United Nations ECE/ENERGY/GE.6/2023/4–ECE/ENERGY/GE.5/2023/4

Economic and Social Council

Distr.: General
27 July 2023

Original: English

Economic Commission for Europe

Committee on Sustainable Energy

Group of Experts on Cleaner Electricity Systems
Nineteenth session
Geneva, 3-4 October 2023
Item 7 of the provisional agenda
Reliability and cyber resiliency of smart integrated energy systems

Group of Experts on Energy Efficiency
Tenth session
Geneva, 5-6 October 2023
Item 6 of the provisional agenda
Digitalization and energy system resilience

Improving efficiency and reliability of energy systems by means of Big Data analytics

Note by the secretariat

Summary

The energy sector has experienced a shift towards disruptive trends such as decarbonization, decentralization and digitalization, fuelling the energy transition that creates a major impact on the utility industry worldwide. These ambitions drive the imperative for Artificial Intelligence in general and more specifically, Data Analytics.

Deployment of Big Data analytics, machine learning, and Artificial Intelligence in utilities and energy providers is growing at a rate that may outpace the maturity of the organizations. In fact, organizations may have already engaged in advanced algorithmic deployment. Yet, without a strategy, organizations may not have developed workable plans for the curation of data, training datasets, or analytics results.

This document is the result of the collaborative work of the product of the Group of Experts on Cleaner Electricity Systems and the Group of Experts on Energy Efficiency, conducted on the platform of the Task Force on Digitalization in Energy. It is an expansion of the unofficial document GEEE-9/2022/INF.3 entitled “Policy discussion – Challenges of Big Data and analytics-driven demand-side management” that contains the original set of questions posed for further consideration by the Task Force on Digitalization in Energy in collaboration with the subsidiary bodies of the Committee on Sustainable Energy.

Mention of any firm, product, service or licensed process does not imply endorsement or criticism by the United Nations. The designations employed do not imply the expression of any opinion whatsoever on the part of the United Nations Secretariat concerning the legal status of any country, territory or area, or of its authorities.

I. Introduction

1. Technologies that are driving digitalization of the utility sector include integrated energy systems via distributed energy generation and consumption in the form of distributed solar photovoltaics (PV) and wind, utility-scale energy generation and storage, electric vehicles (EVs) and EV charging infrastructure, and the proliferation of smart grid with integrated advanced smart metering and other digitally controlled infrastructure and equipment.

2. All these new technologies generate data. The proliferation of these digital technologies indicates there will be an increase in the amount of data to be collected, managed and analysed. This trend towards Big Data creates new opportunities for expansive and robust decision support systems. Big Data and Artificial Intelligence (AI), however, are still nascent research areas in the energy utility industry due to a lack of resources and expertise, whilst in other industries, such as online commerce and telecommunications, Big Data and AI research are developing as fast as the technology that supports them. To start the integration of AI in the energy sector, the algorithms need historic datasets, at least of a couple of years' worth. Hence to make use of AI in the main, utilities need to start collecting data very early in the analysis process.

3. As a result, new business models, new smart integrated energy systems, utility capabilities and consumer commitments and behaviour, especially on the demand side, will be enabled by these emerging technologies. The upcoming new integrated energy systems with abundance of renewable energy sources (RES) and variable generation will meet the supply-demand balance only by the help of Big Data and AI. With proper research, funding, and policy support, the utility industry can realize international collaboration and fair competition in this technology space.

II. Context

4. Beginning 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 (see Annex). Today, the term Data Analytics is used in nearly every industrial and commercial sector. Several factors have led to the current focus on Data Analytics in the energy sector. The declining costs of information and communications technologies as well as advances in computing power all lead to an increasing availability of data and new opportunities for analysis (push factors). Additionally, the increasing transitory nature of renewable energy sources, and the dynamic nature of the offerings due to new actors constantly entering the market, increase the complexity and create new needs for Data Analytics (pull factors).[1]

5. Since 2005, AI is a topic that has seen a significant publication growth across all topics including Energy and Computer Science, until 2010 when Data Analytics in the electricity sector as a field of research surpassed even the AI-related literature.

6. AI has as many definitions as applications for its use. In this context, it is the leading technology that uses data analytics to automate the decision-making process around customer engagement strategies, optimize forecasts on energy use and energy flows for local generation and storage, enhance theft detection and fraud, trade commodities with higher prediction accuracy, and efficiently manage and secure the energy grid against cyberattacks before they happen.

