BIG DATA FOR SMART ENERGY SYSTEMS

HANKEN School of Economics
Department of Marketing
2019

FORWORDS

This report is a result of the SESP (Smart Energy Systems Research Platform) project, funded by Regional council of Ostrobothnia, municipalities, University of Vaasa and HANKEN School of Economics.

The overall aim of the two-year SESP project (2017-2019) is to build up the new energy system research platform. The platform consists of physical laboratory environment for electricity distribution monitoring, protection, control and simulation, a Big Data collection and management system and business concepts that are based on the developed platform.

The SESP project consists of 9 work packages. Work package 8 is about *New Service Concepts and Business Models and* contributes to the smart energy systems research platform, and Vaasa´s Smart Grid solutions by developing a *system* for *new service development and a framework for understanding, analyzing and developing new business models.*

WP 8 has listed five deliverables of which this report met up to the fifth one*, to present an information package for companies explaining what Big data is and how it can be used for business development* (D8.5).

This report is based on findings presented in WP6, WP7 and a literature review on best practice for managing and exploiting big data. A special thanks goes to research assistant Patrick Latvasalo who has dig into definitions of Big data and analysed how companies can benefit from this type of data.

Peter Björk
Work package 8 leader

1. Smart house, smart living and big data – important components in smart energy systems

In the process of deep learning for customer convenience, family well-being and energy use optimization, the marriage between Big data analytics, machine learning, and human insight is to be benefitted from. Smart houses or smart homes, central components in smart living and micro grids, are residences that uses internet-connected devices and automated systems to enable the remote monitoring and management of appliances and systems, such as lighting, heating / cooling, home security, window and door operations, air quality, ventilation, and home entertainment. Smart homes can also be featured by domotics, home robotics, and personal health surveillance systems.

There are today plenty of companies offering different types of solutions for smart homes. A closer analysis of the solutions offered prove that the home automation and monitoring systems today available on the market, as for example the ABB-free@home solution, mainly focus on one residence at a time leaving micro grid operation to other types of companies and solutions.

Micro grids, designed to operate autonomously or synchronized with the grid, link smart houses / homes together with renewable energy sources and storage solutions. In this Big data on three levels and of two types can be collected, stored and analyzed. There are signal data on household and grid level, and we can separate quantitative data from qualitative (text) data, primarily. Furthermore, houses and houses in the grid are only one arena where energy is produced and consumed. In a bigger perspective, people are travelling to and from their homes, have hobbies etc., meaning that smart living with a focus on sustainable energy consumption brings in a third level of data, the society level.

Based on this structure six areas for data collection is to be identified (Table 1)

Table 1: Levels and types of Big data

Data for Big Data		
	Quantitative data	Qualitative data
Household level data	Energy use / hour Use of appliances	Factors facilitating / hindering change of behavior Analysis of what works well and complaints
Grid level data	Energy production and consumption, storage and optimizing	Analysis of what works well and complaints
Society level data	Societal factors influencing energy production, distribution and consumption. Market price information Weather condition Transportation systems	Measures of opinions in the society and trends

The three levels presented in Table 1 are integrated and should be analysed together. However, this report focuses on the "Household level". We define Big data and discuss how Big data can be used for business development.

2. Big data analytics a tool for business development and smart energy management

Big data analytics for business development and smart energy management is according to Zhou, Fu and Yang (2016) a step by step process consisting of seven phases, starting with Data collection, Data cleaning, Data integration and Data mining, for in the next step be used for Representation, Visualization and Application. The last two phases are intelligent decision making and real-time interaction, and smart energy management. In full use, Big data analytics in smart grids benefit all actors in the grid, utilities, power transmission and distribution companies as well as end-users, i.e. households. Utilities may use big data for generator planning and optimization, efficiency improvement and renewable energy planning. Power transmission and distribution companies benefit from big data analytics by improved grid planning, real-time sensing, fault detection and asset management. The end-users may use data and information to forecast load, and benefit from demand response and take advantage of dynamic pricing.

