Collaboration in Higher Education for Digital Transformation in European Business

Impact Assessment of Big Data Analysis and Application Cases

– A Cross-Country Comparative Analysis

Margareta Teodorescu, margareta.teodorescu@fh-bielefeld.de
University of Applied Sciences Bielefeld/Germany
Jan Budik, budikjan@gmail.com & **Jan Luhan**, luhan@fbm.vutbr.cz,
Brno University of Technology/Czech Republic;
Luca Alfieri, luca.alfieri@ut.ee, Tartu University/Estonia;

This working paper is a result of the EU Erasmus Plus project "Collaboration in Higher Education for Digital Transformation of Corporate Businesses" (CHEDTEB) and essential ideas come from the intensive exchange with IT experts and company managers of the working groups "Big Data" and "Blockchain technology". Further information about Big Data, algorithms, AI and Blockchain technology can be found on our project webpage.

http://www.chedteb.eu/

The authors are grateful for the lively discussions within various Big Data workshops and would like to thank in particular following colleagues for their valuable ideas, and suggestions:
Martin Fridrich, Lukáš Novák – both from Brno University of Technology/Czech Republic,
Viire Täks, Sherif Sakr both from Tartu University/Estonia,
Rainer Lenz, Bernd Kleinheyer, Tahir Lushi - all of University of Applied Sciences Bielefeld/Germany

Co-funded by the Erasmus+ Programme of the European Union

Obsah

I. Introduction

As the digital revolution is on the way to fundamentally affecting and transforming industries, today's organizations are able to access more data than ever before. Companies are now creating new ways to harness all the data they generate in order to improve the operational efficiency of their business activities, optimize business processes, and find solutions to challenges that will drive their profits and growth.

However, the huge amounts of data companies generate every second are of no value, as long as they do not know how to put Big Data to work. This paper aims at exploring – in a cross-country framework – how Big Data analytics has forced companies from different sectors to reconsider and experiment with their business models and how this contributes to their innovation and performance. By using a comparative study of companies in selected EU countries – Germany, Estonia and the Czech Republic –, the focus mainly lies on exploring the impact that digitization and Big Data technologies have had on businesses as a whole and how companies have succeeded in making digital technologies a competitive advantage. The analysis is centered on assessing the effect that digitization and Big Data technologies have had from an economic and competitive perspective.

This paper applies case study-use methodology to compare the companies in the three EU member countries. This involves use cases from Germany, Estonia and the Czech Republic – five, two and three cases, respectively – to compare and analyze different sectors of the economy. The main aim is to demonstrate – based on selected cases – the potential of Big Data analysis as one of the key factors in digital transformation. Contrary to single-case settings addressing specific business cases, this article focuses on a compilation of possible applications and implementation scenarios that represent various perspectives.

After a short introduction, the second chapter presents firstly five case studies from different sectors of the German economy – transport and logistics, medicine, consumer goods, media and e-commerce. The chapter continues with two Estonian case studies from the financial services sector (FinTech) and the public service and finishes with further three cases from the Czech banking, telecommunication and agricultural sectors. The comparison has a very wide basis, as it aims at identifying insights about countries and sectors as well those that have reached different levels of development in terms of digitization.

II. Big Data Application Cases

This chapter provides an overview of some of the most innovative Big Data solutions from three member countries of the European Union. It starts by presenting five use cases from Germany (GE), followed by two uses cases from Estonia (ES) and three from the Czech Republic (CZ).

GE1 - Smart Port Logistics – *Hamburg Port Authority (HPA)*

As Germany's largest universal port, Hamburg is of vital importance for supplies to European domestic markets with a consumer population of up to 450 million. It makes up about a tenth of the city's total area and is located right at its heart. However, this central location of the port poses a number of tough challenges for its operator, Hamburg Port Authority (HPA)[1]. Among others, these problems include how to effectively manage burgeoning commodity flows within a limited space, avoid truck waiting times and deliver precise information for better decision-making. One of the main challenges is how to use the existing infrastructure, given the limits as to how many roads, railways and waterways can be extended in the confined port area.