7. An implementation of AI is Big Data analytics. This requires skills to curate, manage, and analyze the data. The role of a Big Data analyst typically goes further than those of traditional business intelligence analysts. 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

[1] Frederik vom Scheidt and others, "Data analytics in the electricity sector – A quantitative and qualitative literature review", *Energy and AI*, Vol. 1 (2020).

analytics' is often described as the use of predictive and prescriptive approaches (sometimes also referred to as AI) to advance those insights into action. The use of advanced analytics in the context of this document is around the measurement and management of grid events and customer demand. However, it is interesting to note that advanced analytics used in corporate settings can provide a return between 8 and 9 per cent reduction in operating costs (aspects of people analytics such as improvements in work-related accident investigations, management and prevention, recruitment, training, performance management, and employee retention).

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 centres, advent of the internet (and the world-wide-web) means that 2.5 quintillion bytes (2.5 trillion gigabytes, GB) of data are created every day.

9. Although the scientific community has no standard definition for Big Data, indeed, there are between 3 and 10 characteristics. The term Big Data is not only about the volume of data, but also refers to the high speed of transmission and the wide variety of information that is difficult to collect, store and process using the available classical technologies. 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 on a timely (near real-time) basis.

III. Identified challenges

A. Data sharing and democratization of data

10. Being connected online is increasingly becoming a daily necessity. Not only for the convenience, but also for access to necessary data and information needed to progress business ambitions. Connectivity, as a technology, is only the starting point, and in worse-case situations may exacerbate existing barriers or create new ones.

11. Data sharing and democratization of data are fundamental to the concept of digital inclusion that is defined as "equitable, meaningful, and safe access to use, lead, and design of digital technologies, services, and associated opportunities for everyone, everywhere".[2] In order for people to embrace new technologies and get the full benefit of them, these technologies, and their associated data sets, need to be useful and authentic.

12. For data to be made widely available it must be shared amongst many stakeholders, including those at the margins of an industry or society. Issues of cybersecurity, confidentiality, ownership, and privacy concerns need to be resolved for this to be realized. Translations between countries and regions are also an issue, namely language relevance. It is worth noting that access to training for the necessary digitalization skills is closely related to the proliferation of the local language on the world-wide-web. The main benefactor of these cross-cutting skills are the utilities and energy providers, whether or not the energy provider is the generator or the owner of the grid asset.

13. Timely and complete access of relevant consumption and customer data is a challenge not yet resolved, especially in areas where digitalization is declared as 'the next engine of growth'. The economies of Central and European countries are good examples where digitalization (and data sharing) can make global-level impacts, and indeed with the 2016 actualization of the European General Data Protection Regulation (GDPR) there are specific guidelines for how to handle various types of data. Yet, without greater cooperation and policy coordination amongst regions, securing and realizing the full benefits of digitalization and Big Data is still only an ambition.

14. Starting with the understanding that 'Democratization of Data' is the ongoing process of enabling all stakeholders, regardless of their level of technical knowledge, to work with

[2] See: https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Definition_Digital-Inclusion.pdf (accessed 7 May 2023).

data effectively, and to make informed decisions based on that data, there are still challenges that require deliberate consideration.

Data curation

15. 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.

16. As one of the most prominent Internet-of-Things (IoT) applications for utilities, Advanced Metering Infrastructure (AMI) provide 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 European Union (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.[3] The penetration of smart meters in Asia-Pacific stood at 69 per cent in 2019 and is expected to grow to 82 per cent in 2025. The 10 fastest growing markets during 2020-2026 will all be in Central, Eastern and South-Eastern Europe.[4]

17. Smart meter deployments are not restricted to electricity, however. For example:

(a) Regions that experience high water stress need innovative ways to manage and control water usage.[5] It is argued that 700 million water smart meters connections are expected by 2030, up from 196 million as at the end of 2021. Geographies with the largest expected deployments by 2030 include China (31 per cent of total share), North America (29 per cent) and Europe (28 per cent);[6]

(b) Smart gas meter deployments are also increasing as connection and connectivity technologies improve across commodities, as well as to support governmental policies to build infrastructure for the efficient distribution and use of both residential and industrial natural gas. Providing different readings than their water and electric counterparts, natural gas meters provide pressure, volume, and temperature of gas giving another perspective of usage within the connected premise. Common to the other commodities, natural gas meters can give readings on unexpected meter events, which can be correlated with the other commodity meter readings to give a holistic view of usage and status of the premise (and potentially for customer health and safety);[7]

(c) Connected technologies of phasor measurement units (PMUs), supervisory control and data acquisition (SCADA) systems concurrent with AMI are core enabling technologies for the smart grid and provide invaluable meta-data (‘data-about-the-data’). New connectivity technologies, including 5G massive Machine-Type Communications (mMTC), non-mMTC Low Power Wide Area (LPWA) technologies, and 4G Cellular are expected to replace RF mesh networks, the current primary communication technology.

18. Deployment of these systems means that utilities can harness the power of remote metering for connection and disconnection services, outage avoidance, and energy usage monitoring with a highly granular view on grid infrastructure and asset 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. The sheer volume of data created by these smart meters (and other IoT devices) creates an expanded attack surface that is increasingly challenging to monitor and secure. As an example, 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

---

3 Nicholas Nhede, “Smart meter penetration in North America will reach 81% by 2024”, Smart Energy International, 5 July 2019.

