multi-platform support superior to commercial GIS, avoiding duplication of data while enhancing data validity and reducing human error.

The integration of smart substations in smart grid architecture supporting intelligence aggregation in utility operation and management is addressed in [8]. Complying with IEC CIM standards supports substation analytics and system integration with enhanced value adding information exchange. As a result, smart connectivity is provided for Intelligent Electronic Devices (IEDs), promoting interoperability at all utility system levels and enabling operating and functional information exchange suited for individual IEDs, utility and decision support systems.

In [16], an electric utility company utilizes big data via analytics and the proposed software framework is based on CIM IEC standards, in order to convert utility big data into operational decision support, promote efficiency and save costs. Figure 4 illustrates how utility big data applications can interact with each other using a CIM-based integration architecture, through a common enterprise service bus.

Given the above-mentioned examples of how CIM is applied both in industry and in research, it can be concluded that its implementation is beneficial for an effective visualization of the electrical smart metering data due to the following reasons:

- Shared common services through a message bus interface: mapping of the CIM class structure to an application’s external interface;
- Facilitation of a real time environment for dynamic data messaging;
- Enhanced data analytics;
- Identifying and resolving issues in order to ensure the quality of data;
- Platform independence.

CIM modelling is feasible via multiple formats [19]:

- UML-CIM: standard defined in UML using classes, attributes and relationships;
- XML-CIM: class structure mapping and data encapsulation format. Encoding of plain text enables human-machine interaction;
- RDF (Resource Document Framework): an XML schema that defines relationships between XML nodes (outside parent/child class relationships). Nodes are assigned unique RFD IDs and resource attributes. In addition, an RFD schema is needed to provide the vocabulary for describing an object oriented type of system for RDF. The combination of RDF and RDF schema supports a class hierarchy structure XML schema through inter-class properties;
- CIM, XML and RDF: their combination is used to model the entire CIM power system. Provides a readable format for both humans and machines, due to the platform independent plain text format;
- XML Messaging: data exchange is provided through the XML data structure (CIM messages), associated with an
XML schema that defines classes and attributes interpretation.

Therefore, competitive, privacy or security concerns prohibiting open exchange of complete model data can be alleviated by layered data exchange via CIM with restrictions ensuring only the required data is shared.

IV. BIG DATA AND DATA VISUALIZATION

In the energy sector, the progressive penetration of Distributed Generation (multiple sources of small scale power generation) brings deep changes in the design of the grid [20]. At the same time, the penetration of a large number of power electronic devices (PVs, heat pumps, smart meters) has brought major changes in the volume of collected data.

Big data analytics can bring new opportunities in the management of a smart grid, in terms of data storage, analysis and mining, as mentioned in [5]. The work also states the importance of GIS as a traditional and complex source of big data, characterized by spatial attributes.

In various parts of the existing literature big data is often referred to as data of very high volume, variety and massive continuous flow. Variety, volume and velocity, also known as the "3Vs" are some of the most common big data characteristics. Thus, the notion of real-time is closely related to the speed required for processing and analyzing the data [5] [16] [21].

But these parameters cannot provide meaningful decision support to ensure a smart, resilient and reliable power grid, unless valuable knowledge is extracted from big data. In [5] the core of smart grid big data management is presented: data mining techniques and knowledge representation/visualization.

Another study [16] proposes additional solutions to enhance utility big data, apart from the "3Vs". Utilization and analytics are recommended for implementing a user-centered application framework, while a presentation of the collected big data is useful for the visualization framework.

The design of a data platform supporting enterprise level Big Data Integration (BDI) is addressed in [21]. The methodology proposes data integration and big data analytics schemes into a common data repository platform featuring scalability, real time data and security. The objective is to overcome conventional solution challenges and to create user friendly and powerful data query visualization and analytics tools.

Some examples of big data use cases are presented in [22]. The big data reference architectures of social networking services, such as Facebook, Twitter, LinkedIn and Netflix, are shown. The implementation of such use cases requires a great variety of technologies regarding data integration, storage, analysis and visualization, with the purpose of a better understanding of consumers needs.

The necessity of being able to display in near real-time the acquired information from different sources has led many researchers to develop GIS-based systems. [23] and [24] address the issue of real time data integration by using open source desktop GIS software, such as QGIS integrated with Grass. Other studies approach the combination of different types of GIS tools to display information: ArcGIS and QGIS [25], QGIS and Pmapper [23], QGIS, GRASS and MapServer [24]. An overview of GIS software tools utilized for visualization purposes is presented in [26] and [27], the latter concluding that QGIS is a better choice for data visualization and spatial analysis.

