VESTLANDSFORSKING

Big Data in Electric Power Industry

Opportunities and Challenges for Sogn og Fjordane region

Rajendra Akerkar, Minsung Hong

VESTLANDSFORSKING-RAPPORT NR. 5/2021

TITTEL Big Data in Electric Power Industry	**RAPPORTNUMMER** 5/2021 **DATO** 09.06.2021 **GRADERING** Open **TAL SIDER** 43
UNDERTITTEL Energy Big Data	
PROSJEKTTITTEL TEKNOLØFT Sogn og Fjordane	**PROSJEKTNUMMERR** 6490
FORSKARAR Rajendra Akerkar, Minsung Hong	**PROSJEKTANSVARLEG** Øyvind Heimset Larsen
OPPDRAGSGIVARAR Norges forskningsråd	**EMNEORD** ICT, Energy, Power, Big Data

SAMANDRAG

As Big Data continues to disrupt almost every industry, the electric power sector has also started to catch up. This report presents the classification, sources and features of electric power Big Data (energy Big Data) in general, with special focus on the Sogn og Fjordane region. The current status of electric power Big Data is analysed, and challenges and opportunities are explored. This is a feasibility study for how we can identify and go on with interesting and useful data-driven R&D for the electric power industries and businesses. Target audience of this report is partners in the Teknoløft project and public and private energy organisations.

ANDRE PUBLIKASJONAR FRÅ PROSJEKTET

Rajendra Akerkar, Minsung Hong. Big Data in Aquaculture: Opportunities and challenges for Sogn og Fjordane region. 6/2021, ISBN 978-82-428-0435-8.

VESTLANDSFORSKING, Postboks 163, 6851 Sogndal

ISBN: 978-82-428-0434-1

På framsida: Electric hydropower collage

Table of Contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
AMI	Advanced Metering Infrastructure
AMR	Automated Metering Reading
AMS	Advanced Metering System
API	Application Programming Interface
CAPEX	CAPital EXpenditure
CIM	Common Information Models
DFS	Distributed File System
DPR	Digital Protective Relay
DR	Demand Response
DSM	Demand Side Management
E&P	Exploration & Production
ESB	Enterprise Service Bus
EV	Electric Vehicle
GDPR	General Data Protection Regulation
GIS	Geographical Information System
GPS	Global Positioning System
HAN	Home Area Network
HSE	Health, Safety and Environment
IED	Intelligent Electronic Device
IHD	In-Home Display
IoT	Internet of Things
JSON	JavaScript Object Notation
NPT	Non-Productive Time
ODBC	Open DataBase Connectivity
OPEX	OPerational EXpenditure
PaaS	Platform-as-a-Service

PMU	Phasor Measurement Units
PPDM	Professional Petroleum Data Management
PV	PhotoVoltaic
QoS	Quality-of-Service
SaaS	Software-as-a-Service
SCADA	Supervisory Control and Data Acquisition
SER	Sequence data of Event Recorder
SOA	Service-Oriented Architecture
RTEP	Real-Time Electricity Pricing
XML	Extensible Markup Language

1. Background

1.1. The country

Norway is a key energy nation in Europe with an entirely exceptional set of resources[1]:

- the petroleum industry, including oil companies, petroleum refiners, fuel transport and end-user sales at gas stations
- the gas industry, including natural gas extraction, distribution and sales
- the electrical power industry, including electricity generation, distribution and sales
- the coal industry
- the nuclear power industry, and
- the renewable energy industry, comprising alternative energy and sustainable energy companies, including those involved in hydroelectric power, wind power, and solar power generation, and the manufacture, distribution and sale of alternative fuels.

Norway[2] has only one per cent of Europe's population, but 20 per cent of the hydropower resources, 50 per cent of the water reservoirs (stored water for hydropower production), 40 per cent of the gas resources and 60 per cent of the oil resources.

As assets' yields become harder to access and even harder to forecast, it is vital that the industry is collecting and maintaining its data effectively. Big data is relevant to the whole energy sector. However, since the industry sector is too vast, in this report, we will limit our discussion to hydroelectric power and smart meters.

Norwegian power production is almost 100% renewable and emission-free. 95 per cent of the power production stems from the 1,600 hydropower plants which are spread all across the country, and some 3.5 per cent stems from wind power. The latter is expected to grow significantly in the coming years, due to very good wind resources in Norway. The Norwegian power grid is 330,000 km long and has a security of supply rate of 99.998 per cent. To take a leading role as the first renewable and all-electric society in the world, unique renewable energy resources in Norway make a good point of departure on the path towards an emission-free society, including green growth and new jobs in other relevant industries.

---

[1] https://www.energinorge.no/

[2] https://energifaktanorge.no/en/norsk-energiforsyning/kraftproduksjon/

Norway's interconnections with neighbouring countries are of vital importance for supply security, and it also enables the Norwegian renewable energy industry to participate actively in the European market. This is particularly important due to the fact that most years, Norway is a net exporter of electricity.

1.2. The region

The Sogn og Fjordane region is rich in renewable energy, and it`s annual electricity production of 15-17 TWh comes mainly from hydro and wind power. This report focuses on an impact of Big Data in the regional electric power sector including the following two power companies.

Sogn og Fjordane Energi[3] (SFE) is a power company that operates in Sogn og Fjordane in Norway. SFE is divided into three divisions: power production, sales to end users, and grid management. SFE builds on a long history as a hydropower producer, and has built itself up as one of the largest producers of renewable energy in Western Norway. The production operations are grouped in the company SFE Produksjon, which includes operation, maintenance and rehabilitation of existing production facilities, power development and development activities, and a power trading environment that drives physical and financial sales of the power production. The company has an average, annual power production of close to 2 TWh (average production), which corresponds to consumption in around 100,000 homes. SFE is the operator of 25 of 27 fully and partly owned power plants in Sogn og Fjordane.

Sognekraft[4] is a power company. The grid company of Sognekraft is merged with the grid companies of Aurland Energi and Lærdal Energi to the company Sygnir. The new company is now operating the grid in the municiplities Sogndal, Vik, Lærdal and Aurland. The company's core business is production, distribution and sale of electric power. The company has a power production of 608 GWh, and also has operational responsibility for power plants owned by others (125 GWh). Sognekraft sells about 700 GWh of power to end customers.

This report represents a literature study and an overview based on some initial dialogue with SFE and Sognekraft.

[3] https://sfe.no/

[4] https://www.sognekraft.no/

The Big Data in electric power makes the energy production, consumption and pertinent technical revolution deeply integrated into the Big Data philosophy, and promotes the systematic application of Big Data technology in the regional sector, and also accelerates the development of related industries and business model innovation.

2. Big Data Technologies

2.1. Traditional paradigm to the Big Data paradigm

Big data has changed the way that we adapt in doing businesses, managements and explorations. Data-intensive computing is coming into the world that aims to provide the tools that we need to handle Big Data problems. Table 1 outlines the shifts required to move from traditional to the Big Data paradigm.

Table 1 Traditional paradigm to the Big Data paradigm

Traditional Paradigm	New Paradigm
Some of the data **E.g.,** An online transaction records main data fields, a timestamp and IP address.	**All of the data** **E.g.,** Clickstream and path analysis of web-based traffic, all data fields, timestamps, IP address, geospatial location where appropriate, cross channel transaction monitoring from web.
Clean Data **E.g.,** Data sets are typically relational, defined and delimited.	**Chaotic Data** **E.g.,** Data sets are not always relational or structured.
Deterministic **E.g.,** In relational databases, the data has association, correlation, and dependency following classic mathematical or statistical principles.	**Complex coupling** **E.g.,** Data can be coupled, duplicative, overlapping, incomplete, have multiple meanings all of which cannot be handled by classical relational learning tools.
Examining of Data to Test Hypotheses **E.g.,** Defined data structures induce the generation and testing of hypotheses against known data fields and relationships.	**Discovery of Insight** **E.g.,** Undefined data structures induce exploration for the generation of insights and the discovery of relationships earlier unknown.
Lag-time Analysis of Data **E.g.,** Data needs to be defined and structured prior to use, and then captured and collated. The period of extracting data will vary but often involves a delay.	**Real-time Analysis of Data** **E.g.,** Data analysis takes place as the data is captured.