4 Berg Insight AB, “Smart Metering in Europe - 17th Edition”, October 2021.

5 World Resources Institute, “17 Countries, Home to One-Quarter of The World’s Population, Face Extremely High Water Stress”, 6 August 2019.

6 Transforma Insights, “Water Smart Meters: 700 million connections by 2030 to solve issues related to water scarcity and loss”, 17 August 2022.

7 Reports and Data, “Smart Gas Meter Market […] Forecast to 2028”, July 2023.

reasonably inferred by either direct or indirect means."[8] Whilst there is agreement across the industry on the general definition of PII, there are a variety of ways to secure, handle and use PII. Today's customers are keenly aware of the importance of data privacy and the need for personal protection as they move closer to their role in the transformation of the smart grid. Motivated by the need for public protection, multiple privacy regulations have recently been enacted, including GDPR and the California Consumer Protection Act (CCPA). These actions join longstanding data security provisions from other sectors such as health, finance and commerce.

20. In addition, data that organizations gather, process, and store as part of routine business operations but do not use for other purposes, also referred to as 'dark data',[9] represents an untapped potential for businesses. By applying the principles of making data findable, accessible, interoperable, and reusable (FAIR principles)[10] into the data curation practice, organizations can turn dark data into valuable assets that could be used to enhance grid efficiency and improve forecasting of energy supply and demand.

21. Within the context of energy customer demand analytics, customer segmentation, energy clustering and other demand-profiling efforts, analytics models require as much information as possible for discernment within the confusion matrix. The challenge remains on how customer analytics can be truly actionable if information is limited due to regulatory constraints over potentially sensitive data. As the energy sector moves along its digitalization trajectory, a balance needs to be struck between access to the necessary data for identifying and discerning energy demand needs across various demographic segments, and the privacy of data required to keep sensitive information from going public.

Data availability

22. 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 or energy use. Whilst 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. In other words, although new power plants built on the principles from digitalization guarantee greater efficiency and higher availability of services, and with the rise of digital twins which can help with modelling, forecasting, and testing for optimal performance, there is still a significant lack in cross-discipline and cross-sectoral data access.

23. 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 vendors exist as do customer-oriented usage data technologies. Often these systems are incompatible due to proprietary software and data formats.

24. The FAIR Guiding principles for scientific data management and stewardship intend to provide guidelines to improve the machine-actionability of data for increasing reuse and useability of data and to deal with increasing attributes of data volume, velocity, and variability.

Data integration and smart energy management

25. When data is created, collected, and curated it generates a history or a provenance. Data provenance describes its origin, how that dataset came to be, what operations were performed on it, who performed them, when they were performed, and why those operations were performed.

[8] NIST Special Publication 800-79-2, "Guidelines for the Authorization of Personal Identity Verification Card Issuers (PCI) and Derived PIV Credential Issuers (DPCI)", July 2015.

[9] Gartner, "Dark Data", Information Technology Glossary. Available at: https://www.gartner.com/en/information-technology/glossary/dark-data (accessed on 27 May 2023).

[10] GO FAIR Initiative. Available at: https://www.go-fair.org/fair-principles/ (accessed 29 May 2023).

26. As data management systems and technologies are deployed heterogeneously, on a time-forward schedule and with different software architectures, data integration between energy systems is often constrained. Often it is due to problems of provenance.

27. Additionally, utility companies are still using outdated hardware and software that are often functioning on the edge of obsolescence. That leaves older technologies isolated from new platforms, as they are unable to interact with new generation of hardware and software. The utility sector, in particular, faces these issues, as the investments in digitalization are mainly direct toward new grid systems and technologies (e.g., poles, cables, transformers, PMUs), rather than upgrading older data management systems.

28. Looking deeper at the technology, integrated data platforms must support decision-making in a hybrid environment. This is the one in which there is integration with the applications that reside internal to the organization as well as potentially working with public cloud environments. The data lifecycle is complex and diverse, starting at the point of collection and continuing through to its end-use consumer. These aspects of hybridization and governance require a thoughtful, appropriately implemented, and unified data management solution.

29. Smart energy management systems combine end-use devices, distributed energy resources and advance control and communication systems with standardized data management models. Setting up data architectures to optimize systems integration and analysis is an on-going challenge. Innovative approaches towards Big Data analytics and AI as an organic process within the smart energy management systems are still needed to ensure future proofing of systems and robust data integration.