The technology solution providers that HPA worked with were T-Systems, Telekom Innovation Laboratories, DAKOSY and SAP[2]. They helped with managing network traffic flows more efficiently by introducing the cloud-based SPL (smartPORT logistics) information and communication system, which provides information about more than 140,000 truck journeys in the port and optimizes commodity flows. As a result, updated information and forecasts about traffic within the port can be accessed via mobile devices as well as telematics units installed in driver cabs. The system also gives drivers approaching Hamburg information about traffic conditions such as bottlenecks and free parking spaces. This improves the traffic situation within the narrow confines of the port and cuts delivery times.

Traffic management usually generates large volumes of data. More than 8.000 trucks undertake approximately 140.000 journeys at the site every day. Furthermore, more than 100 rail companies operate services at the facilities, making the Port of Hamburg Europe's biggest rail hub. As such, the volume of traffic and transport data is enormous. The challenge increases even more, as there is a large number of parties involved (e.g. port authorities, shipping agents, carrier companies, parking lot operators, terminal operators, etc.), all of whom have their own systems and types of data. While problems related to the technical infrastructure have been resolved, capturing and applying data, especially traffic data accurately, remains a challenge.

Although individual companies have optimized their own processes and systems, the traffic system as a whole has not been optimized yet. SPL enables the port authorities as well as transport and logistics companies to monitor transport consignments in real time so that goods can be moved

1 Bitkom: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data*, p. 183 (2016)

2 Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele*, p. 55 (2015)

more efficiently and more safely. Thus, the movement of trucks and containers is more efficient, as the right information is available at the right time and the right place. Drivers can get an updated overview of the traffic and their destinations, all delivered directly to their mobile devices. The traffic flow and the goods transport chain are interlinked (including estimated time of arrival, or ETA), and are thus optimally coordinated. The next step will be to extend the system to include more participants and sources of information, such as data about the availability of containers. Shipping companies, terminal operators, ship-owners and storage facilities, some of which are still using their own systems, could be integrated in the future.

The project provides each participant at the commercial life of the port with major advantages. All shippers have access to traffic and infrastructure information, which makes it easier for drivers and dispatchers to communicate with each other whilst the different fleets are monitored on a permanent basis. The port is improving the efficiency of its road network and developing additional services; port companies are also boosting their activities.

In the future, Smart Port Logistics can be extended to more logistics areas of the port such as rail freight and shipping traffic, as these can be optimized through traffic management, too. The current optimization problems in these areas can be mapped with the Big Data in the cloud, just as in the case of road traffic. The same know-how, with little modifications, can be applied to ships and trains as well.

GE2 Personalized Treatment for Tumor Patients – *National Center for Tumor Diseases (NCT)*

The common goal of both doctors and scientists that work together at the National Center for Tumor Diseases (NCT)[3] in Heidelberg is that of defeating cancer. Every year 10,000 patients are treated here and – as every tumor is unique – new forms of therapy are constantly being developed. The Center has implemented the IT-solution *"Medical Research Insights"* based on SAP HANA, which helps scientists in developing individual methods of cancer treatment.

The challenge the doctors and scientists at the Center are faced with is that the success chances of the various treatment methods for each single tumor patient can only be vaguely predicted. The therapies – which are costly and

3 Bitkom: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data*, p. 196 (2016)

beset with side effects – may not always improve the patient's medical condition[4].

However, complex diseases need complex analyses. Altogether, there are more than 100 different types of cancer. Moreover, each type of cancer differs in its characteristics and development. That is why it is very important that patients get individual and personalized treatment. In order to develop such an individualized treatment in a fast and effective way, the NCT needs to analyze large amounts of data in real time. In the past, the Center had to browse through a wide range of databases, create lists of patients, print patient files and check manually to see if the criteria for one of the ongoing clinical studies were met.

The solution to this challenge means providing patients from the first day of their diagnosis with the optimal treatment. This implies determining the most promising therapy for each patient's genetic profile and consequently, analyzing a high volume of both structured (tumor documentation, medical files, clinical studies etc.) and unstructured data (doctors' letters, handling instructions, test studies, press releases, etc.).

