Policy discussion – Challenges of big data and analytics-driven demand-side management

Dr. Elizabeth Massey [1] (lead author), Dr. Piyush Verma [2], Mr. Beto Altamirano [3], Mr. Andrei Covatariu [4], Mr. Romanas Savickas [5]

Introduction

1. In recent years, the energy sector has experienced a shift towards disruptive trends such as decarbonization, decentralization and digitalization, creating an energy transition that generates a major impact on the utility industry worldwide.

2. Technologies that are driving digitalization of the utility sector include Distributed Energy Generation in the form of Distributed Solar PV, Utility-scale storage, Electric Vehicles (EVs) and EV Charging Infrastructure, and the proliferation of Advanced Metering Infrastructure and smart meters.

3. Big data, however, is still a nascent research area in the electric utility industry due to a lack of resources and expertise, whilst in other industries, such as online commerce and telecommunications, big data research is developing as fast as the technology that supports it.

4. As a result, new business models, utility capabilities and consumer commitments, especially on the demand side, will be enabled by these emerging technologies. With proper research funding support, the utility industry can realize international collaboration and fair competition in this technology space.

5. Key objectives of this paper are:

(a) Review the current challenges of big data analytics within the context of distribution grid / demand-side management.

(b) Describe policy gaps or barriers to the progress of advanced analytics in the electric utility sector.

(c) Identify key questions that deserve further analysis to address the challenges, gaps, and barriers to progressing state-of-the-practice for utility demand-side advanced analytics and advanced demand-side management.

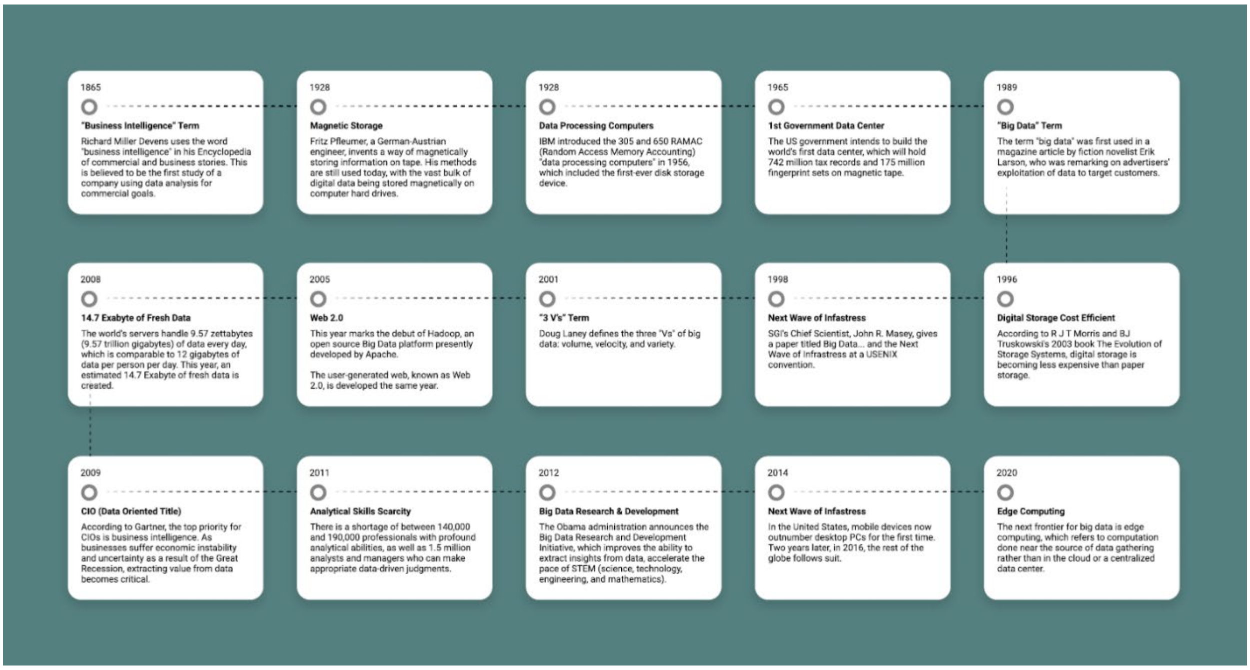
Context

6. Beginning in 1985 when the term, ‘Business Intelligence’ was coined, the term Data Analytics grew out of that work and has been a core part of the last century of evolution in the computing field. Today the term Data Analytics is used in nearly every industrial and commercial sector (Figure I).