2.2. Definition of Big Data

An important point to be noted, while discussing the concept of Big Data, is that the phrase can refer to either huge and distinct datasets, or technologies processing such datasets. In literature, Big Data is classified into two different types: static Big Data and real-time Big Data. Both types of datasets can be structured or unstructured (Akerkar, et al. 2014). We can define Big Data as given in the following Figure 1:

Big data is	**using big volume, big velocity, big variety data asset to extract value (insight and knowledge),**
	and furthermore ensure veracity (quality and credibility) of the original data and the acquired information,
	that demand cost-effective, novel forms of data and information processing for enhanced insight, decision making, and processes control.
	Moreover, those demands are supported by new data models and new infrastructure services and tools which is able to procure and process data from a variety of sources and deliver data in a variety of forms to several data and information consumers and devices.

Figure 1 Big Data definition

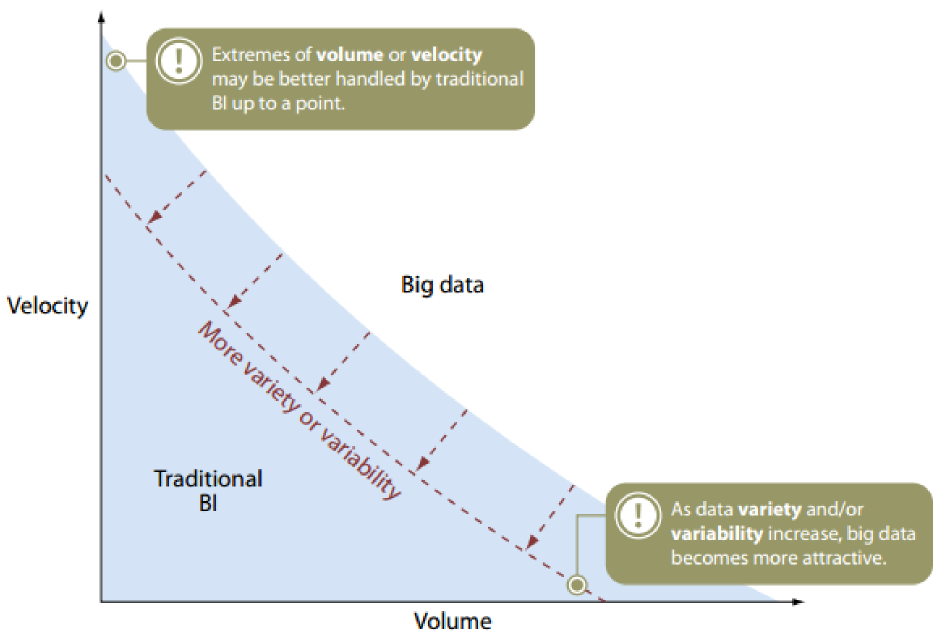

Figure 2 The graphics illustrating Big Data (Hopkins, et al. 2011)

The following Table 2 provides some business benefits which can be realised by using Big Data solutions in the electric power sector:

Table 2 Business benefits using Big Data solutions

Improved Operations	Drive combined insights from a single data management platform for structured, unstructured and real-time data. Reduce the operational Non-Productive Time (NPT) and Health, Safety and Environment (HSE), regulatory compliance cost due to 'real-time risk management'.
Unified Ontology for energy sector	Provide personnel with access to searchable institutional knowledge that compensates for limited expert staffing and achieving accuracy and helping personnel find what they are looking for more quickly. Given the shortage of experts, the time saved in accessing and loading data is important.
Faster Production Rate	Accelerate time-to-production by minimising data bottlenecks that reduce asset team productivity. Enable faster decision-making by Geologist & Geophysicists and operational teams as risk profiling and forecasting is performed.
Asset Development	Improve asset uptime and predict the need for asset related to operational demands.
Enhanced safety and efficiency	Enhanced safety and efficiency in operation by linking different relevant data with physical models.

Using a combination of Big Data and advanced analytics in electric power production and supply activities, experts can accomplish strategic and operational decision-making.

2.3. Big Data Opportunities

The opportunity that Big Data presents to all industry sectors is in the potential to unlock the value and insight contained in the data industries already held via the transformation of information, facts, relationships and indicators. The value of Big Data for industries is limited by their ability to efficiently utilise Big Data and the ability to derive useful information from this data. With every opportunity, there come barriers and business must overcome these barriers to explore the benefits of Big Data. Important areas that Big Data may influence are described below.

Data management

There are potential savings in time and money if industries implement smarter data management practices that were aware of the needs of Big Data analysis. Data sources from differing enterprises and operational areas would be of greater benefit to multiple industries and for multiple purposes if there were greater transparency. For example, through better business process management, redundant data collection processes can be reduced by reusing data collected from separate processes.

Personalisation of services

We have moved from an era of experiencing things at a macro level to experiencing things at a personal level. Big data analytics may produce value by revealing a clear picture of a customer. Big data is able to achieve this due to its characteristic granularity. This granularity may assist in unlocking the possibility of personalised services tailored to the individual and delivered by industry. The granularity in Big Data opens up new opportunities for personalising services. When a service provider knows something specific about a user then there is an opportunity to tailor the service offered accordingly. This will be most useful when the data in question relates to the user's needs, and when the personalisation is done in a manner that is prominent for the transaction being undertaken or service being used.

Predictive analytics

The alliance of multiple datasets from disparate sources in combination with advanced analytics technologies will advance problem-solving capabilities, and in turn, will improve the ability of predictive analytics to reveal insights that can effectively support decision-making. In short, Big Data opens up the field of reliable predictive analytics. By assessing the relationships embedded in large datasets it is possible to construct a new generation of models explaining how things are likely to evolve in the future. This approach can be blended with scenario planning to develop a series of predictions for how a system will respond to distinct choices. The state of the art in predictive analytics can deliver forecasts for some domains with a very high degree of precision, offering an auditable, scientific basis for making decisions in complex systems.

Productivity and efficiency

The analysis of Big Data sources can be used to identify cost savings and opportunities to increase efficiency and reliability, which will directly contribute to an improvement in productivity. This can in turn help to boost further innovation.

3. Characteristics of Electric Power Big Data

The existing data collection and management systems in the electric power sector are relatively isolated because these systems come from different departments and the data management of different departments is isolated. The electric power Big Data exists problems of multi-source heterogeneity, information redundancy, different time granularity, inconsistency of statistical models and uneven data quality. These problems will pose challenges to the integrated management, analysis and processing of electric power Big Data.

Electric power Big Data is a subset of Big Data in the electric power industry. So it has the "4V" features of Big Data, including volume, variety, velocity and value. For the electric power Big Data, their "4V" characteristics are reflected in the following aspects (Zhou, Fu and Yang 2016), as shown in the Table 3 below.

Table 3 Characteristics of energy Big Data in 4V

4V	Description
Volume	Introducing smart metering devices and sensors in smart energy systems as well as combining with other data sources might present many new opportunities as well as many tough challenges. The first challenge is the massive amount of data. This challenge is not only reflected in the storage side but more importantly in the analysis and processing of the electric power Big Data. For example, one smart meter, with a resolution of seconds to minutes, generates much fewer data than one phasor measurement unit (PMU), with a resolution of milliseconds; yet an advanced metering infrastructure (AMI) may generate a large volume of data coming from millions of customers.
Variety	This means increasing complex data types. In energy systems, the data are not only traditional structured relational data but also many semi-structured data like the weather data and Web services data, as well as unstructured data like customer behaviour data and the audio and video data. "The energy Big Data is a mix of structured, semi-structured and unstructured data (IBM 2014)." With the increasing utilisation of social media and call centre dialogues in the energy sector to support decision makings, the energy Big Data has become more varied. - Structured: standard and data models

	- Unstructured: images, log curves, well log, maps, audio, video, etc. - Semi-structured: processed data such as analysis, interpretations, daily drilling reports, etc.
Velocity	It refers to the speed requirement for collecting, processing and using the electric power Big Data. The speed of data collection and processing is very fast ranging from 5- or 15-mins interval to sub-second interval. For the many real-time tasks in energy systems, such as equipment reliability monitoring, outage prevention or security monitoring, the typical analytics algorithms that need many hours or more time to run are not competent. An example is a continuous stream of data, as opposed to a once-in-a-while event-triggered data from a sensor. Although the majority of power system sensors are event-triggered, there are also sensors that produce data streams at high rates. Also, the data circulation and processing speed of power enterprises are very fast, which realises real-time processing and analysis of a great deal of data in a fraction of a second. - Real-time streaming data from various equipment, and sensors - Relevant data fragments need to be automatically detected, assessed and acted upon.
Value	Energy Big Data itself is meaningless unless valuable knowledge that supports effective and efficient decision makings on the energy management process can be discovered. Also, the value of electric power Big Data is sparse, which means that the knowledge mined, and the value obtained from large amounts of data may be limited. Therefore, in the era of Big Data, we should pay more attention to the overall data rather than the sample data (Mayer-Schönberger and Cukier 2013). Moreover, for the power monitoring data, the abnormalities of the data are the important data, because the abnormal data is the key basis for condition based maintenance. - Enhancing production - Reduce costs, such as non-productive time (NPT) - Reduce risks, especially in the areas of safety and environment

According to (CSEE Committee 2013), there are also the "3E" (energy, exchange and empathy) characteristics of electric power Big Data. Energy (data-as-an-energy) means that energy savings can be achieved by Big Data analytics. Exchange (data-as-an-exchange) refers to the Big Data in the energy system that needs to be exchanged and integrated with the Big Data from other sources to better realise its value. Empathy (data-as-an-empathy) means that better energy services can be provided, users' needs can be better satisfied, and consumer satisfaction can be improved based on electric power Big Data analytics.