30. When considering the potential drivers of data integration, for the utility sector these use cases can range from failure probability modelling to enhancing customer experience. Whilst the various operational and customer-oriented use cases have their unique challenges, having a standardized data architecture and data management processes which are optimized for analytics can expedite breakthroughs on cutting-edge solutions.[11]

31. Several standardized data models are available as reference:

(a) In the United States, the Electric Power Research Institute (EPRI) published a utility-centric synthesized framework for large organization data management assessment. The purpose of this framework is for senior management to holistically evaluate the people, processes and technologies in their organizations that support data management and data analytics;[12]

(b) The Common Information Model (CIM), developed and maintained by DMTF (formerly known as the Distributed Management Task Force[13]), is an international standard schema that provides a common way to represent computational and networking elements in a system and their relationships to other systems and elements, a large and robust framework for data and equipment communications management. The information model, the use of which requires a cross-disciplined set of skills and an extensive experience, defines and organizes common and consistent semantics for equipment and services. This is done through object-oriented class abstractions, inheritance, and connection associations. Management of services such as fault diagnostics, system configuration, accountancy, performance, and security are provided by the CIM model. Access to the standards documents, binary libraries for object class definitions and relationship hierarchies are available through membership.

B. Utility analytics sector skills availability

32. Within the context of a distribution network, the data is generated and collected from AMI and smart meters, weather stations, SCADA systems, and transportation (for example,

[11] Brad Gall, Chad Tucker, Beth Massey, "Shared Services Common Data Model to Deliver Advanced Analytics", *Proceedings of 2022 IEEE International Smart Cities Conference* (Pafos, Cyprus), 2022, pp. 1-5.

[12] Electric Power Research Institute, "Data Governance and Utility Analytics Best Practices", 30 April 2014.

[13] DMTF, "Common Information Model", https://www.dmtf.org/standards/cim (accessed on 23 May 2023).

EV charging) systems. For each of these types of systems, there are a plethora of vendors to choose from, each with their own proprietary method to store and share data.

33. High resolution data provided by technologies such as AMI meters is determined by the 'flags' selected within the meter. One might believe it is important to collect as much data as possible. However, it is more valuable to collect the interval data with the most built-in information. For example, smart meters report voltage readings which can also give information about the health and loading of the associated transformer. Vendor depending, the number of chosen flags in the meter can increase the amount of data collected and subsequent costs to store. The cost-benefit analysis for how much and which data types contain the most value, is a worthwhile effort that needs to be developed early in the process. Such a cost-benefit analysis should also consider the lack of standardization among vendors, the skills needed to navigate the various and heterogenous systems from which the data is collected, the environments required to optimally format and manage the data needed for analytics, and allow for the additional work required for environmental setup and commissioning.

34. 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 operations personnel that know the day-to-day business and data analysts that know the algorithms. All utility segments would benefit from working together so that redundancy in algorithms and datasets is limited.

35. It is noted that, "Remarkably little is settled around the wise use of technology. For example, for products that can be controlled by their manufacturer – like smart home devices and new cards – [it is] unclear what the manufacturer's responsibilities are."[14] Expertise and supporting policies around concerns on privacy and responsibility to the consumer for device failure whether malicious or not and whether fatal or not are currently lacking in both design and implementation.

36. Challenges around the upskilling of the current workforce to effectively use tools and techniques currently available and drive improvements in education require further research into other key areas including data translation into operational needs, data monetization and cybersecurity.

Data translation into operational needs

37. 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 time series data or even Geographic Information System (GIS) data 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 as measurable objectives can create actionable insights.

38. In the utility sector, organizations working in traditional energy generation (i.e., fossil-based) have to adjust their cost margins by improving the efficiency of the plants. Some studies project that proper use of advanced analytics can bring about savings of between 5 and 7.5 per cent.[15] This can be due to an improvement in uptime, predictive maintenance based on failure rate forecasting, and optimization of fuel consumption coupled with a focus on performance monitoring to reduce (or eliminate) over-production.

39. It was estimated in 2018, that data-directed technologies can drive up operations and maintenance cost savings to more than 12 per cent. The costs of sensors and data capture devices have significantly decreased as much as one-tenth the price from ten years' ago. As more communication improvements such as 5G or the future 6G technologies continue to boost transmission speeds of 1,000 Gbps for 6G compared to the 600 Mbps of 5G, data from IoT solutions can be analyzed, and decision-making support given in real-time.

[14] Jonathan L. Zittrain, *The Future of the Internet -- And How to Stop It* (Yale University Press & Penguin UK, 2008).

[15] McKinsey & Company, "The Digital Utility: New challenges, capabilities and opportunities", June 2018.

40. The dynamic nature of this level of data analysis brings a step change to the definition of strategic use cases. Use cases such as failure probability monitoring, outage detection and prediction, smart grid security and theft detection, transactive energy management, preventive maintenance for equipment, optimizing asset performance, demand response management, real-time customer rates and billing, and enhanced customer experience seem to rise to top of the list among utilities that have advanced their high-resolution data capture programmes.