7. Advanced analytics represents capabilities to manage and analyse data, typically going further than those of traditional Business Intelligence (BI). In this context ‘big data analytics’ is the examination of a set of data using algorithms and other sophisticated modelling and statical analysis techniques to produce actionable insights from this data. A related term ‘advanced analytics’ is often described as the use of predictive and prescriptive approaches to advance those insights into action.

[1] Director, Connected Analytics.
[2] Co-Chair of the Group of Experts on Energy Efficiency, Chair of the Task Force on Digitalization in Energy; Harvard Kennedy School.
[3] Harvard Kennedy School.
[4] Senior Research Associate, Energy Policy Group.
[5] Vice-Chair of the Group of Experts on Energy Efficiency; Senior Advisor, UNEP-CCC, Copenhagen Climate Centre.

Figure I
100 Years of Data Analytics



Source: developed by the authors

8. As the capabilities of computers and computing power have grown the amount of data collected, stored and processed on a daily basis has increased. The growth of data centers, advent of the internet (and the World-Wide-Web) means that 2.5 quintillion bytes of data are created every day (2.5 trillion GB).

9. Although the scientific community has no standard definition for big data; indeed, there are between 3 and 10 characteristics of big data[6]. In this context the term '*big data*' is defined as extremely large, heterogenous data sets from a variety of new data sources that traditional data processing software cannot handle.

Challenge 1 – Data Sharing and Democratization of Data

10. For data to be made widely available it must be shared amongst many stakeholders. Issues of cybersecurity, confidentiality, ownership, and privacy concerns need to be resolved for this to be realized. Translations between countries and regions are an issue. Timely and complete access of relevant consumption and customer data to non-generating or non-asset owning utilities (such as Community Choice Aggregators (CCAs) in California) is a challenge not yet resolved, especially in areas where de-regulation is relatively new and the local regulations on data sharing have yet to catch up with the changing regulatory landscape.

11. Starting with the understanding that 'Democratization of Data' is the ongoing process of enabling all stakeholders, regardless their level of technical knowledge, to work with data effectively, and to make informed decisions based on that data. The following key questions are relevant to explore this challenge further.

Data Curation

12. Data Curation is the process of collecting, organizing, characterizing, cleaning, enhancing, optimizing, and preserving data. Data that is optimized for analytics use needs a structure which focuses on leveragability of that data and the algorithms. The questions here

[6] See: https://unece.org/sed/documents/2021/06/working-documents/improving-efficiency-buildings-through-digitalization

focus on identifying a standard by which energy usage, digital asset and network information data be classified and organized.

13. Personally Identifiable Information (PII) is defined as, 'Any representation of information that permits the identify of an individual to whom the information applies to be reasonably inferred by either direct or indirect means.' Whilst there is agreement across the industry on the general definition of PII, there is a variety of ways to secure, handle and use PII. The challenge remains on how customer analytics can be truly actionable with limited information.

1. *What regulatory guidelines exist for how potentially sensitive data collection is prioritized and on what criterion?*
2. *What standards or guidelines exist to determine which energy data needs to be collected and which data needs to be kept private? What collection processes will preserve the accuracy of the data?*
3. *What efforts are currently being made in and among countries to drive improvements for data curation?*

Data Availability

14. When it comes to the utility sector, Data Availability refers to the process where utility companies (both distribution systems operators and supply companies, in geographies with unbundled services) and their users have continuous, secure, and readily usable data, associated with their electricity use. As the energy sector has had a high inertia in adopting digital technologies, as well as in harvesting and managing the data output, data availability is relatively poor.

