BIG DATA CONCEPT IMPLEMENTATION FOR ENERGY CONSUMPTION

IMPLEMENTAREA CONCEPTULUI DE META-DATE ("BIG DATA") PENTRU UTILIZAREA ENERGIEI

Oana CONSTANTINESCU (VASILIU)[1] , Cristian GHEORGHIU[2]

Abstract: *Production, distribution and consumption of energy optimisation is one of the major concerns of the communities, being considered as one of the main benefits of data processing, in order to allow better targeting of actions and better management of available information. Communities are mainly focused on managing energy demand, measuring energy efficiency and adopting new methods of supply and use. At the heart of this activity is data, considered to be of equal importance regardless of the source that provides it.*

Keywords: Big data, data management, digitization

Rezumat: *Optimizarea producției, distribuției și consumului de energie constituie una dintre preocupările majore ale colectivităților, fiind considerată ca unul dintre principalele beneficii ale prelucrării datelor, cu scopul de a permite o mai bună orientare a acțiunilor și o mai bună gestionare a informațiilor disponibile. Colectivitățile se orientează mai ales spre gestionarea cererii de energie, măsura eficienței energetice și adoptarea noilor forme de producere și utilizare. În centrul acestei activități se află datele, considerate ca având aceeași importanță indiferent de sursa care le furnizează.*

Cuvinte cheie: Meta-date, gestiunea datelor, digitalizare

1. Intrduction

Meta-data, also known as "big data", are of high interest for actors involved in various fields of activity. The rapid increase in the amount of data used raises the issue of their storage, given that traditional storage techniques such as relational databases do not allow the storage of such large amounts of data. Their exploitation is also a major problem: the analysis of these metadata in a reasonable amount of time is mainly accompanied by the development of analytical applications that process the data to extract useful information.

[1] PhD. Std., University Politehnica of Bucharest, e-mail: oanamconst@gmail.com
[2] PhD. Std., Teaching Assistant, University Politehnica of Bucharest, e-mail: gheorghiu.cristian.k@gmail.com

These analyzes, often called Big Analytics or Analytique, are based on distributed calculation methods. They apply different algorithms that fall into the following categories: Data Mining, Machine Learning, Decision Making or Visualization.

Analytic involves the implementation of statistical and mathematical methodologies in the work processes of a company, regardless of whether they are operational or supportive, even in its strategy or business model. To the extent that the raw material of the analysis is digital data, this implementation is adapted to the digital transformation of society, the economy and therefore the enterprises.

In recent years, a number of new concepts have emerged such as: Business Intelligence, data analysis, business analysis, big data (meta-data), intelligent data, which have generated the emergence of a large number of theories, analysis and related efficiency and opportunities for the development of new specific products and services.

Recent advances in digital technologies and services have led to increasingly complex data collection and processing systems that could revolutionize business, scientific and social processes, enabling decision-making based on a comprehensive catalog of rapidly available data. The "meta-data" phenomenon (Big Data) is booming; it has enormous potential in areas ranging from health, food security, climate and energy efficiency, smart transportation systems and smart cities.

Meta-data already affects many aspects of contemporary life and has the potential to change the way we manage the economy and the very structures of society. With the publication of the Digital Single Market Strategy for Europe in May 2015, the European Commission has recognized that value generation at different stages of the data value chain will be at the heart of future knowledge-based economies.

2. Challenges and issues of the digital transformation of the energy sector

2.1. Issues related to the implementation of the digital technology

In the energy sector, digital technology makes it possible to develop new perspectives around six axes:

- the emergence of a new networked and decentralized energy model;
- the possibility of a new energy mix in which renewable energies occupy a leading place;
- the “Business to Business” (B2B) model of the sector becomes “Business to Business to Client” (B2B2C) by integrating the expectations of the final customer into the value chain;
- development of new digital services;
- the digitization of networks (smart grids) offers new ways for optimizing the management, operation, maintenance and development of networks;

- the development of predictive capacities will be an activity of the future (for example: anticipating local network constraints of consumption and production forecasts at very short time scales).