The technical solution for NCT was developed by SAP and is called SAP HANNA. It enables its users – including doctors, scientists and other medical personnel – to work more efficiently and simply. The application has been developed together with the users and makes possible clustering patients according to various characteristics. The solution allows combining patient data from different sources and visualizing all information of the patient's history in a graphical timeline. By using this new solution, employees are able to record and analyze large amounts of data about every patient in real time. All the information from Doctors' letters, MRT results, genetic analyses or cancer registry files flows together centrally. In this way, it is possible to predict within a short time interval which therapy promises to be the most successful.

The volume of data, which is being gathered, processed, analyzed and evaluated at the Center, is huge. NCT uses more than 15,000 protocols to treat more than 10,000 patients every year. Every patient generates up to 1200 data points. The use of Big Data technologies enables the integration of all this data with a high volume of data from various sources. The analysis of the patient data is extended to the analysis of unstructured data, e.g. doctors' letters. These analyses, which took weeks in the past, can be now carried out within minutes and the findings can be used for the immediate benefit of the patient.

4 Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele*, p. 107 (2015)

The biggest win for NCT and its patients consists in the substantially increased transparency of patients' histories as well as their faster and much better matching with appropriate studies. Thus, besides patients´ data, a real time identification of tumor types is possible and, in this way, the most effective treatment method for the patient can be generated. As a result, the diagnosis of tumors is accelerated and the life expectancy of the tumor patients is extended.

The Personalized Treatment for the Tumor Patients project at the National Center for Tumor Diseases provides an example of how medical work can be revolutionized through intelligent combinations of technical expertise and practical design at the user interface. Only one single and simple interface enables NCT's employees to browse and filter a high volume of data. The SAP HANA-platform provides patient care support by delivering detailed information and timelines of treatment activities, operations, chemotherapies and home visits.

NCT and SAP are planning to further continue their cooperation and extend the analysis of patients' data based on the SAP HANA-platform. NCT wants to use data analyses in real time to support each type of diagnosis and every step of the ongoing therapy, aimed at improving the life and future of people who suffer from cancer.

GE3 - Demand Signal Management – *Beiersdorf AG*

Beiersdorf AG[5] is a global consumer goods company offering well-known brands such as NIVEA, Eucerin and Hansaplast. As the company wants to react individually to its customers' needs, understanding the demand and the buying behavior of the consumer in every single market on a global level becomes crucial for the coordinated positioning of the brand portfolio.

The challenge Beiersdorf AG is faced with is that it needs a consistent and comprehensive overview of their own brands as well as of the relevant competitors in the respective markets in order to plan and run appropriate marketing activities[6]. Currently, understanding and interpreting the different signals of demand is done manually, which is very time-consuming and prone to error. Thus, the technical challenge consists in implementing a central platform which can detect and illustrate patterns and signals of demand. These signals originate from various data sources and include data from market research as well as sales and panel data.

5 Bitkom: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data*, p. 180 (2016)

6 Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele*, p. 93 (2015)

Therefore, data harmonization becomes an important condition for an informative reporting and analysis system.

The Demand Signal Management application is based on SAP HANA and offers Beiersdorf AG a central platform for the aggregation of all market relevant data. In addition to simply harmonizing the data, the Demand Signal Management provides new analytical perspectives by analyzing new product attributes. It also shows the main reasons for the market share trends of their own brands and of the brands of competitors as well. In this way, the decision process is becoming faster and more effective. Understanding about brand development in the single markets increases and the brands can evolve further in a more specific way. As a result, both revenue potentials and the market value can increase.

Beiersdorf AG works with very large amounts of data. The market data of its own products and brands and those of the competitors from over 60 countries (over 500 databases) is collected, harmonized and analyzed. Processing data brings new insights and findings, which makes it possible to focus on fast-growing brands and emerging markets. The availability of the analyzed data in real time allows an appropriate reaction to competitors' activities as well as the tracking of in-house product launches. Conversely, the application of sophisticated Big Data technology provides a global overview of the brand strength, brand positioning, and the market shares of their single brands as well as those of their competitors.