4. Sources of Electric Power Big Data

The type of data that can eventually form Big Data in power systems can be classified into domain data and off-domain data (Akhavan-Hejazi and Mohsenian-Rad 2018). The domain data can be further categorised by their sources. Here are a few examples:

Telemetry and supervisory control and data acquisition (SCADA) data

It enables a continuous flow of measurements on grid equipment status and parameters and other grid variables. The SCADA data can have various sources such as renewable energy resources which generate a huge amount of data, such as real-time production and equipment status, and is able to be used for specific purposes. For instance, the data from condition monitoring systems of many wind turbines can be utilized in predictive maintenance strategies.

Oscillographic and Synchrophasor data from intelligent electronic devices (IEDs)

It makes up of voltage and current waveform samples in time or frequency domains that can create a graphical record. Modern IEDs provide high-accuracy, time-stamped power system measurements primarily in the two formats. Synchrophasor measurements are a growing part of real-time operations at utilities during power system disturbances. Combining synchrophasor and oscillographic data during post-disturbance analysis enables to provide a wide-area context for a better understanding of the conditions leading up to and following a disturbance (Nakafuji, et al. 2017).

Consumption data

This data is most often the smart meter data. Energy consumption patterns in the consumption data can be analysed and extracted, so very valuable conclusions can be made for managers and governments. In particular, the forecasting of the energy consumption of buildings and campuses in terms of time series has immense value for energy efficiency and sustainability in the context of smart cities (Pérez-Chacón, et al. 2018).

Asynchronous event data

Even data often comes from devices with embedded processors generating messages under a variety of normal and abnormal conditions.

Metadata

Metadata is any data that can describe other data. Grid metadata is highly diverse and may include internal sensor data, calibration data, and other device-specific information.

Financial data

This data may include day-ahead and real-time market bids and price data, bilateral transactions, and retail rates.

Moreover, power grid operation relies also on different forms of off-domain data, i.e., the data that is not specific to or necessarily intended for the power sector. There are many forms of existing or emerging off-domain data that are yet to be exploited for the power grid operations and energy enterprise. Examples include traffic data, social media data, trade indices, and image and video streams.

The digitalisation of electric power systems boosts the rate of data generation. However, the characteristics of data flowing throughout the system reveal the challenges afore-mentioned (i.e. volume, variety, velocity). Therefore, Big Data is an essential technology to realise the paradigm shift in the electric power sector, and electric power Big Data (EBD) is the term that refers to data related to the sector. The sources of EBD can be categorised into two groups, namely electric utility data and supplementary data (Refaat, Abu-Rub and Mohamed 2016).

Electric utility data

Electric utility data contains all of the data and information that a utility can reveal from a smart grid system. The data includes many kinds of data such as SCADA data, phasor measurement units (PMU) data, smart meter data, IEDs' data, asset management data, digital protective relay (DPR) data, digital fault recorder (DFR) data, sequence data of event recorder (SER) data, AMI data, control and maintenance data for equipment, and automated metering reading (AMR) data (Nafi, et al. 2016).

Supplementary data

Supplementary data includes all other data sources which are beneficial for Big Data applications such as time-reference data, geographical information system (GIS) data, global positioning system (GPS) data, weather and lightning data, seismic reflection data, animal migration data, financial market data, social media data and regulatory reporting data.

Figure 3 summarizes comprehensive Big Data in the electric power sector.

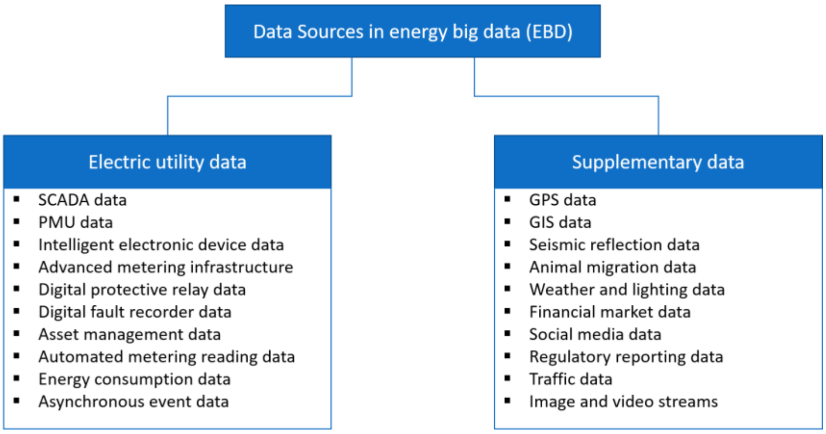

Figure 3 Energy Big Data

Also, regional power companies collect a large amount of sensor data from different hydroelectric powerplants, including but not limited to: vibration, temperatures, water levels, voltages, power levels and several other units. This naturally gives access to a huge source of information that has the potential of being used for Big Data analytics, predictive and optimal planning of maintenance and other big-data analysis in combination with other sources of information, such as weather observations, financial market prices and other spot prices in the electric power market. It is also possible to use the sensor data alone for trending and real-time surveillance of our power plants. We believe that an unexploited benefit can be achieved by pulling on the other information sources mentioned above, to further increase robustness, performance and profit in the hydroelectric power-industry. The Table 4 presents an example of sample data sources in the regional electric power industries.

Table 4 Example of electric power data sources

Case	Data Type	Data Source (Internal/External)	Data Format	Data Collection Method	Data Collection Rate (GB /month)	Data Volume	Data Retention Period	Data Storage	Sensitive/ Private Data
Smart home network	Smart Meter Data	Internal	AWS Data Bases	APIs based on REST technology	About 30GB/month	About 15TB	10 years	S3	O
	Smart meter data at consumer grid points, power plants and transformer substations	External on physical servers, internal partly utilizing cloud services/partly physical servers	Smart metering data (AMS): Hadoop DFS, XML timeseries database.	Smart metering data (AMS): Radio GSM mesh network Substations: Python API + IoT standard utilizing network protocol	About 5GB/month	About 15TB	More than 10 years for substations data Less than 10 years for consumption metering points due to GPRS	Smart metering data (AMS): Data management, physical servers controlled. Substations' data: Private database.	O (It can still be utilized for research projects.)
Smart energy grid	Sensor data	Internal	Historian DB (GE)	API ODBC	Less than 100GB/month	Less than 5TB	8 years		O
	Weather statistics	External		APIs based on REST technology			More than 30 years		X
	Weather forecast	External		APIs based on REST technology			About 10 years		X
	Energy spot prices	External		FTP-server	Less than 1GB/month	Less than 1TB	About 20 years		O
	Energy market prices	External		FTP-server	Less than 1GB/month	Less than 1TB	About 20 years		O

5. Big Data Flows in Electric Power Industries

The deployment of smart grids and prosumer participation in producing and consuming energy markets has resulted in enormous data management issues for electric power generation or/and distribution companies. Governments and power distribution companies around the world realise the challenges of managing Big Data in the electric power sector (Potdar, et al. 2018). The Figure 4 below illustrates the energy data lifecycle from data collection to strategic decision-making.