This will lead to a profound transformation of the energy sector. The implementation of new sensors in the transmission and distribution networks will lead to a considerable increase in the volume of data to be managed.

The use of this new data is an important source of value creation. For network administrators, this means, for example, managing the combination of measurement data (index, subscriber load curves, etc.) and network data (power, voltage, etc.) with external data (especially weather).). This value creation concerns all stakeholders in the sector, from producers to the end customer. For the latter, it is especially about meeting the new expectations: better control of consumption, more comfort, greater availability of customer service, personalization of services, fast and remote intervention, etc..

The obstacles that slows down the digital transformation are of human nature, cultural and organizational:

- *data collection is still limited to traditional channels* - data collected by companies for the purpose of customer knowledge and marketing strategy comes overwhelmingly from traditional billing systems and customer relationship management tools.

- *unstructured data is poorly collected or exploited* - in addition to traditional customer information (contact details, behavior and expenses), companies sometimes collect less traditional data (textual data from the Internet, opinions transmitted or shared online, graphic data, sound data, images or videos). The analysis, especially in a context of "meta-data", involves the processing of big data from various sources. Their treatment must be carried out in a limited time, sometimes even in real time. The problem is not only storing a considerable amount of customer data but also selecting from the continuous data stream those data that will be kept based on specific skills and tools that are not yet available.

- *lack of availability of analytical skills* - utility providers do not traditionally have internal analytical profiles and skills to exploit all data useful to their business: processing historical data in real time, contextualizing and obtaining useful and exploitable results from them.

- *lack of data processing and exploitation tools* - data from the web, mobile phones and other connected equipment have made traditional analysis techniques obsolete. They currently lack the power, speed, flexibility and have become too expensive. The combination of statistical, mathematical and computer methodologies make it possible to partially solve these processing difficulties, so that calculations are generally performed faster. However, specific technologies and processing tools are needed to extract information from unstructured data, which is the weakest link in the analysis. It is necessary for this data to be "translated", to become more reliable, to be indexed and combined with existing data, in order to be integrated into decisions. The analysis requires an architecture and an information system that responds to the challenges of volume, computing and processing speed, diversity of formats and that ultimately allows a fluid flow of information within the company.

- *data analysis is too little oriented towards predictive or real-time objectives* - a small part of companies use customer data for predictive purposes and some of them do so to optimize technical processes by increasing execution speed and storage capacity (elements key to exploiting the growing volumes and the ever-faster flow of data and information). In industry and especially in the network and infrastructure industry (energy, communications, transport), the collection and analysis of real-time data from several sources for predictive purposes is essential. This applies in particular to data from sensors connected to the network, data obtained through monitoring, environmental data, financial data, which make it possible to significantly improve the quality of services and the operational performance of companies. Based on these data, preventive diagnoses and predictive maintenance can be performed, real-time identification of patterns leading to preventive detection of incidents and failures to determine the areas most at risk and to identify the main cause of problems.

- *the absence of measures to recover investment in data mining* - meta-data marks a crucial moment in the exploitation of customer data and is a particularly useful lever for developing and increasing profitability. However, awareness within companies is very low. Quantifying and measuring the value of data analysis projects, based on predefined key performance indicators and an estimate of investment return, would make it possible to establish a roadmap with data analysis projects to be carried out as a priority.

- *trust of actors, a major risk to data reliability* - with the advent of the digital economy and the development of the Internet of Things (IoT) concept, the amount of data increases exponentially and forms a gigantic mass of structured and unstructured data. The source of this data is often a person: the customer, the citizen, the user or the employee. For this reason, the issue of data security is the key to the future of the implementation of relevant analytical solutions. Data collection and analysis will need to be accompanied by user confidence to gain useful value. The most frequently asked questions concern the ownership of the data, the use of the data by the data controller, the recipient of the data, the security measures implemented to protect them against unauthorized access, the location of such data, etc.