15. If big data is not readily available, and not accompanied by power system data, the opportunity for advanced data analytics is dramatically reduced and often made infeasible. A variety of Meter Data Management (MDM) vendors exist as do customer-oriented usage data technologies. Often these systems are incompatible due to proprietary software and data formats.

1. *Which data should be shared (raw/pre-post-processed) and how broadly (among national regions, internationally)?*
2. *How can the customer's usage data be made more approachable to bring customers into the energy transition conversation? Wi-Fi-enabled appliances with incentives.*
3. *What standardized, cleaned and analytics-ready synthetic datasets are publicly available for research, pilot program design or testbed definition?*
4. *How is the continuous process of data sharing currently being managed and what is the cost?*

Data Integration and Legacy Systems Management

16. As data management systems and technologies were deployed in various moments in time, by various vendors, using different software architectures, data integration between different energy systems is limited.

17. Additionally, utility companies are still using outdated hardware and software that are, however, functioning. That leaves older technologies isolated from new platforms, as they are unable to interact with new generation of hardware/software. The energy sector, in particular, faces these issues, as the investments in digitalization are mainly direct toward new systems and technologies, rather than upgrading old ones.

1. *What standard processes exist for integration of heterogeneous data warehousing systems from multiple vendors?*
2. *What standards exist for interoperability between systems?*
3. *What standards exist for low-cost integration of fragmented demand-side systems and services from multiple electricity operators, providers, and vendors?*

Challenge 2 – Utility Analytics Sector Skills Availability

18. As one of the most prominent IoT applications for utilities, Advanced Metering Infrastructure (AMI) and Smart Meters provides benefits to utilities in both Operational- and Customer-focused areas. Even during the COVID-19 pandemic, global deployment of AMI systems and digital meters has continued and even increased. The number of installed meters is expected to exceed 227 million units in 2026 in the EU-27+3 region (from 150 million units in 2020) and yearly shipments of smart electricity meters in North America will grow from 8.8 million units in 2019 to 19.9 million units in 2024. The penetration of smart meters in Asia-Pacific stood at 69 percent in 2019 and is expected to grow to 82 percent in 2025. The 10 fastest growing markets during 2020-2026 will all be in Central Eastern and South-Eastern Europe. Connected technologies of synchro phasor measurement unites (PMUs), supervisory control and data acquisition (SCADA) systems along with AMI are core enabling technologies for the smart grid. Deployment of these systems means that utilities can harness the power of remote metering for connection/ disconnection services, outage avoidance, and energy monitoring highly granular view on grid status and operations. Whilst installation of these systems can have relatively low initial costs, there are high maintenance costs; not the least of which are the properly skilled (experience) and trained (education) personnel.

19. Many utilities do not have their own analytics department, and there is a growing need to collaborate among utility departments to determine whether algorithms are reaching the right conclusions. There is a gap between utilities that know their business and data analysts that know the algorithms. All utility segments would benefit from working together so that algorithms and datasets are not duplicated.

20. Large-scale test beds are needed to evaluate various solutions with the continued proliferation of smart and connected IoT devices, adding to the education and skills concerns for utilities.

21. Challenges around the up skilling of the current workforce to effectively use tools and techniques currently available and drive improvements in education require further research into the following key areas.

Data Translation into Operational Needs

22. Data translation can be defined as the process of converting volumes of data from one syntax to another and performing value lookups or substitutions from the data during the process. Translation can include data validation as well. One example of data translation is to convert vendor specific timeseries data or even GIS and customer flat files while performing data validation on the source data. Translating data into operational ambitions requires a strategic vision on which objectives can be established. Measurable objectives can create actionable insights.