2.1. Big data defined

Big data as a concept and phenomenon has rendered a vast amount of different explanations and definitions. A summary of a selection of explanations of Big data presented in Table 2 bring us to the following definition,

> *Big data consists of a huge data set characterized by high volume, velocity and variety, and can be used to create actionable insight for value creation, co-creation and facilitation.*

Table 2. Big data - definitions

Author	Year	Definition
Bello-Orgaz, Jung & Camacho	2016	Big data refers to datasets that are terabytes to petabytes (and even exabytes) in size, and the massive sizes of these datasets extend beyond the ability of average database software tools to capture, store, manage, and analyse them effectively.
De Mauro, Greco & Grimaldi	2016	Big Data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.
Hashem, Yaqoob, Anuar, Mokhtar, Gani & Khan	2015	Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.
Wamba, Akter, Edwards, Chopin & Gnanzou	2015	Therefore, we define 'big data' as a holistic approach to manage, process and analyze5 Vs (i.e., volume, variety, velocity, veracity and value) in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages.

Chen, Mao & Liu	2014	In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time.
Chen & Zhang	2014	More commonly, Big Data is a collection of very huge data sets with a great diversity of types so that it becomes difficult to process by using state-of-the-art data processing approaches or traditional data processing platforms...More generally, a data set can be called Big Data if it is formidable to perform capture, curation, analysis and visualization on it at the current technologies.
Kitchin	2014	In other words, big data consists of massive, dynamic, varied, detailed, inter-related, low cost datasets that can be connected and utilised in diverse ways, thus offering the possibility of studies shifting from: data-scarce to data-rich; static snapshots to dynamic unfoldings; coarse aggregation to high resolution; relatively simple hypotheses and models to more complex, sophisticated simulations and theories.
Rao	2014	Big data refers to the growth and availability of large volumes of data, both structured and unstructured. Such an exponential volume of data could not be analysed by the traditional software used to handle databases.
Batty	2013	'... any data that cannot fit into an Excel spreadsheet ...' This implies that big data is not a new concept but exists in every era where the tools for data processing are always being stretched by increasing size…Much if not most of what we now call big data is produced automatically, routinely, and by various forms of sensors.
Kumar	2013	The term big data referred to the phenomenon of advancing trends in technology that enabled a fresh approach to assessing the world and arriving at decisions. It represented an approach wherein decision making was largely dependent upon data and analysis instead of experience and intuition. Some experts also employed the term 'big data' to describe datasets whose size was beyond the capability of normal database software device to capture, store, handle and assess.
Gartner	2012	Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.
TechAmerica Foundation's Federal Big Data Commission	2012	Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.

2.2. Big data for smart energy systems – customer insight

Restructuring the business landscape and being customer-oriented command, a business approach leaning on customer insight. Therefore, the discussion in this section has a focus on the end-users and how big data can be used to develop energy services (EaaS) of value.

Generally, it can be concluded that big data has both potentials and challenges when it comes to smart energy system development. Even if the development is technology driven, customers must be involved. Consequently, Big data can be used to:

1) **Find new ideas for business development.** Big data brings forth the possibility to identify emerging trends and by that be a source for inspiration and development. Crowdsourcing and open innovation linked to Big data can also be used by companies to enhance their research, development and design capacities through distributing the tasks to the public willing to help voluntarily. New ideas in companies can result in either incremental or radical innovations (or both), and the information stored, digital in format, can be company internal information and/or collected externally. New value can be created from ideas that have its origin in data retrieved from reconstructed and merged big data resources (Chen, Mao & Liu 2014: 204). By utilizing big data during the idea generation process, it might provide a way to identify new markets and suitable innovations intended for use in these markets (Chen & Zhang 2014: 317-318). Additionally, it can be benefitable to utilize different technological devices and support systems in conjunction with Big data to help with the collecting, transferring and storage processes. To exemplify, Big data combined with IoT (Internet of Things) allows organisations to gain real time access to data that might provide insights for novel ideas and innovation that provide value to the organization (Hashem, Yaqoob, Anuar, Mokhtar, Gani & Khan 2015: 113). As society and development is moving forward at a faster pace it becomes critical for businesses to notice symptoms, warnings and signs of the future to come, i.e. weak signals, that will help the company to prepare, react and reorient their strategies for the future changes in the market (Holopainen & Toivonen 2012: 198-201).