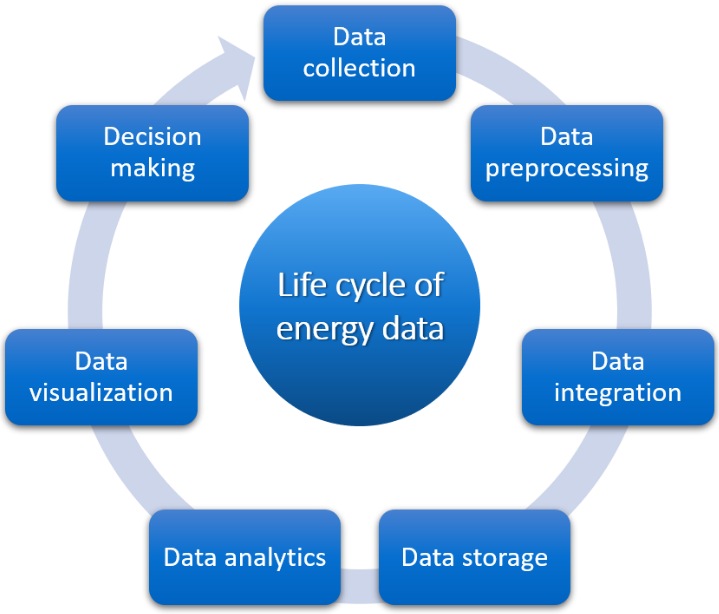

Figure 4 Lifecycle of energy Big Data

Data collection

This first step in energy data management involves capturing data from several sources in the electric power sector. The most common and well-known source of data originates from the AMI that captures data from smart meters from IEDs installed at end-user buildings. Depending on the population of a given location, this data can quickly escalate in size. That is a massive amount of data just from one source, which is the smart meter. There are other sources of electric power data, such as sensor data, voltage data, power quality data, control devices, mobile terminals, metadata, event-related data (e.g., breakdowns, voltage loss), reliability data, operating system data, energy grid equipment data, historical data and third-party data. In addition to this, smart energy grids rely on weather data to forecast demand and trigger on-demand renewable energy generation along with fault detection and user electric power consumption predicting

(Zhou, Fu and Yang 2016) (Daki, et al. 2017). Other than weather data, numerous other kinds of Big Data sources (e.g., mobile phones, Electric Vehicles (EVs), connected thermostats, real estate data, and customer energy behaviour profiles) integrate with the electric power sector to provide better forecasting and prediction services (Fehrenbacher 2012). Geographic information systems (GIS) also have an important role in the sector. Data from GIS sources can provide valuable information that can be used in decision-making systems since it provides local geographic information for many issues, such as identifying solar farm locations (Sánchez-Lozano, et al. 2013), electrification of rural areas etc. Since data is collected from so many various sources, devices and platforms, it requires the following tasks such as cleansing, proper integration and storage.

Data pre-processing

The pre-processing step includes the following two main tasks:

- *Cleaning*: Energy-related data is acquired from several sources as discussed earlier. On many occasions, such data might be inaccurate, meaningless and incomplete. Filtering such impure data is essential before analysing them. Data cleansing refers to such filtering processes to keep the data consistent and constitutes five distinct steps. The first step determines the faulty or abnormal data; next step is to detect these data; then correct the error; document it and modify the entry process to alleviate future data errors (Chu and Ilyas 2016). During the five steps, various points should be considered such as data format, completeness and its rationality to minimise errors during analysis.
- *Redundancy Elimination*: Redundant data means surplus data that can be neglected during the transfer and analysis process. Identifying such data is the first step. While it generally increases processing time and requires additional storage, eliminating the redundant data reduces the data transfer cost and results in energy and/or storage space savings (Sai and Chen 2017) (Chen, et al. 2017).

Data Integration

Energy Big Data consists of a variety of data collected from various sources as mentioned. Since the collected data originates from different sources, it may be not uniform and pose a significant problem for data analysis. Hence, the data should be appropriately integrated (Guerrero, et al. 2017). For example, data could be in different formats, which need formatting to a standard format before data analysis. In most cases, different service providers use different types of smart meters, each running different software, and following a proprietary storage format. Such inconsistency becomes a massive problem for the transmission and generation companies when trying to utilise such data (Zhou, Fu and Yang 2016). One solution to the data integration issue is standardisation. Standardisation will help in the longer-term when all the smart meters (new and legacy) and other smart grid infrastructure could talk to each other and share data seamlessly (Neumann, et al. 2015). Another solution for this problem can be a mediator software which can convert the data and maintain uniformity. There are several Big Data analytics tools available in the market that can streamline data integration. For instance, service-oriented architecture (SOA) can be utilised for data integration purposes. Enterprise service bus (ESB) and common information models (CIM) are some examples of SOA realisations (Daki, et al. 2017).

Data Storage

In the traditional energy grids, historical data is saved for forecasting purposes and does not require excessive storage space. However, in smart energy grids, there are other data resources requiring storage for effective demand management using renewable energy sources (Qiu and Antonik 2017). Another important issue is the in-/out-put data speed because there is a need for real-time data analysis in smart energy grids. So, energy data storage must have two essential features: a powerful data access interface system for speeding up data transfer and large and reliable storage space for storing various data. Thus, to deal with electric power Big Data, a desirable storage system would be organised into three parts: first part stores the data on a disc array, second is the connection and network subsystem to bridge between discs and servers, and the last part is the management software to share data between various other servers. Another way is a DFS like Google's GFS, Hadoop's HDFS and Taobao's TFS which is cheaper and provides higher performance (Mosaddegh, et al. 2016). Such DFSs allow sharing of resources among multiple users (Cao, et al. 2017). Additionally, NoSQL databases (e.g., MongoDB and Cassandra) also provide efficient mechanisms to store massive electric power Big Data.

Data Analytics, Mining and Knowledge Discovery

In the electric power sector, a massive amount of data is gathered regularly. Such data is analysed to discover different (or abnormal) patterns of usage within energy grids (Qiu and Antonik 2017). Various kinds of analysis are conducted on electric power Big Data, for examples: customer behaviour analysis (Park, Kim and Kim 2014), load analysis (Eltantawy and Salama 2014), state analysis (Weng, et al. 2016), operation analytics (Moradi, Eskandari and Hosseinian 2014), fault analysis (Jiang, et al. 2014), and signal analytics. Sort of analysis can be classified into two broad categories in terms of the response time required for data processing. The first category does not require fast response time and includes load analysis for forecasting long-term demand and consumer behaviour analysis. The other needs a very fast response time for real-time analysis. It includes various cases such as smart meter data for real-time price forecasting, identifying and analysing faults within the energy grids. Due to the characteristics of Big Data such as variety and velocity, analytical tools used for the analysis of electric power Big Data require high robustness, high scalability, high-velocity and fault tolerance features. And selecting a proper tool is significantly critical for Big Data analytics from the utility perspective (Diamantoulakis, Kapinas and Karagiannidis 2015). There are useful real-time processing tools like Splunk and Storm for analysis that provide fast execution time, fault tolerance capability and parallel computational ability. Hybrid processing tools, such as Apache Flink and Spark, are other options for the data analysis, where high execution time is not required.

Data Representation and Visualisation

Data visualisation as the next step after data analysis assists the decision-makers in understanding the analysis because it provides analytical results in forms of visual representations like graphs (Srinivasan and Reindl 2015) (Stefan, et al. 2017). Visualising results are more efficient than reviewing pure numbers to gain information and insights from data. It offers visible patterns which aid to identify and detect sources of concerns and opportunities. Stakeholders can make effective decisions from the visual interpretation of data than textual or numeric forms. Thus, visualisation is also significant. Furthermore, it is equivalently essential to the end consumer since it can easily offer the consumer their usage patterns. There are several tools and techniques available in the market that the utilities and consumers can use to visualise electric power Big Data. For example, the use of 2-D and 3-D visualisation tools to present the load forecast, user consumption, generation from the renewable source and power quality. Tableau is one useful tool to visualise the data intuitively. GIS software such as ArcGIS, QGIS, MapInfo, GRASS, gvSIG and Maptitude are good alternatives for visualising smart grid data on the maps (Stefan, et al. 2017).

Real-Time Decision-Making

Data analysis paves the way for decision-making. Stakeholders take critical decisions in real-time based on real-time analytics from the analytics engine, for example, important decisions like real-time pricing, on-demand renewable generation, estimating capacity constraints, forecasting demand and provisioning real-time supply that require real-time decisions. It helps to find the faulty sections in electric power grids and to take corrective actions to restore the grid back to complete functionality, consequently consumer confidence and revenue increases. Several real-time demand response (DR) algorithms facilitate timely and accurate decision-making.

Figure 5 shows an overview of the data and energy flows on the energy market including smart grid systems, prosumers and aggregators. Smart grids enable consumers to optimize their electric power usage and align it with their needs and, when appropriate, with their electric power generation and storage preferences and result in making profits (Parag and Sovacool 2016). Prosumers using photovoltaic (PV) or small hydro systems in a domestic household can produce electricity, and excess electric power from the prosumers can be stored, sold, or/and shared with neighbours in the same neighbourhood. Although, such energy from renewable sources is based on a non-continuous nature and is influenced by weather conditions.