41. The challenge remains on whether a utility can take advantage of the data collected to formulate a strategy around known (or unknown) needs.

Data monetization

42. To a large extent, data monetization is the ideal goal in achieving widespread deployment of data analytics. Utilities and energy service unknowingly possess a wealth of extensive and very valuable customer and operational datasets of information with more data becoming increasingly available through the deployment of smart devices.

43. 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. Using methods such as 360-degree customer profiles can address the increase in churn rates, which is as much as 25% in some markets.[16] Solutions such as automated voice analytics in call centers, integrated with communications systems (e.g., mobile applications in the field) and corporate websites, along with consumption analysis and dynamic rate pricing, will allow companies to meet their customers where they are. This holistic level of integration increases lifetime value for customers and reduces churn for the utility.

44. Especially for those utilities that are distribution-only providers, increasing the efficiency of their retail portfolio through better assessment of customer creditworthiness and consumption variation can help to minimize defaults and avoid fraud. Studies show that the impact of an integrated strategy for customer analytics increases the profit margins of companies by 5-10 per cent in addition to increasing customer satisfaction.

45. Additionally, there is a significant quantity of data gathered for regular business activities which is not used for any other reason; this dark data could be important to other businesses and could be a new revenue source. Using insights based on experiences, utilities can provide new revenue generating products and services and enhance product and operational performance to create a more compelling and sustained customer relationship.

Cybersecurity and grid resiliency

46. 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 high risks of cyberwarfare, utility companies must set proper prevention and mitigation strategies, while also developing business continuity plans after cybersecurity breaches.

47. It was argued that “by the end of 2023, modern privacy laws will cover the personal information of 75 [per cent] of the world’s population.”[17] As modern-day customers become more knowledgeable about their role in the digital grid journey, they will want to know what kind of data is being collected and the intended (and actual) uses. Utilities and energy providers, especially those that cover multiple geopolitical boundaries in various jurisdictions, will need a strategic customer education programme focused on cybersecurity and how the applications are fit for the local jurisdiction.

48. The same analysis concludes that by 2024, “organizations that adopt a cybersecurity mesh architecture will reduce the financial impact of security incidents by an average of 90 [per cent].” This type of architecture can be extended to cover identities outside the traditional

[16] Pablo Boixeda, “Optimizing the Energy Sector with Data Analytics”, Cloudera, 20 December 2022.

[17] Gartner, “The Top 8 Cybersecurity Predictions for 2021-2022”, 20 October 2021.

security perimeter. The increase of remote working across organizations will drive adoption of these architectures in the coming years.

49. It is also argued in the report that by 2025, ever-increasing percentages of organizations will adopt strategies around: cybersecurity risk policies for third-party transactions (60 per cent), legislation for regulating ransomware negotiations (30 per cent; up from 1 per cent in 2021), a dedicated cybersecurity committee representative on the board of directors (40 per cent), and create a culture of organizational resilience to survive coincident threats from cybercrime, severe weather, civil unrest and political instability (70 per cent).

50. Further analysis and additional information on impacts of cybersecurity to the digital landscape of the energy sector, is contained in the document "Key considerations and solutions to ensure cyber resiliency in the smart integrated energy systems" (ECE/ENERGY/GE.6/2023/3−ECE/ENERGY/GE.5/2023/3).

C. Big Data analytics modelling research and development efforts

51. Access to on-going research and application efforts for utility-scale data analytics in the utility industry is a key pillar to progressing Big Data analytics and indeed digitalization of the energy sector as a whole. At the same time, access to and use of the results from the research efforts of responsible national bodies, is often difficult.

52. Data maturity is an important concept for energy providers. As utilities find that data is evermore central to all enterprise strategies, helping to drive innovation, and continues to be integrated across departments, it is a necessary priority to develop and implement a strategy for a data maturity roadmap. In this way utilities can drive optimization in their internal processes as well as innovate over and provide reliable services for their customers.

53. The below Table shows a maturity curve of data capabilities. Progressing along the levels of this curve, an organization increases its data management, algorithm development and delivery capabilities. Studies show that companies with mature data management and robust (i.e., repeatable, verifiable) analytics processes can boost their profitability by an average of 12.5 per cent of total gross profit.[18] Advancing up the levels of the curve, demonstrates a maturity in an organizations' data analytics strategy and implementation. As more investments are made into the analytics capabilities, the faster an organization can progress along this curve. Studies show that even conservative investments yield good movement and most utilities and energy provider organizations transition between phases fairly quickly even with modest investments in their data analytics.[19]