2.2. Observability and state estimation in the context of the smart grid for electricity distribution

In distribution networks, among the functions of the smart grid that play a key role, the functions of network observation and management can be mentioned. These functions cover the whole chain: data collection and measurements from the network, consumption, information analysis, decision support for operators and sending action orders on network components. The objective to be achieved is to ensure the optimal operation of the entire electrical system from an economic point of view, the quality and safety of the power supply, reliability, environmental impact, etc.

These functions, some of which may be automated, use information and communication technology (IC&T) and are able to assist the operator in solving both network operation problems through remote control actions and planning operations.

In practice, remote control (remote control) comprises, on the one hand, remote signals and telemetry (remote monitoring) which transmit data from devices to the operator and, on the other hand, remote controls and remote settings, which transmit operator data to devices.

Remote monitoring makes it possible to know the status of the network in real time. It gathers the measurements of the network characteristics (voltage, frequency, current, temperature, etc.), the position signals of the different devices (switches, fault crossing indicators, remotely controlled operating devices), but also their possible fault tripping. The control centers are equipped with synoptic images (generally graphical presentation, which allows the visualization of a set of related information or a complex system) created according to the needs of the operator. As a result, the network operator can view the operating diagrams, the values of the measured parameters, the detailed content of the alarms, the installation states, etc. The remote control offers the possibility to open and close the network devices or the remote electrical system, as well as to perform some adjustment actions or to activate the automations.

SCADA systems in distribution networks currently use very few sensors which does not allow the implementation of advanced automatic functions. In the current and future context, in order to take advantage of the flexibility of the network and local resources, the operator will need to know precisely the state of its system. For this it will need a special function called "status estimator" which, using appropriate algorithms, can provide accurate and reliable data, with minimal error, to advanced network functions. This function, based on measurements and pseudo-measurements (models, knowledge and data on power withdrawn from the network), should allow the estimation, in "real time", of the state of the distribution network. Such status estimation functions can be widely used in the case of transmission networks (at regional and national level, even continental), but they cannot be transposed directly into distribution networks (at local level) because these networks have strong differences. structural and operational.

2.3. Data processing and meta-data in the energy field

After the data collection stage using different sensors, the following processing is required:

- *Control* - to ensure the validity of the raw data collected.

- *Correction* - for certain types of data estimation, correction mechanisms can be implemented. These mechanisms are coordinated with all parties involved. For example, in the event of a failure of a particular customer's measuring device, the energy consumed can be estimated using methods recognized by the competent bodies. Climate corrections may also be made using methods accepted by all stakeholders and approved by the regulatory authority.

- *Storage* - is an important step that involves securing the storage of validated data and correcting it with the help of the IT systems. These systems must be secured in accordance with the legal and regulatory framework. Thus, the data processing chain of smart meters will comply with the security criteria developed by the competent authorities and the data will be able to be evaluated and certified.

- *Adapted processing* - depending on the objectives, an adapted data processing can be performed. For example, it is possible for the process consumption data to deduct billing data for suppliers who have supply contracts, to aggregate consumption and production data on the perimeter of a concession, and so on.

- *Availability* - data is used for the operation, maintenance and development of the public distribution network. External actors may also receive the data necessary to carry out their activities or missions (final customers, suppliers, communities, third parties, regulators, etc.), in accordance with the legal framework, in particular as regards to the protection of sensitive data, commercial and personal type.

3. Meta-data analysis tools for the transition to smart grids

The smart grids emergence is a consequence of European goals to encourage more rational energy use, optimized grid management and a greater connection of renewable energy sources at all levels of the grid voltage.

In many countries, regulators have performed cost-benefit analysis (CBA) of the implementation of advanced measurement technologies to validate the same implementations. The European Commission's recent report, entitled "Cost-benefit analyzes and the state of the implementation of smart metering in the EU-27", presents studies on the costs and benefits identified by different European countries.

All over the world and especially in Europe, distribution network operators are implementing or have implemented advanced measurement systems, as well as new measurement and control technologies on distribution networks. However, a significant part of these network operators use a very small part of the capacity of these new metering systems, either for remote operations or for monthly readings. Although they save money on manual meter reading, they do not realize the enormous potential of these technologies. Another part of the operators made the effort to implement technologies to collect data daily. However, very often, these data end up in data warehouses for which no effective analytical tools would allow their exploitation, are sometimes very expensive (worth millions of euros) and remain rarely used.