2) **Collect data from test bench and pilot areas.** In test situations a lot of data can be collected, and a lot of potential synergies can be found between Big data and Internet of Things-applications, which can be utilized for concept development and test (Chen, Mao & Liu 2014: 177). Concepts can be tested and analysed by collecting Big data using sensor-technology that is applied to smart houses and smart micro grids. Especially the availability of real time data supports concept development and test (Chen, Mao & Liu 2014: 194). By collecting data of consumers' actions, e.g. energy use, through Big data instead of their opinions, e.g asking about energy consumption through a questionnaire or survey, it is possible to obtain more relevant information of what consumers really think and how they finally behave (Auger & Devinney 2007: 377-379). Big data can also be utilized for improving development operations arriving from factors within the organization and its products. Big data can potentially be used to detect security threats and safety loopholes in products or services during the testing phase (Chen, Mao & Liu 2014: 204

3) **Constant improvements.** IoT applications combined with Big data allows companies to monitor and analyse their services and products in use in order make adjustments that makes them more efficient and optimized (Chen, Mao & Liu 2014: 198). To exemplify, the installation of IoT applications into a smart city can provide the city with valuable big data that can assist in decision making regarding city's services and their possible enhancement (Chen, Mao & Liu 2014: 198). It is possible to apply IoT sensors into several different locations and situations for the purpose of collecting big data in real-time which will provide valuable insights for companies, organizations and cities (Sun, Song, Jara & Bie 2016: 769). Furthermore, big data might also be used to discover the most optimal price point for services which will in turn benefit both the organisation and its customers in the end. Additionally, the data can help organisations with identifying sectors where resources should be allocated to in order to maximize efficiency of operations (Chen, Mao & Liu 2014: 202).

Business development in energy systems can take many forms, organisational development, process development, business network development, product development, and customer insight development. With a focus on Big data, there are challenges to look out for.

3. Big Data for Business development - challenges

There are some challenges with using Big data that relates to the general issues of handling large and complex data sets. One might encounter issues with compatibility when combining different kind of data sets from different sources that comes in the form of excessive and irrelevant data and data that behaves in a distinct manner (Chen, Mao & Liu 2014: 203). Additionally, discovering valuable information through Big data can be difficult as there exist challenges related to how one collects, store, search, share, analyse and present the data (Chen & Zhang 2014: 318). Furthermore, storage space and costs for Big data is a great concern as the amount of data is increasing at a rapid exponential rate, which generally might lead to situations where valuable data is lost (Chen & Zhang 2014: 319).

The possible issue with certain big data is the underlying biases and subjective information that it may, in some instances, contain (Kitchin 2024: 8-9). Therefore, it is important to keep this concern in mind when analysing and using Big data for business decisions, especially when the decisions will affect new and future investments and ventures. Moreover, sometimes the source of the data might be unknown or not verified, which might lead to poor quality of the data collected (Hashem, Yaqoob, Anuar, Mokhtar, Gani & Khan 2015: 110). The quality of the data impact on value creation (Chen, Mao & Liu 2014: 203-204). In this respect it becomes important to educate all employees about Big data within an organisation in order to prevent trust issues and gain the full potential of the value it provides (Wamba, Akter, Edwards, Chopin & Gnanzou 2015: 243). However, processing and transforming Big data into usable and valuable information requires a lot of time, therefore the time aspect is critical to consider when planning to utilize Big data in business operations (Cai & Zhu 2015: 3). Moreover, the possible issue with collecting data later or withholding collected data for a long time is that the data might get outdated and invalid, which in turn might negatively affect decision-making in organisations (Cai & Zhu 2015: 3).

There is also always the risk of obtaining too much of data that are refined into an amount of ideas that are challenging to handle and to realize in practice. This risk is especially prevailing in the process of idea generation when the goals and objectives (of the organisation) are not to be lost in an ocean

of new directions. With a new service on a new market the amount of data might be unexpected and therefore companies should be prepared. For example, storing and utilizing vast amount of big data requires adequate IT systems (De Mauro, Greco & Grimaldi 2016: 125). In the same vein, handling and processing of Big data has its challenges and one might encounter issues when transferring and converting Big data (Chen, Mao & Liu 2014: 202-203). Consequently, organisations might have to invest in data centres to be able to continue using Big data adequately in the future (Chen, Mao & Liu 2014: 178). Nevertheless, the challenge today is to keep the Big data processing systems up to date with the rapid increase of data that is created every second (Hashem, Yaqoob, Anuar, Mokhtar, Gani & Khan 2015: 112).

