21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2015 Grid of the Future Symposium

A Future Oriented Data Platform for the Electric Power Grids

T. XIA
Dominion Virginia Power
USA

Tao.Xia@dom.com

SUMMARY

Grid of the future relies on future oriented data platform to operate and manage the power grid and the components it contains in an adaptable, secure, reliable and automatic way. This paper starts with a brief introduction of big data problem faced by humankind and specifically in the power industry, emphasizing the importance of data storage, processing, and analytics tools employed to facilitate the conversion of raw data to management decisions. Big data typically represent three distinct characteristics of data: volume, variety, and velocity, which are discussed in details with power grid data examples. The day-to-day difficulties that the electric utilities are consistently confronted with are also discussed to better understand the ramifications of unsolved big data issues. To fulfill the U.S. Department of Energy's goal of an adaptable, secure and reliable power grid, the renewable generation, smart grid controls, energy storage, plug-in hybrids, and new conducting materials have been widely deployed across the United States, which will require fundamental changes in the operational concepts and maintenance practices. That in turn is demanding a future oriented data platform to integrate huge amount of data from a large number of sources and a diverse set of applications and must be scalable to support a high volume of various sensors installed across the grid and still provide real time guarantees. The fundamental idea in the data integration and management part of the data platform is that it will be the common data repository for data of all sorts and serve as the one-stop shop for major data needs. In the mean time, state-of-technology data modelling is greatly needed in this future oriented data platform along with a well guided self data service process that will help a utility identify all data sources, timeliness of the data, and what data is really needed to solve the utilities engineering and operations needs. Dominion Virginia Power is in the process of establishing such a data platform that will integrate the operational data with asset information and others to operate its grid more reliably and better manage its asset across the electric power network. Dominion has been implementing a comprehensive data historian with highly contextualized data structure, in conjunction with a Network Model Management (NMM) application, as the integrated solution to enable such objectives. The data flow diagram of the proposed data platform as well as the Asset and Network Model Integrated Solution Architecture being implemented in DVP is displayed and explained and the key concepts are thoroughly discussed. The easy wins of the data platform implementation are being presented and the opportunities of advanced grid analytics enabled by this data platform are introduced. The paper concludes with the affirmation of the success of Dominion data initiative and its constructive role in providing common data repository, state-of-technology data model, and self data service capabilities to the engineers, operators, planner and managers of the grid in this future oriented enterprise level data platform.

KEYWORDS

Big Data, Common Data Repository, State-of-technology Data Model, Self Data Services, Dominion Data Initiative, Data Platform, Asset and Network Model Integrated Solution

Introduction

In 2007, humankind was able to store 2.9×10^20 optimally compressed bytes, communicate almost 2×10^21 bytes, and carry out 6.4×10^18 instructions per second on general-purpose computers. General-purpose computing capacity grew at an annual rate of 58%. The world's capacity for bidirectional telecommunication grew at 28% per year, closely followed by the increase in globally stored information (23%). **[1]** With those unprecedented rapid growth rates, the challenges and issues associated with communicating, storing, processing, analyzing and visualizing the data in almost all industries have become an area of extensive research and development efforts.

Pivoting back the electric power grids, the electric power big data can be found in all sections of electric power generation and management. It is the key technology of electric power to overcome the challenges of limited resources and harsh environment, renewable energy integration, load fluctuation and sustainable development. **[2]** The data analysis systems for future power grid systems face many unique challenges, which include real time guarantees, scalability in term of the amount of data and the diverse of application to support, high reliability, and security. Although significant investments have been made in modernizing the electricity infrastructure, the ultimate benefits of investments will not be fully realized unless powerful data storage, processing, and analytics tools are employed to facilitate the conversion of raw data to management decisions.**[3]**

This paper aims in giving a real world example of a data platform being implemented in an electric utility company to meet the big data challenges.

Power Grid Big Data Challenges

The information technology definition of big data is it is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.**[4]** Big data typically represent three distinct characteristics of data: volume, variety, and velocity (3V). Volume corresponds to the amount of data collected for a specific purpose; variety represents the wide types and ranges of the collected data and the fact that the data cannot be structured into regular database tables; and velocity characterizes the pace of collecting and processing data from various sources. **[3]**

Currently, the primary sources of data in utilities are offered by SCADA (Supervisory Control and Data Acquisition) gathered from various sensing devices, phasor measurement units (PMUs) distributed over transmission and distribution networks, consumption data collected by smart meters, which are deployed at customer premises, and intelligent electronic device data, which represent the data collected from individual grid components. In addition to the data directly obtained from the electricity infrastructure, utilities may collect data from other resources to facilitate system studies such as weather data, geographic information system (GIS) data, manufacturer data, electricity market data, and etc.