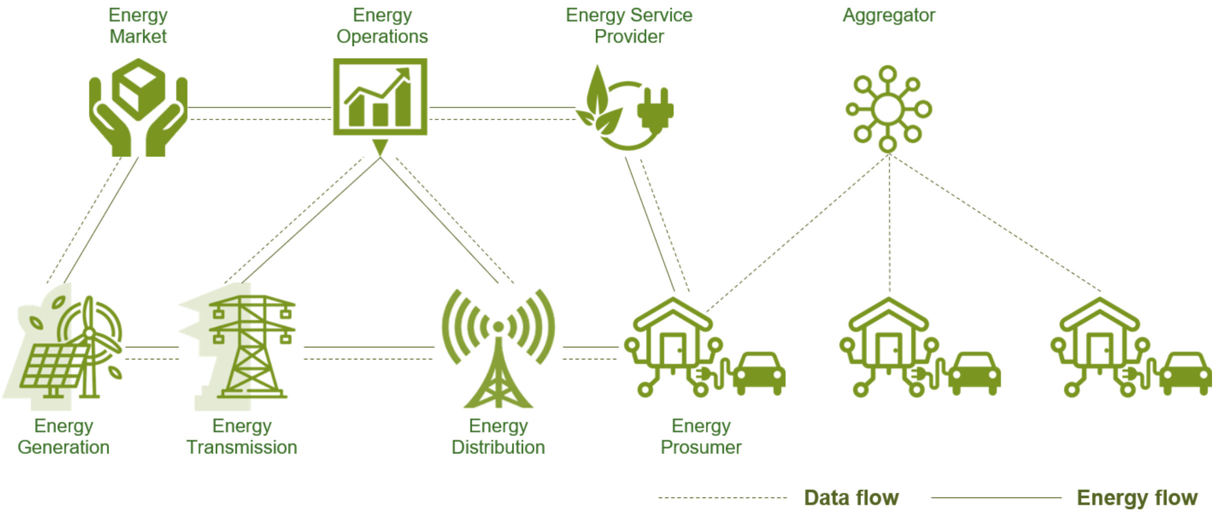

Figure 5 Energy and data flows on energy market, smart energy grid, prosumer and aggregator

Thus, the energy market is efficiently able to comprise grid-connected prosumers who are managed via an aggregator. The aggregator is responsible for analysing energy data flow and making decisions accordingly to manage the community groups of prosumers who mutually offload and use their generated energy or trade-off their surplus electricity to other energy buyers (e.g., individual consumers, energy retailers, or utility grid) (Bellekom, Arentsen and Gorkum 2016). However, the concept of prosumers and energy market and architectures modelling dynamic prosumers are still in its infancy (Anthony, et al. 2019).

On the other hand, today, the concept of micro-grid is being used for the purpose of helping the environment via several renewable and available resources, reducing generation costs and managing energy efficiently. The idea was begun with Advanced metering infrastructure (AMI) to develop demand management, increase energy efficiency and a self-repair electric grid, thus it enables to improve reliability and respond to natural disasters or deliberate sabotage (Dogaheh and Dogaheh 2017). In addition, it allows consumers to reduce their electric power consumption via proper scheduling of different appliances and real-time electricity pricing (RTEP). It might be achieved with smart energy home more effectively.

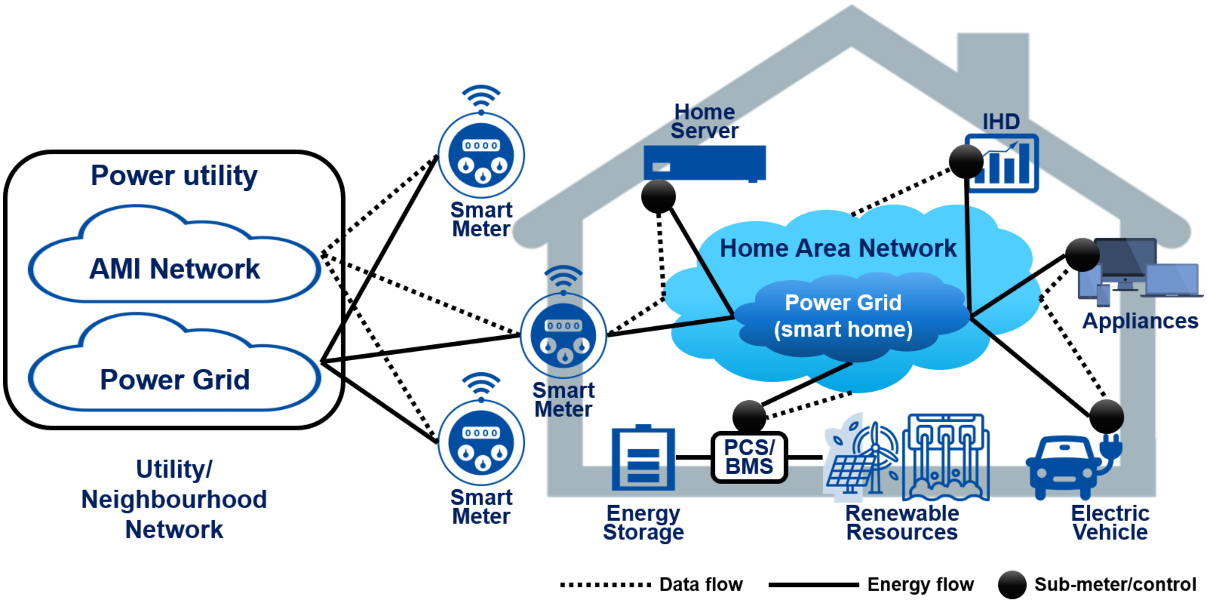

Figure 6 Energy and data flow on AMI network with smart home

Figure 6 illustrates the concept of AMI network and smart home (Home Area Network: HAN) and data and energy flows on these networks. HAN domain and the smart meter domain (AMI) enable consumers to monitor and control their electric power consumption profile and of their appliances via an in-home display (IHD) which consists of a computer, tablet, or smartphone. Smart meters include many interconnected smart meters (i.e., AMI) which installed and monitored by utility companies in order to transmit load information and demand-requirement signals between the smart homes and energy market. In addition, as afore-mentioned, energy produced from renewable resources can be stored by power conversion and battery management systems and consumed by the prosumers for an EV or their appliances. Remained energy will be transmitted to the energy grid through a smart meter.

6. Big Data Challenges in Electric Power Sector

6.1. Main challenges

There are several data related challenges in the electric power sector. It is an imperative task to extract business-critical intelligence and insights from large volumes of data in a complex environment of legacy diverse systems and fragmented and decentralised solutions that are common in the sector. Generally, companies are concerned with challenges associated in managing substantial complexity of data such as data coming from their hydroelectric power-plants, including but not limited to — vibration, temperatures, water levels, voltages, power levels and several other units. Also, data is growing exponentially in the form of both structured as well as unstructured data. The key challenges and possible approach to tackle these challenges are given in Table 5.

Table 5 Big data challenges

Challenges	Approach
Data from different sources (structured, unstructured & real-time)	Leverage the power of Hadoop, NoSQL databases for scalable information management systems in batch and near-real time streams to fulfil need for homogenous, integrated and perspective-based information
Huge volume of domain specific information embedded in each data cluster	Agile Big Data techniques, distributed processing, data mining
Use of different software products for data interpretation and decision making	Agile Big Data techniques for consistent asset models, optimized operational expenditure (OPEX) and capital expenditure (CAPEX), effective monitoring and integration between operation and business system
Difficulty in using data to respond to user needs quickly and efficiently	Analysing productivity, planning, uncertainty in delivery of energy and managing storage
Huge expenses on data management, handling streams of often incompatible data	Empower consumers with web, mobile-enabled dashboards by easy slice and dice of data, planning innovative services and predictive risk modelling

6.2. Other challenges

Following are the challenges related to data collection, analytics, energy services, applications and relevant regularisations:

Data integration

As we mentioned in the previous section, smart grid data are gathered from various sources such as sensors, actuators, transformers and other electrical and computing infrastructure that forms part of the grid ecosystem. Such data needs to be integrated with other data to provide situational awareness and assist in decision-making. Data semantic approaches need to be investigated to address the data integration issue. Therefore, the collected data needs to be semantically tagged to facilitate interoperability among different sources. In addition, semantic technologies can aid to solve problems of electric power Big Data integration. For example, the nature of electric power Big Data can be studied from an integration perspective, the value of data originating from and the structure of different sources may be different. Hence, integrating needs to happen spatially and temporally. Schema-level mapping approaches should also be further investigated to address the upcoming challenges of electric power Big Data.