In order to improve the management of the very complex energy system, all operators need to understand and manage these large amounts of data in order to be able to implement concrete actions. The current problem is both the ability to store this data and the ability to extract important data for the management of the electricity system as a whole.

In practice, extracting value from the new data provided by smart grids and advanced measurement systems is complex. This requires advanced metadata analysis capabilities, in addition to generating more or less dynamic reports or dashboards. Statistical analysis of large amounts and variety of data has the ability to produce results that operational teams can use (monitoring department, networking department, customer department, asset management department, etc.).

First of all, this is a technological challenge. The specific metadata technologies currently available on the market are extremely varied and require the existence of a number of components: high-performance infrastructures, data management technologies, real-time or non-real-time integration tools, specific analytical and statistical tools.

Currently, there are appropriate solutions for each of these technologies. In order to implement this type of architecture, it is possible to take responsibility for the reliable integration of all the components involved, but the number of people competent to apply the technologies is still very limited.

Secondly, this activity is a functional challenge for all business teams: never having available this type of information or these volumes, it is natural that the analyzes to be performed and how they are conducted, from a business perspective, do not they are not easy to define either. Metadata analysis tools are really confronted with the paradigm "we do not know in advance what we will find and what analyzes we will be able to find".

This dual “technical-functional” challenge significantly complicates the implementation of an IT project and in practice can make it difficult to quickly provide tools to effectively support business teams in their investigations.

In order for data management to be useful to these actors, it will be necessary to provide:

- specific and adapted tools for data exploration and visualization;
- tools for ad-hoc analysis of this data, allowing the design of appropriate analysis algorithms;
- tools to perform these algorithmic analyzes on a daily basis and on a large scale in an efficient manner;
- support from scientists who know both smart grids and analytical tools to continuously support the development and adaptation of analyzes.

The implementation of big data analytics capabilities in the energy sector thus seems to be the cornerstone that allows, beyond the primary benefits of remote meter reading, to meet the promises of energy transition in terms of energy savings, energy efficiency, capacity integrate increasingly new uses (such as electric vehicles) for both companies and users.

This activity must comply with precise rules to protect the roles, responsibilities and interests of the various actors in the system, in accordance with the regulations and legislation in force:

- the authorities could play an important role in managing access to (aggregate) data related to their concession, especially to communities;
- providers can play a role in accessing individual data because the data generated by the communication meter belongs only to the customer to the extent and for the period in which the latter has a supply contract associated with this meter.

New services will often be based on data from different sources: data from different energy networks (electricity, gas, heat), data from smart meters, but also data from equipment located downstream of meters, etc.

4. The benefits of data management provided by smart grids for energy suppliers

The new services offered by smart grids open up perspectives for improving the relations between the suppliers and the customers. As new players try to enter this market, competition increases and energy providers need to seize the opportunities offered by smart grids and implement development strategies to remain competitive.

1. Customer relations are a key challenge for energy suppliers

In recent years, suppliers have understood the importance of establishing a stronger and better relationship with the customers. Extensive transformation programs have been launched through projects based on better customer segmentation, in order to better adapt the consideration of customer needs and the response provided, both in terms of support channel, customer discourse on the level support etc.

In this context, the emergence of smart grids is a great opportunity for suppliers: a means to better understand the uses and behavior of customers, to better support them, to communicate better with them and thus strengthen the quality of the relationship.

2. Smart grids offer new perspectives for energy suppliers

Smart grids facilitate the management of energy demand, in particular by providing the customer (individual, company or institution) with information on these uses and consumption, which will allow them to optimize their energy costs. The challenge is to make the consumer active and to empower him by appropriating his consumption data.

However, suppliers will not be the only ones to benefit from the new data resulting from the implementation of new information and communication technologies in energy networks.