In open innovation settings, when external Big data is collected and shared, there is the risks of possible information leakage (Frishammar Ericsson & Patel 2015: 76). However, sometimes the combination of external and internal knowledge might lead to unclarity of the ownership of the data as the borders of internal and external (own and public) data becomes diffused. In worst case, it can lead to legal issues (Kostkova, Brewer, de Lusignan, Fottrell, Goldacre, Hart, Koczan, Knight, Marsolier, McKendry, Ross, Sasse, Sullivan, Chaytor, Stevenson, Velho & Tooke 2016: 1-2). A very big issue today is data and system security. For organisations this means investments in operation, management, soft- and hardware (Chen & Zhang 2014: 321).

To master big data for system development requires special competence and skills that can be difficult to find (De Mauro, Greco & Grimaldi 2016: 126-127). There are today ample evidences of how difficult it is to finding talents with adequate know-how and skills for Big data management (Chen & Zhang 2014: 321).

References

Auger, P. & Devinney, T.M. (2007). Do What Consumers Say Matter? The Misalignment of Preferences with Unconstrained Ethical Intention. Journal of Business Ethics. 76: 361–383. Available on: https://doi.org/10.1007/s10551-006-9287-y.

Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography. 3: 3, 274–279. Available on: https://doi.org/10.1177/2043820613513390.

Bello-Orgaz, G., Jung, J.J. & Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion. 28: 45-59. Available on: https://doi.org/10.1016/j.inffus.2015.08.005

Cai, L. & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal. 14: (2), 1-10. Available on: http://dx.doi.org/10.5334/dsj-2015-002.

Chen, M. Mao, S. & Liu, Y. (2014). Big Data: A Survey. Mobile New Appl. 19: 171–209. Available on: DOI 10.1007/s11036-013-0489-0.

Chen, C.L.P. & Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences. 275: 314-347. Available on: https://doi.org/10.1016/j.ins.2014.01.015.

De Mauro, A., Greco, M. & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. Library Review. 65: (3), 122-135. Available on: https://doi.org/10.1108/LR-06-2015-0061.

Frishammar, J., Ericsson, K. & Patel, P.C. (2015). The dark side of knowledge transfer: Exploring knowledge leakage in joint R&D projects. Technovation. 41–42: 75-88. Available on: https://doi.org/10.1016/j.technovation.2015.01.001.

Hashem, I.A.T., Yaqoob, I., Anuar. N.B, Mokhtar, S. Gani, A. & Khan, S.U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems. 47: 98-115. Available on: https://doi.org/10.1016/j.is.2014.07.006.

Holopainen, M. & Toivonen, M. (2012). Weak signals: Ansoff today. Futures. 44: (3). April 2012, 198-205. Available on: https://doi.org/10.1016/j.futures.2011.10.002.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. GeoJournal 79: 1-14. Available on: DOI 10.1007/s10708-013-9516-8.

Kostkova, P., Brewer, H., de Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R.A., Ross, E., Sasse, A., Sullivan, R., Chaytor, S., Stevenson, O., Velho, R. & Tooke, J. (2016) Who Owns the Data? Open Data for Healthcare. Front. Public Health 4: (7). 1-6. Available on: doi: 10.3389/fpubh.2016.00007

Kumar, T.V. (2013). Netflix: Leveraging Big Data to Predict Entertainment Hits. *IBS Center for Management Research*. 1-19. Available on: https://www.thecasecentre.org/educators/products/view&&id=115033.

Rao, A.S. (2014). Amazon's Big Data Strategy. *IBS Center for Management Research*. 1-15. Available on: https://www.thecasecentre.org/corporate/products/view?id=119853.

Sun, Y., Song, H., Jara, A.J. & Bie, R. (2016). Internet of Things and Big Data Analytics for Smart and Connected Communities. IEEE Access. 4: 766-773. Available on: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7406686.

Wamba, S.F., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics. 165: 234-246. Available on: https://doi.org/10.1016/j.ijpe.2014.12.031

Zhou, K., Fu, C. & Yang, S. (2016): Big data driven smart energy management: from big data to big insights. *Renewable and Sustainable Energy Reviews,* Vol. 56, pp. 215-225.