Bad data detection

In many cases, bad data poses a massive problem because information systems for smart grids are interconnected, and the data flow through each of these systems. Such polluted data should early be detected and prevent the pollution distributing and infiltrating into other systems. Thus, detecting bad data becomes a critical research challenge that requires immediate attention. The data also could enter the systems when an attacker tries to attack by purposely injecting false or incorrect data. From a security point of view, it is a critical challenge because incorrect data will lead to incorrect decision-making (e.g., excessive energy generation or low tariffs). In this way, bad data has a significantly negative impact on smart grid operations. Hence, it should be dealt with extreme caution (Kosut, et al. 2010) (Xie, Mo and Sinopoli 2011) (Tajer 2017).

Standards and interoperability

The smart grid consists of various types of devices, networks, management software like SCADA and a variety of electric power Big Data and results in various communication devices technologies used in the background of the energy grids. Such communication systems or devices have a different feature, communication and processing speeds, and distinct data transfer mode like parallel series. Thus, data transfer becomes a hard task for these devices. Therefore, an interoperability mechanism is required to make the smart grids more flexible and efficient. In addition, there are needs for open standardisation for better interoperability and to ensure secure and sustainable smart. Some of the standardisation used in smart grids such as IEEE 1815, IEEE 2030.5, IEC 61850, IEC 61850-90-7 (Potdar, et al. 2018).

Big data knowledge representation and processing

Big data analytics requires machine learning and artificial intelligence techniques. It is often known that the process and outputs from such techniques lack intuitive physical interpretation (Wagstaff 2012). Thus, it is important to fill this gap by offering suitable domain knowledge interpretation to make a correct decision based on the intelligence derived from those techniques. This task might be also challenging because of the Big Data characteristics of energy data.

Big data security and privacy

Smart energy Big Data might often contain individuals' private information which requires to be protected under various legal regulations (Powner 2011) (Simitis 1994). For example, smart grids and meters collectively generate various data instances that have privacy and security concerns. The data also contain sensitive information that could be used to make decisions affecting the safe operation of the critical infrastructure of an organization or institution. Therefore, security and privacy are very crucial issues. However, it is also very challenging due to the Big Data nature of the smart energy data, distributed and open environment of the infrastructure. On the other hand, from a personal user privacy perspective, consumers will only trust a technology if they know that their data is protected and does not impact their privacy. Such consumer sensitive data needs to be securely transmitted to protect their privacy. Furthermore, for secure and safe data transfer, all the data must be encrypted before transferring it.

Scalable and interoperable computing infrastructure

A smart grid is usually a highly distributed system (Hu and Vasilakos 2016). Energy Big Data is gathered from every corner including electric power generation, distribution, renewable energy powered vehicles and smart meters and so on. Such data includes dynamic streaming and non-streaming data, structure and unstructured data. Also, there is a constant flow of the data between machines and humans. It is very challenging to manage, store, share, process and analyse such data. Thus, A scalable and interoperable computing infrastructure is required.

Cost optimisation of smart grid

One of the particular areas attracting attention from smart grids is studying the cost of data management. Even though there are several studies on the cost of data centres and how to cost-effectively manage large data centres, specific studies on data management cost of smart grids are missing. The smart grid is a complex network, and data is one of the most important elements of the smart grid ecosystems. Thus, it is one significant research challenge (Potdar, et al. 2018).

Real-time Big Data intelligence

According to (Huang, et al. 2014), overlapping of peak demand from individual homes may cause blackouts at some substations within several seconds. And it results in hugely increasing damages in terms of cost and human resource over time. Therefore, decision making in real-time is essential for both system operation and real-time pricing. An intelligent decision making needs to analyse current and historical data. Given the huge volume and high variety of such data, it is challenging enough to process them. With the constraint of real-time requirement, it will be extremely difficult to design new algorithms that can provide real-time intelligence from electric power Big Data.

Distributed and parallel intelligence

The electric power sector is experiencing a data explosion collected from distributed sources from smart grids. A distributed and parallel intelligent approach can effectively address this problem, which can also reduce the raw data accumulation and communication significantly. Although existing aggregation or summarization methods can also achieve the same aim of reducing large raw data, such methods are targeting local raw data without considering overall system target. Furthermore, they can lose very useful information which would be needed for specific applications. A proper distributed intelligence approach should be built upon a solid theoretical basis to approximate the relevant overall performance indicator. For instance, for anomaly detection, distributed local intelligence based on the observations from a single node or several neighbouring nodes should be enabled to estimate the overall sensor network data probability density distribution (Cui, et al. 2012) (Yuan, Li and Ren 2011).

Quality-of-Service (QoS)

As aforementioned, the smart grid is a complex network of heterogeneous devices and distributed systems transmitting different types of data at varying intervals with different levels of speed. The QoS aspect becomes more critical during data transmission. If the captured data reaches the destination in time, the data becomes valid for other systems also. However, network congestion and other factors can cause delay resulting in a poor quality of service. It may even introduce errors or raise the complete loss of data. For example, it becomes a major concern as it may result in suboptimal power quality if it corrupts the smart meter data or impacts the data centre if the air-cooling systems fail. Therefore, the needs and urgency for developing QoS frameworks are paramount. Relevant challenges are as follows: (1) identifying and agreeing upon QoS metrics for data acquisition, transmission, storage and security, (2) defining and enforcing service level agreements, and (3) methods to monitor and implement QoS strategies (Sooriyabandara and Ekanayake 2010).

Demand response

The demand from electric power consumers has been increasing along with the emerging types of machines such as EVs and smart home furniture. Thus, there has been arousing the concern on the stability and reliability of electricity supply. However, since the traditional energy grid lacks real-time response between demand and supply, it might not meet these demands due to its inflexibility (Maharjan, et al. 2013). Therefore, DR is expected to be an aspect in the future smart grid (Jiang, et al. 2016). Demand Side Management (DSM) is one of the most extensive application fields of Big Data analytics in the electric power sector, ranging from consumer segmentation to dynamic pricing (Yang and Zhou 2015). Valuable knowledge can be discovered from the massive electric power Big Data collected in near real-time by IEDs. Such knowledge enables many demand-side decision-making and marketing strategies development (Zhou, Fu and Yang 2016). For example, electric power load forecasting (Park, et al. 1991) is an important research content in smart grid, which predicts future load demand through analysis of historical load data, weather data, and social factors, etc. Energy load classification (Yang and Shen 2013) (Chicco 2012) is the process to classify different load profiles into groups using the various clustering methods. The energy consumption patterns of different users can be identified by the classification, which can support the development of competitive marketing and energy strategies and personalized energy services. For instance, dynamic pricing (Oldewurtel, et al. 2010) (Chao 2011), also referred to variable pricing or real-time pricing, can guide the user's energy consumption behaviours and improve the reliability of power systems by different and appropriate pricing strategies.

Asset management

The electric power industry is one of the typical asset-intensive industries. Both the electric power companies and the energy grid enterprises often face many asset management problems, such as resource sharing, asset retirement monitoring, operation and maintenance management, and inventory management (Zhou, Fu and Yang 2016). Analytics of electric power Big Data enables the efficiency of asset management and collaborative operation. The data of production, operation, marketing and management can be integrated and shared to achieve throughout electric power generation, transmission, transformation, distribution, and consumption. For example, the massive sensor data gathered from energy system infrastructure combined with advanced Big Data analysis and visualization techniques can change the traditional ways of power system operation and maintenance. In addition, the risk and unnecessary expenses of manual operation can be reduced, and the reliability of energy grid systems can be improved. Also, real-time monitoring collection and analysis of energy consumption data can be carried out to reduce the risks of a power failure and grid collapse. Furthermore, the weather data is also important to enhance the reliability and stability of energy systems. Particular weather patterns discovered can be used to predict future outages and identify the problem positions or areas, thus leading to faster failure warnings and recovery resulting in reducing asset damage (Wigle 2014).