At the foundation of the data explosion in electric utilities with regard to the volume aspect of the 3V big data characteristics is the meter read. Most customers had their meter read 12 times per year or once a month. Early AMR (Advanced Meter Read) systems were designed to bring a kWh read once a month for the purpose of providing the utility customer a bill. Today, 15-minute read intervals are what is expected of an AMI/AMR (Advanced Meter Infrastructure/Advanced Meter Read) which is 96 meter reads per day per meter or approximately 2880 reads per month (30 day month) per meter. This equals an increase of 287,900% in just kWh data. **[5]**

As to the variety aspect of the big data problem, data diversity is a major challenge in its own right at the time for this one and half century old industry. It was mentioned very briefly at the beginning of the section about the general often-used data categories for electric utilities, which are only a fraction of the data required to run the power grids safely, reliably and economically. The data formats, attributes, storage methods, and security levels are essentially different and the variations can sometimes be poles apart. Figure 1 shows a list of the major data sources in an electric utility company that has an extremely wide range from Digital Fault Recorder (DFR) to Construction Plans and from Relay Settings to On-line Condition Monitoring (OLCM). The list is still far away from complete albeit long and complex, demonstrating the level of gravity of the data variety characteristic in the utility big data realm.

Figure 1

A persuasive testimony to the velocity characteristic of utility big data challenge is sampling rate burst for the synchrophasor data measured at bus terminals, transformer bushing, transmission line conductors, and various other special locations. In the conventional SCADA data collection system, the measurements are being reported every 2~5 seconds, whereas the PDCs (Phasor Data Concentrator) are receiving data from PMUs at an astoundingly higher rate of 30~60 reports/second, almost a 2 orders of magnitude data velocity increase. That level of data rate expansion basically renders some of the traditional tools and applications for electric utilities non-functional, posing a real issue in terms of collecting, storing and processing this type of high-speed data.

Due to the 3V (Volume, Variety, Velocity) characteristics of the big data and lack of strategic methodology and plan to alleviate or eliminate the issues arising from it, the electric utilities are consistently confronted with day-to-day difficulties list below in the operations of their grid and the equipment within it.

1) Data Silos
2) No semantics layer on top of the data
3) Lack of cross system integration
4) Not all relevant data is shared
5) Difficult to share data and models
6) Excessive time used to validating data/models, not running studies
7) Data accuracy and inconsistency
8) Common data not in sync and up to date
9) Impossible to propagate data change to all pertinent data destinations

The big data challenges in power grids discussed above, high level and detailed alike, are demanding a future oriented data platform to meet the challenges and offer solutions to the issues.

Data Platform Requirements

The U.S. Department of Energy's (DOE) has set the goal for an electric grid as "adaptable, secure, reliable, resilient, and can accommodate changing loads, generation technologies, and operating business models". To fulfill that goal, the widespread deployment of renewable generation, smart grid controls, energy storage, plug-in hybrids, and new conducting materials will require fundamental changes in the operational concepts, situational awareness tools and maintenance practices. The whole system becomes highly dynamic and needs constant adjustments based on real time data. Even though thousands of sensors such as phase measurement units (PMUs) and smart meters are being widely deployed, a data platform that can support this amount of data in real time is greatly needed. **[6]** This future oriented data platform has to be able to integrate huge amount of data from a large number of sources and a diverse set of applications and must be scalable to support a high volume of various sensors installed across the grid and still provide real time guarantees. Moreover, the system needs to be highly reliable and highly secure because the power grid is a critical piece of infrastructure.

The key technologies of this future oriented electric power data platform can be divided into data integration and management techniques, data analysis techniques, data processing techniques and data visualization techniques. The integration and management technologies of this electric power data platform, which could create new application functionalities for enterprises, combine data from two or more information systems. In another words, the technologies integrate the data which are from different sources, of different formats, with different features, in logical storage media. **[2]** The fundamental idea in the data integration and management part of the data platform is it will be the common data repository for data of all sorts and serve as the one-stop shop for major data needs.