1. *Once made available, big data needs to be judiciously integrated to allow for cost-effective management and utilization. What operational and customer needs use cases can be standardized across the utility industry to be shared?*
2. *How can these shared use cases be prioritized through a standard decision-making process based on criteria such as impact or risk?*
3. *What educational models allow for faster recruitment of skilled talent that can quickly extract value from data analytics within the utility industry, demand-side context?*
4. *To implement big data analytics, legacy utility solutions that handle large amounts of data must be completely redesigned. What efforts can facilitate the modernization of legacy systems?*
5. *Because solution providers and customers may have different backgrounds and expertise, the need for a common vocabulary when making big data decisions is critical. What operational and educational models could help standardize language across the utility industry?*

Data Monetization

23. To a large extent, data monetization is the ideal goal in achieving widespread deployment of data analytics. Utilities and energy service providers have a wealth of customer and operational insights to provide new revenue generating products and services and enhance product and operational performance to create a more compelling and sustained customer relationship. Data monetization is the stage of data maturity where utilities and other energy supplier companies leverage big data for new revenue opportunities. For example, leveraging actionable insights from user data and customer behaviours can drive a utility to upscale their customer relationship and rethink their customer experience.

1. *What licensing approaches, for the use of data with mutual benefits for the provide and user, have been standardized?*
2. *What measures are considered that balance a level of product safety with minimal increase in barriers to markets where utilities can monetization their data?*
3. *What can consumers get out of and in exchange for their data and other grid data?*

Cybersecurity and Grid Resiliency

24. As the utility sector is gradually increasing its level of digitalization, increasing risks related to cybersecurity emerge, both operationally and commercially. In this context and considering the current geopolitical context, (e.g., the higher risks of cyberwarfare), utility companies must set proper prevention and mitigation strategies, while also developing business continuity plans after cybersecurity breaches.

1. *What support for best practice or state-of-the-art methods or techniques are being used to prevent successful cyber-attacks on the grid?*
2. *Most utilities have models for grid monitoring and identification of natural and adversarial threats. Expansion of these efforts is encouraged. What best practices in data analytics and cyber security based on artificial intelligence (AI) and machine learning can be shared among utility segments?*
3. *What scenarios or use cases can be combined to optimize the results of the analysis?*
4. *What test-bed models can be shared and standardized to enhance cross-sector collaboration?*
5. *Considering the low level of data curation as well as operability, can cyber threats delay a standardization across utility companies even more? Can these risks also increase costs for increasing interoperability?*
6. *What key cybersecurity threats are at highest risk to affect the energy sectors' data systems, be it on a generalized level or on a more specific one?*

Challenge 3 – Big Data Analytics Modelling R&D efforts

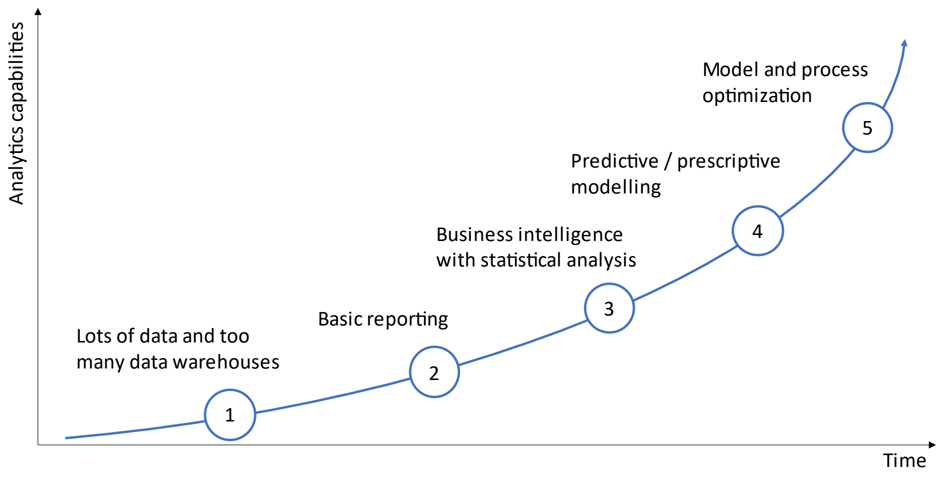
25. Access to on-going national and sub-national level research and application efforts for utility-scale data analytics in the utility industry. Whilst most countries have a national or regional energy 'ministry' (e.g., Department of Energy (United States), Directorate General Ener (EU), Ministry of Power (India), International Energy Agency, etc.), getting access to and using the results from the research efforts of these large bodies is often difficult.