Table
Data and analytics maturity

<table>
<tr><th>Level</th><th>Data and analytics maturity</th><th>Analytics capability</th><th>Commentary</th></tr>
<tr><td>1</td><td>Much data and too many data warehouses</td><td rowspan="2">Reactive reporting – concerned with current issues
• Transactional lists and printouts
• Historical and cost monitoring focus
• No integration of data or operational applications
• Data is scattered in heterogenous storage platforms</td><td rowspan="2">An organization has made some investment in analytics infrastructure; however, it may be outdated or there was no data management strategy to start from. The result is that data is managed in an ad hoc way and decision-making is often reactive based on today's priority rather than based on empirical value-driven evidence.</td></tr>
<tr><td>2</td><td>Basic reporting and minimal automation</td></tr>
</table>

[18] Richard Carufel, "What's your data really worth? It depends on your data maturity level", Agility PR Solutions, 26 March 2020.
[19] Electricity Advisory Committee, "Big Data Analytics: Recommendations for the U.S. Department of Energy", February 2021.

Level	*Data and analytics maturity*	*Analytics capability*	*Commentary*
3	Business intelligence with statistical analysis	Planned analysis – short-term planning • Diagnostic reporting • Data storage and access are automated • Cross-operational integration • Uneven analytics competencies	A growing maturing within departments and across the organization, as operational planning turns more strategic and analytics competencies increase.
4	Predictive / prescriptive modelling	Strategic analysis – consistent delivery • Departmental scorecards, dashboards • Maturing analytic capability • Consistent and effortless production • Insight to action	
5	Model and process Optimization	Process optimization – foresight • Actions based on future planning • Full integration and use of external data • Real-time analytics as a differentiator • Organizational scorecards, dashboards • Widespread analytic capability	Processes across the organization are optimized and integrated. Decisions are based on future planning and deep insights from empirical and robust forecasting. Real-time analytics using advanced capabilities are consistently and organically deployed, and organizational dashboards have been standardized, for example with agreed data-oriented nomenclatures.

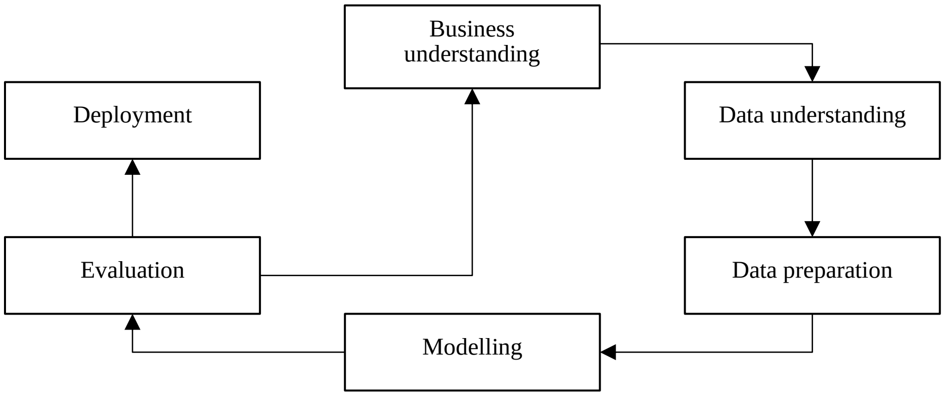
54. Moving along the curve means that organizations need to change their paradigm for decision-making. This requires integrated thinking, cross-functional applications, and cross-operational team collaboration.

Big Data, advanced analytics model research, development, and deployment efforts, and outreach

55. 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. At the same time, the number of RES and other distributed energy resources continue to increase penetration into the grid globally, increasing 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.

56. Utilities must ensure that their analytics models follow the rules of validity, reproducibility, and transparency. Industry standards can offer an organized method for planning and implementing data mining initiatives, such as the Cross-industry standard process for data mining (CRISP-DM) framework, as described in the Figure below. It consists of phases for business comprehension, data comprehension, data preparation, modelling, evaluation, and deployment. Utilities should also think about combining machine learning and AI techniques to handle complicated, high-volume data on cloud-based platforms for scalability.

Figure
CRISP-DM framework



Source: adapted from https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

57. It is essential in today’s world of growing demand for highly skilled technical engineers, that businesses prioritize training and development of their current teams. Continuous training and skill development, focusing on a culture of data-driven decision-making, hiring for a blend of technical (data science, machine learning, statistics) and sector-specific (energy systems, regulation) expertise, and fostering collaboration and knowledge-sharing among team members are key practices to building an effective in-house analytics team.

58. Models that have the potential to benefit others, such as those that anticipate energy consumption, forecast load and the output of RES, or those that analyze grid stability, should be shared either open-source or for revenue reasons. In addition, there are numerous techniques, including regression, time-series analysis, machine learning (including neural networks), and simulation models, that could serve as the foundation for these models. Most importantly, though, is publicly accessible datasets and open-source software, which can promote standardization and transparency among various parties. Shared models, data science techniques, data and open-source software can significantly encourage greater use and collaboration.