The power grid data in such a high variety usually do not contain structured data, but rather they are most likely unstructured data, semi-structure and mixed types. Being unorganized, the data will still remain in silos and it will be extremely difficult to sync up relevant data, propagate data changes to all pertinent destinations and other advanced data applications. Data contextualization grants us great advantages of data cross-referencing, data validation, easy data search and endless others. Also, well-modeled data can be processed in different ways, depending on the analysis that needs to be done or on the information that must be found in the initial data. Open standards such as IEC 61850 and CIM (Common Information Model) play an important role in power grids data models to bridge data segregations and enhance data sharing capability. In a nutshell, state-of-technology data modeling is undoubtedly required in this future oriented data platform.

Many utilities rely on their software vendors to develop the data visualizing and analyzing tools and programs needed to generate the information the utility needs to manage its operations. This long established practice may not be as efficient nowadays since the paradigm of data characteristics in the electricity industry has greatly shifted in the past decade. Factors such as vendors' unfamiliarity of different circumstance specific to individual organizations also exacerbate the dependency on the vendor for application development, especially for the applications that require much customization. Self data service methods can be implemented by a utility to further enhance the value of their data to lower operational costs and improve customer service. A well guided self data service process will help a utility identify all data sources, timeliness of the data, and what data is really needed to solve the utilities engineering and operations needs. A successful data platform has to contemplate the flexibility of self data services.

Proposed Data Platform Strategy

Striving to meet the requirements explored previously, Dominion Virginia Power in the process of establishing a data platform that will integrate the operational data with asset information to operate its grid more reliably and better manage its asset across electric power networks. Dominion has acquired an over-the-shelf commercial data historian, in conjunction with a Network Model Management (NMM) application, as the integrated solution to enable such objectives.