The general benefits to the company generated by this Big Data solution are, without question, multiple and diverse. Firstly, faster reporting by automated data harmonization reduces time differences between global and local reporting. Secondly, the combination of different KPIs from various sources enables the companies to get a better view on the data and serves as a basis for relevant business decisions. For this purpose, the reasons for changing trends and market shares need to be detected and – by means of a real time application – applied as well. In the future, Beiersdorf AG plans the integration of even more data sources aimed at maximizing system potential.

GE4 - Big Data-driven Marketing – *ProSiebenSat.1 Media AG*

The German media company ProSiebenSat.1 Media AG[7] sells classic TV advertising. Furthermore, the company is involved with several e-commerce companies, which means providing air time for commercials promoting the e-commerce offering. In this way, ProSiebenSat.1 takes great interest in precisely identifying which share TV commercials have within the value creation of an advertised e-commerce company.

7 Bitkom: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data*, p. 215 (2016)

In this context, the company is interested in systematically answering specific questions:

- Which measurable influence does television advertising have on online behavior?
- How many users visit an e-commerce website because of the TV commercial?
- How much turnover do users who verifiably visited the website due to a TV commercial generate in a certain period of time[8]?

Using a Big Data approach, ProSiebenSat.1 Media AG developed a solution to measure the TV influence on website traffic. From a previous joint project with inovex GmbH ProSiebenSat.1 compiled a Hadoop cluster containing collected traffic data. The cluster was used by the inovex data science team for analytic purposes. An agile implementation ensured that all challenges that usually accompany an innovative project could be solved in time and within budget. The first stage of the solution is already in productive use and further stages are currently in a proof of concept phase.

As standard solutions for the analysis of web traffic generally cannot relate events from outside the internet – such as broadcasts of TV commercials – to online behavior, ProSiebenSat.1 decided to apply a custom Big Data solution. From a business value perspective, the new solution provides a meaningful, measurable proof of the economic value of TV advertising for the promotion of e-commerce offerings. Besides, it ensures creating predictions for the efficiency of planned TV advertising as well as optimizing media planning for TV advertising.

With this project, ProSiebenSat.1 pioneered Big Data-aided advertising evaluation, connecting online and TV effect analysis for the first time. As yet, no comparable implementations have been published in the media industry. The important innovation characteristics of this Big Data-driven marketing solution are:

- Using the existing Big Data pool to answer business relevant questions;
- General validation of TV advertising impact;
- Comparability of TV and online advertising impact;
- Specific objectification of the advertising services ProSiebenSat.1 offers its e-commerce partners in joint ventures;
- Strengthening ProSiebenSat.1's position in the market against its competitors.

8 Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele*, p. 75 (2015)

By implementing this Big Data solution ProSiebenSat.1 becomes the leading data hub for TV analytics.

The company is also planning future ambitious projects and developing several Big Data and data science initiatives such as app tracking, synchronizing TV and online advertising campaigns, and integrating weather data.

GE5 - BRAIN Platform – Meaningful Personalization – *OTTO*

Otto[9] is one of the world's leading mail order and e-commerce companies, aiming to provide the best customer experience in e-commerce. One of the most important features that distinguishes OTTO from its competitors and defines its success is the personalization of the experience, tailored according to the individual customer's needs.

The challenge in providing meaningful personalization (besides the need of a detailed picture of the user) resides in collecting, assessing, processing, and evaluating huge amounts of data from various sources. Complex algorithms and advanced statistical methods require an enormous processing power and the results have to be made accessible in real time, when the customer is shopping online.