Predictive maintenance

With the advance of energy power, the companies and prosumers of the electric power industry have more and more turbines resulting in generating electric power Big Data. Moreover, the turbines are becoming bigger and more powerful. Thus, they need to add more resources for maintenance purposes, increasing operation and maintenance costs. A predictive maintenance system based on traditional technologies requires having one system in each power plant since they use neither private nor public cloud where centralise all the data. Thus, it is difficult to analyse electric power data gathered from different places, in a unique location. In addition, they need to hire qualified people to manage each predictive maintenance system. However, having all the data produced by the turbines placed in a central system involves a huge computational cost for being capable of processing all the information fast enough to inform a future failure. Furthermore, the notification has to arrive in time to repair it before the turbine' component breaks down. Another problem in traditional technologies is scalability.

Since the volume of data to be processed hugely increases, it is difficult to add more computational resources to handle it. In this regard, although the scalability might usually be provided by buying more powerful hardware (vertical scalability), it is very expensive and limits scalability (Canizo, et al. 2017).

Energy data policy and regulations

In (IqtiyaniIlham, Hasanuzzaman and Hosenuzzaman 2017), authors reviewed development policies and challenges faced by European smart grids that are supported by the European Union. Major challenges discussed are system integration, customer involvement, legal hurdles and technology development. From the integration perspective, different and distributed systems need to be able to share the data with each other to ensure seamless operation. Regarding this, one of the technical barriers being the data models along with the metering infrastructure and standardisation of communication protocols. In addition, compatible storage and decentralised distribution and management are critical elements to ensure successful operations (IqtiyaniIlham, Hasanuzzaman and Hosenuzzaman 2017).

7. Recommendations

For the regional electric power industries, it is important to emphasis the application of Big Data technology and processes instead of designing and developing essential technologies by itself. Key prerequisites for data-driven innovation are access to data, ownership and security risk. It is important to achieve a balance between the interests of regional suppliers and operators when it comes to ownership of, access to and responsibility for data. The recommendations proposed in the following lines are to guide regional electric power suppliers and operators about how to utilise Big Data to manage day-to-day operations.

Data collection and governance

Although the volume of electric power Big Data is large and the electric power Big Data contain a lot of valuable knowledge, their value and data quality might be not so high in most cases. The timeliness, integrity, accuracy and consistency of electric power Big Data need to be improved. It also requires complete data governance strategies, organisation and control procedures. The prerequisites of many electric power Big Data-intensive applications are the high quality of standardization and format uniform.

Data integration and sharing

Currently, there are still many barriers to integrate and share electric power Big Data from various sources. Various standards and models of data definition, storage, and management are often adopted by different electric power businesses, and there are also many redundant data collection and storage. On the one hand, a lack of accessible data suffers researchers that are working on smart energy Big Data management. Recently, there have been some initiatives on electric power Big Data integration and sharing, such as Green Button Data[5] and WikiEnergy.

Data processing and analysis

Traditional data analysis techniques in data mining, machine learning, statistical analysis, and data management and visualization may encounter difficulties in dealing with the electric power Big Data. Effective and efficient processing and analysis techniques for electric power Big Data are the premise and important support for smart energy management. In addition, modelling and simulation always involve a huge amount of data and a lot of parameters including spatial and temporal granularities.

[5] Green Button Data, http://www.greenbuttondata.org/

Regarding the different modelling elements and parameter settings at multiple scales, dimensions, modelling and simulation results should be properly interpreted to support the various decision makings.

Security and privacy

The security mechanism of the IT infrastructure for smart energy systems needs to be further improved. Also, protecting private and sensitive customer data is a key issue in electric power Big Data analytics along with General Data Protection Regulation (GDPR). Consumers should have the right to own their data, and their personal data such as household electricity usage should be protected. Also, it has to be only used as the consumer allows. Industry self-regulation, technical means, and strengthened legislation should all be considered to enhance data security and privacy.

Information technology infrastructure

The explosive growth of electric power Big Data and the speed requirement for collecting, processing and utilising electric power data have brought a series of challenges for traditional IT infrastructure. The infrastructure needs to be improved in the capabilities of data storage, data processing, data interaction, data exchange, data visualization and network transmission to better support Big Data-driven smart energy management.

Professionals of Big Data analytics and smart energy management

Energy sector with Big Data is becoming more and more a multidisciplinary field. Thus, cooperation among the energy experts, data scientists, IT professionals, engineering specialists and management experts are essential for the regional (smart) energy sector. Big data analytics on smart energy management is a relatively new field, and professionals in these areas are still lacking. Therefore, courses and programs in management science, data science, energy science, computer science and social science should be encouraged and developed to train specialists that qualify for the various jobs in the electric power sector.

8. Conclusion

With the advance of Big Data technologies, there are many opportunities and challenges when we adopt and apply the technologies to various industrial sectors. It also applies to the electric power sector. There are a variety of electric power data sources in the regional power sector. To encourage the effective and efficient utilisation of Big Data technologies in power industries in the Sogn og Fjordane region, this *Teknoløft* project report analysed relevant Big Data technologies and investigates pertinent opportunities and challenges. We considered the data lifecycle in the knowledge discovery process and decision making by conducting pertinent literature review. Also, flows of electric power Big Data in energy grids and smart home networks are illustrated.

Finally, five major open issues and fourteen other challenges on the utilisation of Big Data in the regional electric power business are presented. Six recommendations are suggested to deal with identified issues and challenges. The recommendations are provided in order of priorities. As a next step, the recommendations can be converted into small and/or medium scale projects for actual business in the Sogn og Fjordane region.

9. References

Akerkar, Rajendra, Guillermo Vega-Gorgojo, Grunde Løvoll, Stephane Grumbach, Aurelien Faravelon, Rachel Finn, Kush Wadhwa, Anna Donovan, and Lorenzo Bigagli. 2014. *Understanding and mapping big data.* BYTE project.

Akhavan-Hejazi, Hossein, and Hamed Mohsenian-Rad. 2018. "Power systems big data analytics: Anassessment of paradigm shift barriers and prospects." *Energy Reports* 91-100.

Anthony, Bokolo, Sobah Abbas Petersen, Dirk Ahlers, John Krogstie, and Klaus Livik. 2019. "Big data-oriented energy prosumption service in smart community districts: a multi-case study perspective." *Energy Informatics.*

Bellekom, Sandra, Maarten Arentsen, and Kirsten van Gorkum. 2016. "Prosumption and the distribution and supply of electricity." *Energy, sustainability and society.*

Canizo, Mikel, Enrique Onieva, Angel Conde, Santiago Charramendieta, and Salvador Trujillo. 2017. "Real-time predictive maintenance for wind turbines using Big Data frameworks." *In 2017 IEEE International Conference on Prognostics and Health Management (ICPHM).* 70-77.

Cao, Zijian, Jin Lin, Can Wan, Yonghua Song, Gareth Taylor, and Maozhen Li. 2017. "Hadoop-based framework for big data analysis of synchronised harmonics in active distribution network." *IET Generation, Transmission & Distribution* 3930-3937.

Chao, Hung-po. 2011. "Efficient pricing and investment in electricity markets with intermittent resources." *Energy Policy 39, no. 7 (2011):* . 3945-3953.

Chen, Wen, Kaile Zhou, Shanlin Yang, and Cheng Wu. 2017. "Data quality of electricity consumption data in a smart grid environment." *Renewable and Sustainable Energy Reviews* 98-105.

Chicco, Gianfranco. 2012. "Overview and performance assessment of the clustering methods for electrical load pattern grouping." *Energy* 68-80.

Chu, Xu, and Ihab F. Ilyas. 2016. "Qualitative data cleaning." *Proceedings of the VLDB Endowment 9.* 1605-1608.

CSEE Committee, Electric Power Information Special. 2013. "White paper of China's electric power big data."

Cui, Shuguang, Zhu Han, Soummya Kar, Tung T. Kim, H. Vincent Poor, and Ali Tajer. 2012. “Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions.” *IEEE Signal Processing Magazine* (IEEE Signal Processing Magazine) 106-115.

Daki, Houda, Asmaa El Hannani, Abdelhak Aqqal, Abdelfattah Haidine, and Aziz Dahbi. 2017. “Big Data management in smart grid: concepts, requirements and implementation.” *Journal of Big Data* 1-19.

Diamantoulakis, Panagiotis D., Vasileios M. Kapinas, and George K. Karagiannidis. 2015. “Big data analytics for dynamic energy management in smart grids.” *Big Data Research* 94-101.

Dogaheh, Z., and M. Dogaheh. 2017. “Optimizing the performance of smart grids in relation with residential energy centers equipping with Solar Power Units (PV).” *Sindhological Studies* 14-22.

Eltantawy, Ayman B., and M., A. Salama. 2014. “A novel zooming algorithm for distribution load flow analysis for smart grid.” *IEEE Transactions on Smart Grid* 1704-1711.