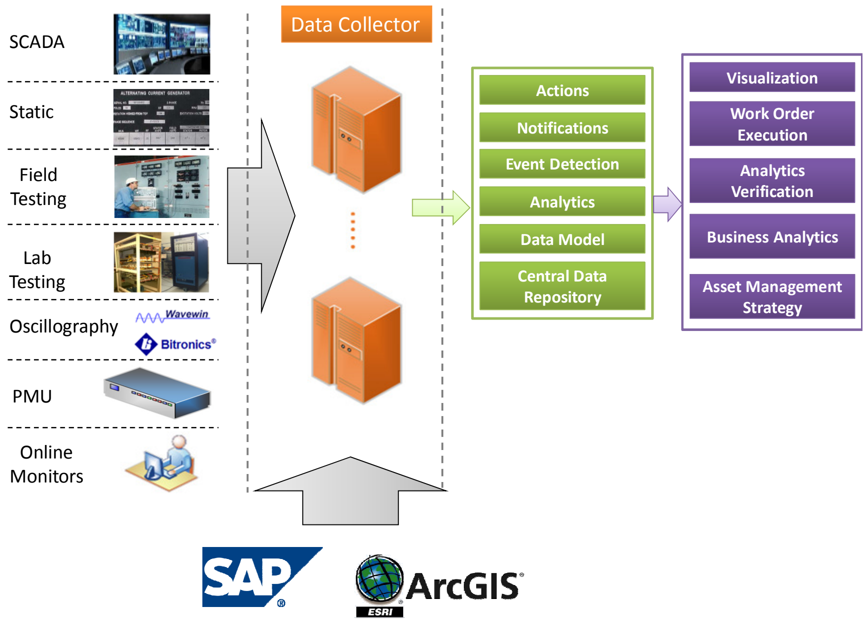


Figure 2

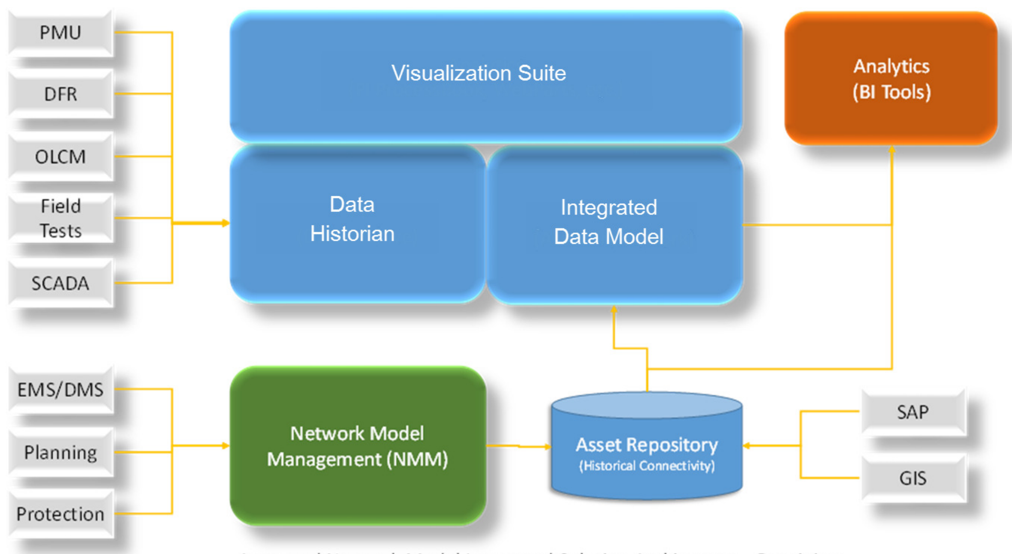
The data flow diagram of the proposed data platform being implemented in DVP is shown in Figure 2. Time-series data from a wide range of sensors and sources are integrated and brought into the data collector (the new Dominion data historian) by various interfaces and adaptors, including real-time dynamic data such as SCADA, PMU and on-line monitors, off-line data such as field testing and lab testing results, file-based data such as oscillography COMTRADE files, and static data such as equipment ratings. Non-temporal data such as SAP that stores work order, asset information, and maintenance record and geographical information such as ArcGIS are also being accounted for in the data platform. It should be mentioned that the diagram above only enumerates major sources for the sake of simplicity.

Once all the data are consolidated in to new data historian, there central data repository can really function as a strong engine to drive a great deal of applications related to enhance data utilization. As listed in Figure 2, the data model can be derived from the integrated data; analytics can be developed to improve grid reliability and operational economy; event detection and notification can be configured; visualization can be set up; dynamic asset management can be optimized by data driven algorithms; and even work order can be automatically generated if abnormality is found somewhere in the grid.

Key to the success of data modelling and NMM implementation in this platform is the power network model management and the integration of asset related data into the contextualized business intelligence data hierarchy. The goal is to correlate asset information with the connectivity nodes in the network models for both planning and operations. Consequently, Dominion is looking to implement some initial use cases to leverage network model and connectivity information in business intelligence data structure such as dynamic equipment health assessment for strategic asset management and optimal VAr advisory for the optimization of reactive power flow.

Traditionally, utilities gather asset health data from online and office line capabilities, as well as historical SCADA information. The advent of centralized Network Model Management

(NMM) capability has allowed utility to not only streamline their network model usage across operation, planning and engineering, but also to make the network connectivity information available for asset management purposes. Given the business challenges that Dominion and more broadly the industry faces today and some of the technology investments it has committed to, the Asset and Network Model Integration Solution Architecture is developed to meet the business challenges. This architecture reflects the industry best practices around the technology, standards, and requirements of utility asset and network model management. The following diagram provides an overview of the architecture. We believe that, implemented correctly, it will provide the right foundation for Dominion's short and long term asset management and operational applications needs as business requirements grow and change over time.



Asset and Network Model Integrated Solution Architecture - Dominion

Figure 3 - Asset and Network Model Integrated Solution Architecture

Specifically, this data flow diagram in Figure 2 and the solution architecture in Figure 3 propose the following key concepts:

- The data historian being implemented in Dominion will be the main repository for operational, offline and field test data with histories. This would allow Dominion to have access to a variety of operational data from a single source. Sources for this include but are not limited to: PMU, DFR, OLCM, Field Tests and SCADA.

- The integrated data model will provide the asset hierarchies so that users will be able to navigate to the desired operational data through a visual and searchable asset structure. This asset hierarchy model will be developed using Dominion's asset data structure and be supplemented by IEC CIM asset related model elements.

- The data historian visualization suite: this is where asset management related analytics can be developed and used by end users. Self data service business model will be the theme behind this concept.

- Network Model Management: this is a stand-alone application that provides a centralized environment for network model management and maintenance, leveraging the IEC CIM connectivity model structure. Vendor applications in this area can import and export a variety of model file formats including the IEC CIM-based connectivity model. Applications to interact with this environment will be those of EMS/DMS, Planning, and Protection.