Given the aforementioned big data related work and the use cases presented in Section II, it can be concluded that the "3Vs" may be a relevant big data definition. However, the knowledge acquired from big data has a heavy impact in a decision making process. Therefore, a more accurate definition should include techniques on how and what big data can be actually used for, in the direction of developing a reliable, secure and effective power grid: data mining, visualization and data analytics.

V. VISUALIZATION SURVEY ANALYSIS

Based on the studies in Section IV, some of the most popular desktop GIS tools are summarized in Table I, according to a few key characteristics that explain the motivation for choosing the appropriate tool for the purpose of this research: license type, supported platforms and plugins/extensions/add ins that could make possible a near-real time implementation.

Various tasks carried out by different kinds of GIS software are presented in Table 1.1 in [28]. Due to privacy concerns and the possibility for local manipulation of data, desktop GIS type has been chosen for the development of this project.

ArcGIS proprietary software is popular for its analytical functions, scripting tools and the possibility for user developed functions in multiple programming languages. The Tracking Analyst extension makes it possible to reveal and analyse data patterns [29], while GeoEvent Processor is able to process time critical events [30] [29].

In MapInfo, the first desktop GIS product, additional tools can be implemented through its dedicated MapBasic programming language, such as the animation layer add in, which is used for tracking frequently updated data, as in the case of real-time applications [31] [32].

Many of the common functionalities of a desktop GIS can also be found within the MapTitude commercial software. It does not provide any real-time related analytic capabilities, but these can be customized using the GISDK application development platform [33].

GRASS and gvSIG are open source GIS software that come in handy for storing and managing spatio-temporal data and solving planning issues [26]. 3D visualization of data and animations are achieved through their user interfaces and the customized extensions.

Open source QGIS software runs on various operating systems and supports data formats from both ArcGIS and MapInfo [28]. Its browser interface makes it possible to access, organize, and visualize data within the supported spatial layers [34]. Similar to MapInfo and MapTitude, its functionalities may be extended by creating additional plugins.
using Python or C++. Therefore, it is possible to integrate features for real-time display of the data, as it has been done in [7].

VI. CONCLUSION AND FUTURE WORK

This paper addressed the motivation and challenges for building an accessible and effective system to display real time and historical visualizations based on data acquired from smart meters. Due to the advantages of CIM, it will be a part of the future work to model the components of an electrical grid, in order to manage smart metering data and to represent GIS data. It is expected that the implementation of CIM will result in enhanced possibilities for data analysis techniques, which constitutes one of the main steps that have to be defined towards building the visualization platform. The choice of a suitable GIS tool was done according to the requirements in the defined use cases, especially the challenge of integrating real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGIS is considered to be a suitable choice due to the fact that it is open source, supported in real-time data visualization. QGI

ACKNOWLEDGMENT

This work is financially supported by the Danish project RemoteGRID which is a ForsKEL program under Energinet.dk with grant agreement no. 2016-1-12399.

REFERENCES


TABLE I

SURVEY OF GIS TOOLS

<table>
<thead>
<tr>
<th>Desktop GIS Software</th>
<th>License</th>
<th>Implementation Language</th>
<th>Platforms</th>
<th>Visualization features (plugins/extensions/add ins)</th>
<th>User Interface and Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcGIS</td>
<td>Proprietary</td>
<td>Python, C++, R</td>
<td>Windows OS</td>
<td>GeoEvent Processor, Tracking Analyst (extension)</td>
<td>ArcMap, ArcCatalog</td>
</tr>
<tr>
<td>QGIS</td>
<td>Free</td>
<td>Python, C++</td>
<td>Linux, Unix, Windows OS, Android</td>
<td>Open Layer (plugin)</td>
<td>QGIS Browser</td>
</tr>
<tr>
<td>MapInfo</td>
<td>Proprietary</td>
<td>MapBasic</td>
<td>Windows OS</td>
<td>Animation Layer (add in)</td>
<td>Color stretch tool</td>
</tr>
<tr>
<td>GRASS</td>
<td>Free</td>
<td>Python, C++</td>
<td>Linux, Unix, Windows OS</td>
<td>Temporal Framework (extension)</td>
<td>Vector Digitizer, 2D/3D visualization with wxNviz</td>
</tr>
<tr>
<td>gvSIG</td>
<td>Free</td>
<td>Jython (Python implementation in Java)</td>
<td>Linux, Mac OSX, Windows OS</td>
<td>3D Animation Manager (extension)</td>
<td>3D Interface</td>
</tr>
<tr>
<td>Maptitude</td>
<td>Proprietary</td>
<td>GISDK (Geographic Information System Developers Kit)</td>
<td>Linux, Unix, Windows OS</td>
<td>Drive-Time Rings, Drive-Time Territories (geographic analysis tools)</td>
<td>MapWizard (thematic maps)</td>
</tr>
</tbody>
</table>