Smart grids allow the relationship between customers and suppliers to be restored. The customer will be able to monitor, manage and control their consumption, but also the sale of the energy they eventually produce. With detailed information about his consumption, the customer can decide to adapt his own consumption behavior. As a result, the level of customer demand for their supplier and its customer relationship could be considerably increased, endangering suppliers who have failed to develop their value propositions and customer relationships.

In an intensified competitive environment, new types of competitors will rely on the development of smart grids and energy services to try to capture some of the value of this market.

3. Smart grids place customer relationships at the heart of supplier issues

Thanks to the data transmitted by smart grids or smart meters, suppliers will be able to know the consumption of their customers in a more detailed way and almost in real time. They will soon be able to identify connected devices, number of users, types of use, etc. and they will be able to answer their customers' questions in an extremely precise way, being able to immediately identify specific problems to associate them with a personalized answer.

Therefore, suppliers will need to optimize the use of this customer data with 2 key issues:

- *to better understand customers*, analyzing very precisely their uses in terms of profiles, their equipment, etc. Subsequently, it will be possible to enrich customer knowledge and improve customer segmentation based on different categories of uses, equipment, etc. Thus, based on these profiles, it will be possible to adapt the portfolio of offers and services to better meet customer expectations.

- *better demand management*. Beyond the amount of data available, its quality will be greatly improved. With real-time data, the provider will be able to better manage the quality of its services and its costs. Consumption habits could also be influenced in order to reduce consumption and improve network management.

From this perspective, the key issue will be compliance with data protection rules and respect for consumer privacy.

Datorită datelor transmise de rețele inteligente sau contoare inteligente, furnizorii vor putea cunoaște consumul clienților lor într-un mod mai detaliat și aproape în timp real. În curând vor putea identifica dispozitivele conectate, numărul de utilizatori, tipurile de utilizare etc. și vor putea răspunde la întrebările clienților lor într-un mod extrem de precis, putând identifica imediat probleme specifice pentru a le asocia cu un răspuns personalizat.

5. Conclusions

Smart grids are a major challenge for energy providers. Although it is an effective way to help them manage customer relationships and thus improve their image, smart grids also create new competitive challenges. Indeed, new players are entering the market and competing offers are growing. Therefore, suppliers need to invest quickly to provide the best services in order to capture customers and retain them in the long run.

The massive flow of data generated by smart grids allows network managers to improve network management and adapt to new constraints caused by the integration of renewable energy and the development of new uses. The use of data from smart grids contributes to a public electricity distribution service at the best cost to the community. In particular, equipment management is optimized and network operations are increasingly responsive. Data analysis also helps to optimize investment policies by:

- *equipment management* - With the digitization and modernization of networks, the distributor can collect the maximum amount of network data and external data in order to gain a better knowledge of the assets in operation. It can then optimize its maintenance by gradually moving from systematic preventive maintenance to predictive maintenance based on mass data modeling. Energy companies are trying to develop a forecasting model by creating a database with historical data on forecasting equipment and algorithms. This development will allow them to better plan maintenance interventions and reduce the occurrence of failures. Finally, this model will evolve towards more precise surveillance by integrating real-time and external data into this forecast model. With the data collected, energy companies can gradually gain a better picture of all assets on a very fine scale, but also of their operation. In the field, technicians and subcontractors will be able to benefit from such access to the geographical and historical data of the equipment, which will facilitate their interventions.

- *Improving the operation of the electricity distribution network* - By collecting data from the entire network, from substations to measuring devices, it is easier for energy suppliers to detect or even anticipate congestion situations in the public distribution network. In addition, data from smart meters will improve forecasts. The continuous improvement of these analyzes helps to better integrate energies from renewable sources, which are variable by nature. Thanks to dynamic feedback and information monitoring, the energy provider may be able to manage incidents and associated interventions more quickly. For example, the manager will be able to quickly detect incidents, diagnose them, identify a solution and plan any intervention at a specific location. Likewise, data analysis makes it possible to better identify counting failures.