59. As research, development, and deployment in Big Data analytics in the energy 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 and the establishment of regional innovation hubs that focus on specific technologies.

60. Worldwide efforts are being made to create communities that share best practices for using advanced analytics models. For instance, the “Digital Europe Programme” of the European Commission highlights the necessity of creating high-impact projects using AI and data analytics. Similar to this, the United States Department of Energy launched the Grid Modernization Initiative to work on revolutionary reforms utilizing data analytics. In order to share information and promote innovation, governments are increasingly forming cross-border collaborations regionally.

61. The gathering of thorough, varied, and high-quality granular data from many sources, such as IoT devices, smart meters, weather stations, etc., is crucial to improving the accuracy of data analytics models. In addition, the incorporation of real-time data streams, the use of feature engineering strategies, and the application of sophisticated machine learning algorithms all have the potential to enhance model performance.

62. The selection of demand-side analytics models frequently entails a trade-off between granularity and privacy, speed and accuracy, and complexity and interpretability. The accuracy, interpretability, and responsiveness of the results are all directly impacted by these compromises. For instance, a very complicated model could produce findings that are more accurate but may be more difficult to understand and take longer to run than a simpler model.

63. One key consideration is the existence of hidden biases which can lead to elements like skewed data, false assumptions, or biased algorithm design. These biases might result in unfair treatment, biased behavior, or misleading insights. This highlights the crucial need for effective data governance procedures, which also include rigorous data gathering, cleaning, auditing, and validation.

64. While some types of biases can be reduced through automation, it is important to understand that algorithms themselves can contain and reinforce biases, especially if they are developed using biased data or deployed without enough control. Therefore, it is essential to have open and accountable methods which are included to the automation of any decision-making processes. Algorithmic biases can be reduced by the use of explicable AI approaches and adherence to AI ethics principles.

Example of advanced analytics and AI application for heat meter failure identification

65. Whilst the synergy between Big Data and AI have not utilized its full potential, there are some pilot programmes that showcase best practices for advanced analytics and provide a pathway that others could follow.

66. In the city of Vilnius, Lithuania the heat provider wanted to establish a programme that analyzed customer heat usage over time to identify anomalous readings which may indicate a malfunctioning heat meter. Statistically, 0.5 per cent of heat meters are identified as ‘broken’ and may send false interval readings. For a service territory of 100,000 customer, this amounts to 500 meters sending false readings. This can create erroneous calculations of actual demand which results in inaccurate billing.

67. Using smart heat meters, interval data was collected over two heating seasons. Considering customer behaviour (based on past historical demand) and seasonality, a Big Data statistical analysis was performed to clean unreliable data and remove external factors such as outliers for off-season heating days and unseasonable temperatures, and to identify the parameters for a business-as-usual scenario. This training dataset was used to train an AI system and then used for utility billing purposes and to identify any heat meter failures.

68. The AI system was tested on monthly heat demand data, the data was cleaned and updated to reflect the number of heating days and the outside temperature for the current month. An Interactive Actual Energy Consumption Map was created using the heat energy customer profiles, and then normalized for influencing factors such as season duration and temperature and apartment size. In this way, the AI system could compare all data for any customer of any year.

69. The data were presented on a GIS map using a color range from green to red, and actual energy performance class from 1 to 10. This map was made publicly available to customers at the utility website, and allowed comparison of similar buildings, thus bringing sharp attention to buildings’ operation, maintenance, and facility management activities, in order to find and fix potential problems. This also supports the measurement and verification process for energy conservation measures such as insulation of buildings and their refurbishment, as customers can compare the actual energy consumption after buildings’ improvements.

70. Although this approach is different from the mandatory Energy Performance Class A-E, which shows only theoretical consumption data, it shows that without Big Data and AI analytics utilities will struggle to properly identify energy consumption issues to provide actual energy consumption information to their customers and make longer-term decisions for capital and operational expenditures on grid assets.

IV. Conclusions and policy recommendations

71. Based on the discussions above, the following conclusions and policy recommendations are presented for consideration:

(a) On data sharing and democratization of data:

(i) Data curation:

- Need for clearly defined, easy to understand national and international standards for handling sensitive data
- Need for cybersecurity and cyber-physical security standards and turnkey operations for residential data owners.

(ii) Data availability: there is still a significant lack of access to cross-discipline and cross-sectoral data. More research into applications of FAIR principles for data is needed, notably ‘dark data’;

(iii) Data integration and smart energy management:

- Need for systems integration to allow for expedited data integration from heterogenous systems
- Need for unified data management standards
- Need for standards and/or protocols for smart energy management architectures
- Need for building strategic use cases
- Need for easy-to-understand and quick-to-use data architectures that can be analytics- and utility-oriented.