Fehrenbacher, K. 2012. *10 ways big data is remaking energy.* 29 January. Accessed June 22, 2020. https://gigaom.com/2012/01/29/10-ways-big-data-is-remaking-energy/.

Guerrero, Juan Ignacio, Antonio García, Enrique Personal, Joaquín Luque, and Carlos León. 2017. “Heterogeneous data source integration for smart grid ecosystems based on metadata mining.” *Expert Systems with Applications* 254-268.

Hopkins, Brian, Boris Evelson, B. Hopkins, B. Evelson, S. Leaver, C. Moore, A. Cullen, M. Gilpin, and M. Cahill. 2011. *Expand your digital horizon with Big Data.* Forrester.

Hu, Jiankun, and Athanasios V. Vasilakos. 2016. “Energy big data analytics and security: challenges and opportunities.” *IEEE Transactions on Smart Grid* 2423-2436.

Huang, Zhichuan, Hongyao Luo, David Skoda, Ting Zhu, and Yu Gu. 2014. “E-sketch: Gathering large-scale energy consumption data based on consumption patterns.” *In 2014 IEEE International Conference on Big Data (Big Data).* 656-665.

IBM. 2014. *Managing big data for smart grids and smart meters.* IBM.

IqtiyaniIlham, Nur, M. Hasanuzzaman, and M. Hosenuzzaman. 2017. “European smart grid prospects, policies, and challenges.” *Renewable and Sustainable Energy Reviews* 776-790.

Jiang, Huaiguang, Jun J. Zhang, Wenzhong Gao, and Ziping Wu. 2014. “Fault detection, identification, and location in smart grid based on data-driven computational methods.” *IEEE Transactions on Smart Grid* 2947-2956.

Jiang, Hui, Kun Wang, Yihui Wang, Min Gao, and Yan Zhang. 2016. “Energy big data: A survey.” *IEEE Access* 3844-3861.

Kosut, Oliver, Liyan Jia, Robert J. Thomas, and Lang Tong. 2010. “Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures.” *In 2010 first IEEE international conference on smart grid communications.* 220-225.

Maharjan, Sabita, Quanyan Zhu, Yan Zhang, Stein Gjessing, and Tamer Basar. 2013. “Dependable demand response management in the smart grid: A Stackelberg game approach.” *IEEE Transactions on Smart Grid 4, no. 1 (2013):* . 120-132.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.

Moradi, Mohammad H., Mohsen Eskandari, and S. Mahdi Hosseinian. 2014. “Operational strategy optimization in an optimal sized smart microgrid.” *IEEE Transactions on Smart Grid* 1087-1095.

Mosaddegh, Abolfazl, Claudio A. Cañizares, Kankar Bhattacharya, and Hongbing Fan. 2016. “Distributed computing architecture for optimal control of distribution feeders with smart loads.” *IEEE Transactions on Smart Grid* 1469-1478.

Nafi, Nazmus S., Ahmed Khandakar, Mark. A. Gregory, and Manoj Datta. 2016. “A survey of smart grid architectures, applications, benefits and standardization.” *Journal of Network and Computer Applications* 23-36.

Nakafuji, Dora, Laura Rogers, Jared Bestebreur, Michael Rourke, and Greg Zweigle. 2017. “Integrating synchrophasors and oscillography for wide-area power system analysis.” *2017 70th Annual Conference for Protective Relay Engineers (CPRE).* 1-10.

Neumann, Scott, Frank Wilhoit, Margaret Goodrich, and VSK Murthy Balijepalli. 2015. “Everything's talking to each other: smart meters generate big data for utilities and customers.” *IEEE Power and Energy Magazine 14*, 40-47.

Oldewurtel, Frauke, Andreas Ulbig, Alessandra Parisio, Göran Andersson, and Manfred Morari. 2010. “Reducing peak electricity demand in building climate control using real-time pricing and model predictive control.” *In 49th IEEE conference on decision and control (CDC).* IEEE. 1927-1932.

Parag, Yael, and Benjamin K. Sovacool. 2016. “Electricity market design for the prosumer era.” *Nature energy* 1-6.

Park, Chan-Kook, Hyun-Jae Kim, and Yang-Soo Kim. 2014. “A study of factors enhancing smart grid consumer engagement.” *Energy Policy* 211-218.

Park, Dong C., M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg. 1991. “Electric load forecasting using an artificial neural network.” *IEEE transactions on Power Systems* 442-449.

Pérez-Chacón, Rubén, José M. Luna-Romera, Alicia Troncoso, Francisco Martínez-Álvarez, and José C. Riquelme. 2018. “Big data analytics for discovering electricity consumption patterns in smart cities.” *Energies* 683-701.

Potdar, Vidyasagar, Anulipt Chandan, Saima Batool, and Naimesh Patel. 2018. *Big Energy Data Management for Smart Grids—Issues, Challenges and Recent Developments.* Springer.

Powner, David A. 2011. *Electricity Grid Modernization: Progress Being Made on Cybersecurity Guidelines, But Key Challenges Remain to be Addressed.* DIANE Publishing.

Qiu, Robert C., and Paul Antonik. 2017. *Smart grid using big data analytics: a random matrix theory approach.* John Wiley & Sons.

Refaat, S. S., H. Abu-Rub, and A. Mohamed. 2016. “Big data, better energy management and control decisions for distribution systems in smart grid.” *IEEE International Conference on Big Data.* IEEE. 3115-3120.

Sai, Xie, and Zhe Chen. 2017. “Anomaly detection and redundancy elimination of big sensor data in internet of things.” *arXiv preprint.*

Sánchez-Lozano, Juan M., Jerónimo Teruel-Solano, Pedro L. Soto-Elvira, and M. Socorro García-Cascales. 2013. “Geographical Information Systems (GIS) and Multi-Criteria Decision Making (MCDM) methods for the evaluation of solar farms locations: Case study in south-eastern Spain.” *Renewable and sustainable energy reviews* 544-556.

Simitis, Spiros. 1994. “From the market to the polis: The EU directive on the protection of personal data.” *Iowa Law Review* 445.

Sooriyabandara, Mahesh, and Janake Ekanayake. 2010. “Smart grid-technologies for its realisation.” *In 2010 IEEE International Conference on Sustainable Energy Technologies (ICSET).* 1-4.

Srinivasan, Dipti, and Thomas Reindl. 2015. "Real-time display of data from a smart-grid on geographical map using a GIS tool and its role in optimization of game theory." *In 2015 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA).* 1-6.

Stefan, Maria, Jose G. Lopez, Morten H. Andreasen, and Rasmus L. Olsen. 2017. "Visualization techniques for electrical grid smart metering data: a survey." *In 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService).* 165-171.

Tajer, Ali. 2017. "False data injection attacks in electricity markets by limited adversaries: stochastic robustness." *IEEE Transactions on Smart Grid* 128-138.

Wagstaff, Kiri. 2012. "Machine learning that matters." *arXiv preprint* arXiv:1206.4656.

Weng, Yang, Rohit Negi, Christos Faloutsos, and Marija D. Ilić. 2016. "Robust data-driven state estimation for smart grid." *IEEE Transactions on Smart Grid* 1956-1967.

Wigle, Lorie. 2014. *How big data will make us more energy efficient.* 2 5. Accessed 6 24, 2020. https://www.weforum.org/agenda/2014/05/big-data-will-make-us-energy-efficient/.

Xie, Le, Yilin Mo, and Bruno Sinopoli. 2011. "Integrity data attacks in power market operations." *IEEE Transactions on Smart Grid* 659-666.

Yang, Shan-lin, and Chao Shen. 2013. "A review of electric load classification in smart grid environment." *Renewable and Sustainable Energy Reviews* 103-110.

Yang, Shanlin, and Kaile Zhou. 2015. "Demand side management in China: The context of China's power industry reform." *Renewable and Sustainable Energy Reviews* 954-965.

Yuan, Yanling, Zuyi Li, and Kui Ren. 2011. "Modeling load redistribution attacks in power systems." *IEEE Transactions on Smart Grid* 382-390.

Zhou, Kaile, Chao Fu, and Shanlin Yang. 2016. "Big data driven smart energy management: From big data to big insights." *Renewable and Sustainable Energy Reviews* 215-225.

KLIMA OG MILJØ

REISELIV

TEKNOLOGI OG SAMFUNN

Sogndal
Fosshaugane Campus

85%
oppdragsfinansiert
Omsetning omlag 30 millionar

30+
medarbeidarar
25 forskarårsverk
Halvparten med doktorgrad.