- *a better approach to investment* - Based on the history of data collected on network operations, the energy provider can develop risk and cost analysis models to make investment decisions. With all the data collected, it will be able to anticipate the resources needed for the security and operation of the public distribution network. Subsequently, these models will be able to consider in more detail the new behaviors and new uses (charging electric vehicles, storage, etc.). These tools and associated data help inform investment decisions to ensure the network works at the best cost for the community.

Another problem that arises is related to the use of personal data of customers. Thus, electricity consumption data relating to a subscriber, customer or tenant are considered personal data. The actors who manage this personal data must comply with various legal obligations:

- informing customers, even obtaining express consent and establishing procedures for exercising the right of access and deletion of data;

- collection and processing of data relevant to the achievement of explicit purposes, with a retention period (conservation);

- implementation of security measures to preserve the integrity and confidentiality of this data.

The main risk identified comes from the new functionality offered by smart meters. The smaller the "measurement step", the more numerous and finer the measurements during a day, measurements that provide accurate information about the lifestyle of the people concerned. A load curve with a step of 10 minutes makes it possible in particular to identify the daily schedule (waking and sleeping hours), hours or periods of absence, etc.

In general, in order to maintain the confidentiality of the data, it is recommended to apply some basic principles:

- the load curve cannot be collected systematically;

- data collection takes place only for the following purposes:

o when justified by the need to maintain the network;

o when the user expressly requests to benefit from specific services (tariffs adapted to his consumption, energy balances, proposal for insulation works, etc.).

The recommendation also expresses data security requirements. In particular, it provides for confidentiality impact studies to be carried out prior to the installation of meters and risk analyzes to be carried out in order to determine the appropriate technical measures to be implemented. These recommendations may be amended to take account of technical and technological developments.

One solution would be to process the data collected using devices or programs installed by users downstream of the electric meters (data taken from the meter or from the electrical panel). Therefore, data processing performed directly through electricity meters is excluded. The approach will always be user-centered.

R E F E R E N C E S

[1] Óscar García, Javier Prieto, Ricardo S. Alonso, Juan M. Corchado, *A Framework to Improve Energy Efficient Behaviour at Home through Activity and Context Monitoring,* Rev. Sensors (Basel), 2017, August;

[2] Per Lynggaard, Knud Erik Skouby, *Complex IoT Systems as Enablers for Smart Homes in a Smart City Vision*, Rev. Sensors (Basel), 2016, Noiembrie;

[3] Chiehyeon Lim, Kwang-Jae Kim, Paul P. Maglio, *Smart cities with big data: Reference models, challenges, and considerations*, Rev. Cities 82 (2018) 86–99;

[4] Sophia Maalsen, *Smart housing: the political and market responses of the intersections between housing, new sharing economies and smart cities*, Rev. Cities 84 (2019) 1–7

[5] http://www.smartgrids-cre.fr/index.php?rubrique=dossiers&srub=gestion-donnees&action=imprimer

[6] Borgia, E. (2014). *The Internet of Things vision: Key features, applications and open issues.* Computer Communications, 54, 1-31.

[7] GhaffarianHoseini A., Dahlan N.D., Berardi U., Makaremi N. *The essence of future smart houses: From embedding ICT to adapting to sustainability principles.* Renew. Sustain. Energy Rev. 2013;24:593–607. doi: 10.1016/j.rser.2013.02.032.

[8] Jin J., Gubbi A. *An information framework for creating a smart city through internet of things.* IEEE Internet Things J. 2014;1:112–121. doi: 10.1109/JIOT.2013.2296516.

[9] Ericsson Networked Society City Index. 2016. Available online: https://www.ericsson.com/networked-society?gclid=CI-Jsoztp84CFRSeGwodOMsCQg

[10] Chourabi H., Nam T., Walker S., Gil-Garcia J. R., Mellouli S., Nahon K., . . . Scholl, H. J. (2012). *Understanding smart cities: An integrative framework.* Paper presented at the System Science (HICSS), 2012 45th Hawaii International Conference on.

[11] Morabito, V. (2015). *Big Data and Analytics for Government Innovation Big Data and Analytics* (pp. 23-45), Springer