- Asset Repository: the main purpose of this Asset Repository (AR) is to store the historical information about the electrical connectivity information and their relationship to assets. This information is critical to support the desired use cases for this project. There are a number of options to implement this asset repository. It can be implemented in the new business intelligence database server environment as a set of relational tables. It can be implemented as a stand-alone database. Or it can be implemented as part of the NMM application server environment. It also depends on Dominion's desire for the long term use of this Asset Repository. Asset data from SAP and GIS will be integrated into AR and then made available to the business intelligence data structure.

- BI Tools: for analytics that go beyond the data in the data historian and the asset repository, Business Intelligence (BI) tools can be deployed to support reporting, query, and analysis of asset data that integrated with the rest of the Dominion Enterprise, such as financials, work management, human resources, supply chain, etc.

Opportunities of the New Data Platform

After the implementation of the new data platform both in data integration and the data modelling, significant benefits can be derived from the consolidated and contextualized data. Early wins such as data enriched one-line diagram, data driven asset management, dynamic equipment monitoring, and other items shown in Figure 3 can be realized with modest efforts and investments, which can well be attributed to the data normalization and structuring in the data platform.

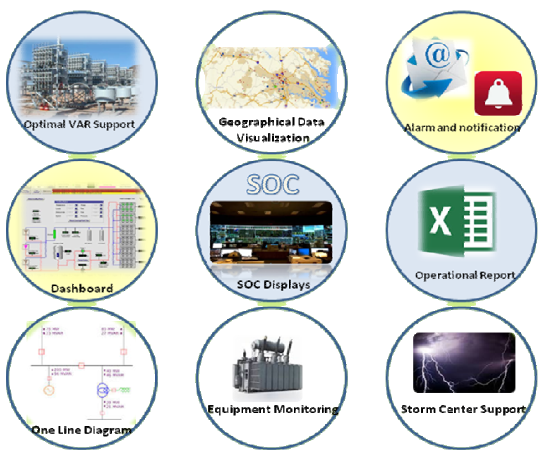


Figure 3

Mixed data from multiple sources, validated network model with mapped asset data, and across-the-system data semantics – all those complex data properties are what the current electric utility data systems typically do not possess to conduct advanced analytics. Within this future oriented data platform, however, the data integration process addresses those issues and makes those features available for the advanced analytics, which opens the gate for a great many leading-edge analytics for a more adaptable, more secure and more reliable power grid, among which are, 1) Advanced/Predictive Restoration Systems, 2) Adaptive Topology Planning, 3) System Dynamics and Transients Modeling Validation, and 4) Wide area Profiling and System Management, to mention but a few. All those aforementioned easy wins and advanced grid analytics are all great examples of opportunities of this new data platform.

As one of the early results of Dominion's new grid data platform implementation, a prototype data enriched one-line diagram has been completed to provide a data visualization and situational awareness tool for users from various departments across the company. One-line diagrams of various types are widely used in assorted applications and by different groups within an electric utility organization. Due to the immense diversity, the one-line diagrams usually only show data within certain domains and are restricted to a certain network not because of NERC compliance but because of firewall constraints.