26. Data maturity is an important concept for utilities. As utilities find that data is more central to all enterprise strategies, drives innovation, and is integrated across departments, a strategy for a data maturity roadmap is a necessary priority. It is a way for utilities to optimize their internal processes as well as innovate over and provide reliable services for their customers.

27. The graphic in Figure II below shows a Data Capabilities Maturity Curve. As the curve moves from the left to the right a company increases its data and algorithmic capabilities. Studies show that companies with mature data wrangling and analytics processes can boost their profitability by an average of 12.5 percent of total gross profit.[7]

[7] See: https://www.agilitypr.com/pr-news/public-relations/whats-your-data-really-worth-it-depends-on-your-data-maturity-level/

Figure II
Data Maturity Curve



28. Challenges around topics such as automation of analytics, cost recovery, and model adoption require exploration into the following key questions.

Data Analytics Model Availability

29. Due to the declining costs of information and communication technology, as well as advances in computing power, lead to an increasing availability of data and new opportunities for its analysis. Data availability is about the timeliness and reliability of access to and use of relevant data. As the number of renewables and other distributed technologies continue to increase penetration to the global grid, increase complexity across the electricity system and create new needs for data analytics and optimized analytics models. Data and data analytics model availability typically have a time limit.

1. *What best practices or industry standards exist to provide utilities guidance for curation of state-of-the-art advanced analytics models?*
2. *What best practices or state-of-the-art techniques are available for utilities to enhance and grow their in-house analytics teams?*
3. *What types of models could be shared to promote standardized practices throughout the energy community, including academia, industry, national laboratories, and policymakers?*

Big Data, Advanced Analytics Model R&D Efforts and Outreach

30. As research in big data analytics in the electricity sector is growing in breadth and diversity it is essential to integrate and structure the fragmented body of scientific work. Currently, data analytics activities span the areas along the entire value chain, from generation and trading to transmission, distribution and consumption. Activities also range over different applications such as forecasting or clustering using various approaches such as Artificial Neural Networks (ANNs).

1. *What efforts are countries making at the national or regional level to establish working groups or best-in-practice communities to allow broader access to state-of-the-art advanced analytics models?*
2. *What input data should be considered and what are the key assumptions?*
3. *What further data can help to improve the accuracy of the modelling and analysis?*
4. *What are the various trade-offs between demand side analytics models and how do these impact the model outcomes?*
5. *What are the hidden biases in the model and what kind of risks or uncertainties do these create?*
6. *Can the decision-making process for these trade-offs be automated to mitigate or eliminate bias?*

Conclusion

31. Based on the above stated challenges, the Task Force on Digitalization in Energy identifies 3 key areas for further consideration:

(a) Integrity – This is a systems' ability to ensure that the system and information is accurate and correct. The questions on Data Curation, Data Integration, Cybersecurity and Data Translation are key aspects of data integrity.

(b) Availability – This is a systems' ability to ensure that systems, information, and services are available as appropriate to the operational needs of the utility. The questions on Data and Analytics Model Availability, and Advanced Analytics Model R&D Efforts, Cybersecurity and Outreach fall into this category.

(c) Confidentiality – This is a systems' ability to ensure that only the correct, authorized users, systems and resources can view, access, change or otherwise use data. The key questions for this area are around Data Democratization, Cybersecurity and Grid Resiliency.

32. The Task Force further suggests the follow-on activities:

(a) Creation of a dedicated working group to investigate the key questions noted above and carry out comprehensive work and deeper analysis of each to extend the mandate of this task force.

(b) Focused research on funding models for those areas in greatest need of attention such as: big data technology advancement (e.g., natural language processing, digital twin modelling, demand / load forecasting, optimized ML, progression of AI capabilities), grid resiliency, Infrastructure investment particularly as it relates to data access, storage, management, and real-time analytics.

(c) Creation and maintenance of a common dictionary of terms in the area of Digitalization in Energy.