(b) On utility analytics sector skills availability:

(i) Data translation into operational needs:

- High-resolution data capture through the proliferation of sensors and smart devices, drive the need for advanced analytics
- New skills are needed for the existing workforce and incoming new hires, to take advantage of greater computing power and robust data architectures for standardization of data sets across departments
- 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
- Expertise and supporting policies around concerns on privacy and responsibility to the consumer are needed for both design and implementation
- Challenges around the upskilling of the current workforce to more effectively use tools and techniques that are currently available and to drive improvements in education, require further research.

(ii) Data monetization:

- Dynamic rate pricing for increased customer value
- Potential model of providing value to consumers in exchange for data.

(iii) Cybersecurity and grid resiliency: utilities will need a strategic customer education programme focused on cybersecurity and how the applications are fit for the local jurisdiction.

(c) On Big Data analytics and maturity:

(i) Big Data, advanced analytics model research, development, and deployment efforts, and outreach:

- The establishment of local and regional innovation hubs is needed to fully test and secure digital and data technologies
- Education of energy communities’ stakeholders is necessary to help consumers understand and acquire agency for their role in the digital transition

- There needs to be greater actionable incentives for the energy sector stakeholders, specifically utilities and energy provider segments to care about leveraging Big Data
- The use cases that justify large investments in data collection, management, analytics, and AI infrastructure need to be demonstrated to ensure clear returns on investments
- The cost recovery model must be considered so that justifiable investments in data analytics can be considered as capital investments with clearly defined customer benefits.

72. The Task Force on Digitalization in Energy further suggests the follow-on activities:

(a) Investigate the above conclusions and carry out comprehensive work and deeper analysis of each, preferably in collaboration with the subsidiary bodies of the Committee on Sustainable Energy, which might accordingly extend the mandate of the Task Force on Digitalization in Energy;

(b) Conduct 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 machine learning, progression of AI capabilities), grid resiliency, infrastructure investment (particularly as it relates to data access, storage, management, and real-time analytics), in accordance with the 2024-2025 mandate of the Task Force on Digitalization in Energy.

Annex

Table
Evolution of data analytics

Year	*Milestone event*	*Description*
1865	Business Intelligence Term	Richard Miller Devens uses the word 'business intelligence" in his Encyclopaedia of commercial and business stories. This is believed to be the first study of a company using data analysis for commercial goals.
1928	Magnetic Storage	Fritz Pfleumer, a German-Austrian engineer, invents a way of magnetically storing information on tape. His methods are still used today, with the vast bulk of digital data being stored magnetically on computer hard drives.
1928	Data Processing Computers	IBM introduced the 305 and 650 RAMAC (Random Access Memory Accounting) 'data processing computers' in 1956, which included the first-ever disk storage device.
1965	First Government Data Centre	The US government Intends to build the world's first data center, which will hold 742 million tax records and 175 million fingerprint sets on magnetic tape.
1989	"Big Data" Term	The term 'Big Data' was first used in a magazine article by fiction novelist Erik Larson, who was remarking on advertisers' exploitation of data to target customers.
1996	Digital Storage Cost-Efficient	According to RJT Morris and BJ Truskowski's 2003 book The Evolution or Storage Systems, digital storage is becoming less expensive than paper storage.
1998	Next Wave of Infostress	SGI Chief Scientist, John R. Mashey, gives a paper titled Big Data ... and the Next Wave of Infostress at a USENIX convention.
2001	"3 V's" Term	Doug Laney defines the three -Vs' of Big Data: volume, velocity, and variety.
2005	Web 2.0	This year marks the debut of Hadoop, an open-source Big Data platform presently developed by Apache. The user-generated web, known as Web 2.0, is developed the same year.
2008	14.7 Exabyte of Fresh Data	The world's servers handle 9.57 zenabytes (9.57 trillion gigabytes) of data every day, which is comparable to 12 gigabytes of data per person per day. This year, an estimated 14.7 exabyte of data is created.
2009	CIO (Data Oriented Title)	According to Gartner, the top priority for CIOs Is business intelligence. As businesses suffer economic instability and uncertainty as a result of the Great Recession, extracting value from data becomes critical.
2011	Analytical Skills Scarcity	There is a shortage of between 140,000 and 1190,000 professionals with profound analytical abilities, as well as 1.5 million analysts and managers who can make appropriate data-driven judgments.
2012	Big Data Research and Development	The Obama administration announces the Big Data Research and Development Initiative, which improves the ability to extract insights from data, accelerate the pace of science, technology, engineering, and mathematics (STEM).
2014	Next Wave of Infostress	In the United States, mobile devices outnumber desktop personal computers for the first time. Two years later, in 2016, the rest of the globe follows suit.
2020	Edge Computing	The next frontier for Big Data is edge computing, which refers to computation done near the source of data gathering, rather than in the cloud or a centralized data center.