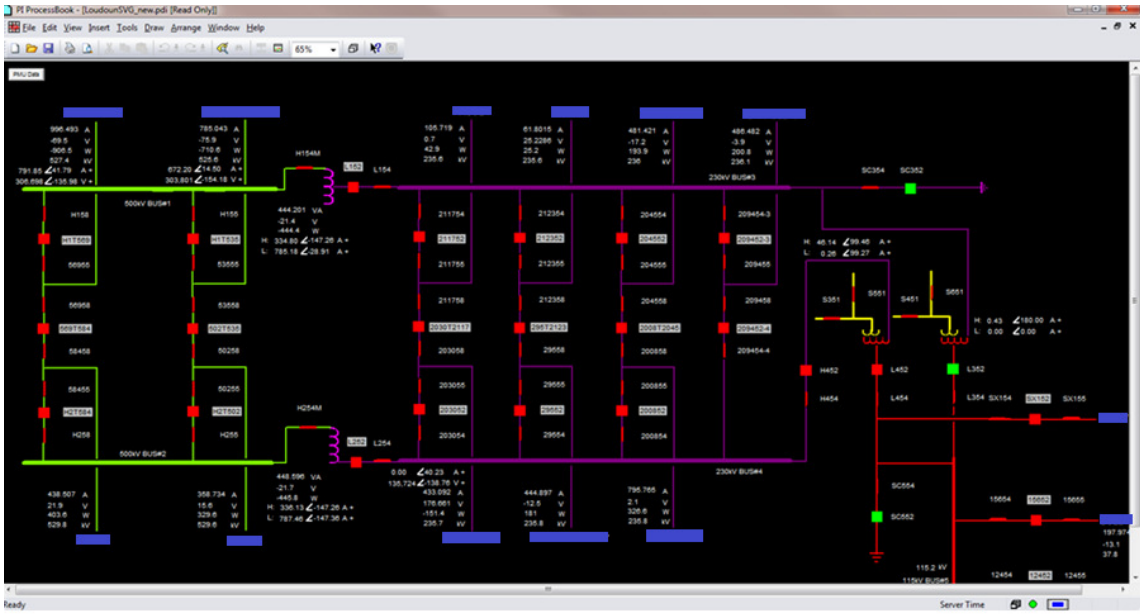


Figure 4

The example one-line diagram in Figure 4 is a replication of an EMS one-line of a substation with names marked out for privacy considerations. The reason it is referred to as data enriched one-line diagram is that the data behind this one-line diagram are not only data from EMS, but also data from other crucial operational data sources/systems as well. In this very diagram, there are data from EMS (SCADA), from synchrophasors (PMU) and from asset database (SAP) ready to be displays at the a mouse click. Notice there is a 'PMU data' button at the top left corner of the diagram. Clicking that will turn on/off the PMU data stream on the diagram as shown in the complex numbers on top of buses and below transformers. All available transformer asset information such as nameplate, manufacture, design characteristics, and electrical test results, is readily available by simply clicking the transformer icons. The benefits also go beyond the data display itself. This one-line diagram also provides the field support personnel with the same view as what is in the control room for better situational awareness across the entire electric utility company. Bringing all major operational data under one roof in an integrated data platform provides vast opportunities for improvements in both operational and maintenance of the power grid for better adaptability, higher security and superior reliability.

Conclusions

Today, utilities face a significant challenge in operating their transmission and distribution network more reliably and managing their aging assets in the network more strategically. With increased integration of renewable energy resources into the grid at both utility and consumer sites, the increased intermittency and variability of generation and load are putting further strain on the grid operation margin and the utility asset remaining useful life, many of which were not designed to operate in such conditions. Therefore, utilities are deploying solutions to collect, manage and analyze much more data from the grid in order to have a better understanding of the grid operation conditions and network asset performance, both descriptive and predictive.

To meet the challenge posed by the big data characteristics shaped by recent technology advancement and grid condition evolution, Dominion is taking a cutting-edge initiative in the electric utility data management arena to expand the capability of the data applications, increase the inclusiveness of data sets of the central data storage, improve the data standardization and modelling, and raise the data anlytics automation level. The initiative aims at providing common data repository, state-of-technology data model, and self data service capabilities to the engineers, operators, planner and managers of the grid in a enterprise level data platform. The proposed data platform has already had remarkable early wins and is making strides towards the U.S. Department of Energy's (DOE) goal for an adaptable, secure, reliable and resilient electric grid in the big data realm.

BIBLIOGRAPHY

[1] M. Hilbert, P. López "The World's Technological Capacity to Store, Communicate, and Compute Information" (Science, 1 April 2011, Vol 332, pages 60-65)

[2] J. Zhan, et al "Study of the key technologies of electric power big data and its application prospects in smart grid" (Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific, Year 2014, pages 1-4)

[3] S. Bahramirad, J. Svachula, J. Juna "Trusting the Data: ComEd's Journey to Embrace Analytics" (IEEE Power and Energy Magazine, Year: 2014, Volume: 12, Issue: 2, Pages: 107 - 111).

[4] F. Shull "Data, Data Everywhere …" (IEEE Software, Year: 2014, Volume: 31, Issue: 5, Pages: 4 - 7).

[5] S.E. McHann "Grid analytics: How much data do you really need?" (IEEE Rural Electric Power Conference (REPC), Year: 2013, Pages: C3-1 - C3-4).

[6] Y. Jian, P. Sharma, I. Gorton, B. Akyoli "Large-Scale Data Challenges in Future Power Grids" (Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on, Year: 2013, Pages: 324 - 328).