A key initiative in achieving this goal is the development of the BRAIN Platform, which integrates classical high-end databases (Teradada) with big-data technology (Spark, 0Hadoop, Hive, Impala and HBase on Cloudera) and a live stream processing engine (based on Ignite and Kafka). The BRAIN Platform provides a real time access to customer data, 360° customer profiling, advanced data science capabilities, and Big Data processing as well as classical data warehousing features. Data is loaded from various sources via ETL (Talend) or Messaging (Kafka) and the processing jobs are managed and orchestrated by Talend and Control-M. Classical warehouse data is stored in Teradata. Big Data is stored as parquet-files on Hdfs to be queried with Hive and Impala. 360°-Profiles are accumulated in HBase using Spark. Together these form an information warehouse. An in-memory database will enable extremely fast reporting. Analytics Stack provides an environment for deeper analysis, data science and machine learning using Spark with Python and Scala. Real time features are implemented according to the lambda architecture. A speed layer processes the click stream data while the user is still online using Ignite. A fast serving layer makes data from streaming and batch available for personalization in the retail outlet.

From a business model perspective, OTTO is an important force in the world of online retail. The company is using the latest technology and is actively shaping the market. As such, there are over 2,1 million items and around

9 Bitkom: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data*, p. 206 (2016)

5,200 brands at otto.de. The company offers convenient shopping on all user devices and is number 1 in Germany in Fashion and Lifestyle (B2C). With over 6 million active customers it receives 20 million orders per year.

ES1 - FinTech enterprises – *TransferWise*

TransferWise is one of the most important FinTech companies in the world and is specialized in peer-to-peer currency transfers. Although its headquarter is located in London (and it could presumably move towards Western Europe because of Brexit) for business plan reasons, the company was founded by Estonians and all the operating teams are located in Tallinn. The company offers lower prices than standard financial institutions (i.e. Western Union) and banks regarding trading in different currencies. TransferWise does not transfer the sender's money directly to the recipient but matches the same amounts of other users. TransferWise then pools these amounts in order to pay out the money transfers using a local bank transfer. This peer-to-peer process avoids the higher charges applied by banks and traditional money transfer services. In the last 8 years TransferWise has reached 4 million users and presently transfers an amount equal to 4 billion dollars on a monthly base. Thus, the users save 4 million dollars per day avoiding bank fees. TransferWise currently employs 1300 employees in 11 countries on four continents including more than 700 in Tallinn.

TransferWise business is based on Big Data technologies taking into account the amount of information that has to be stored and processed to make its matching method of peer-to-peer currency exchange mechanisms. Without the existence of Big Data and methods to handle it, the TransferWise business model could not exist. The problem of handling all this data and report information by every team of the company leads to partnerships with other services enterprises, e.g. Looker. In this way, ~~a~~ part of the job of the data analysts can be reduced and TransferWise can evolve into a more data-driven company.

In 2017 TransferWise launched its borderless account. It enables moving money between 40 currencies and provides in most cases a debit card and local bank accounts. TransferWise also finalized agreements with some standard banks. These agreements allow banks' users to exploit TransferWise services.

Recently, the company signed an agreement with the second-largest bank in France, Groupe BPCE, and aims at reaching similar agreements with further banks as well. All these imply new important challenges in the future, as the complexity regarding the large amounts and the velocity of data that needs to be processed will dramatically increase.

ES2 - Big Data and Machine Learning research at the service public economic policies – *Center of IT Impact studies (CITIS)*

The Center of IT Impact studies (CITIS) is a research center connected with the Johan Skytte Institute of Political Studies of the University of Tartu. CITIS aims to improve the daily e-public services in the country and to support the creation on new policies that could help Estonian enterprises. CITIES exploits auto-generated Big Data from Estonian public e-services. The research institute studies the effects of these services from a political, economic, and social perspective. A better understanding of Big Data issues implies the possibility of creating new e-services and training activities to develop analytical skills.

Among the Big Data projects being currently conducted at CITIS are the "Projects using micro-level tax declaration data for forecasting and policy analysis". These projects aim at building time series automated forecasting models and network simulation models using real-time tax declarations. It exploits data from monthly value-added tax and income tax declarations of the Estonian Tax and Customs Board from 2004-2016. The data also includes information on inter-firm transactions characterizing the network of Estonian companies.

The project produces the following prototypes which were delivered to the Estonian Tax and Customs Board:

- Automated forecasting model for tax revenues and economic indicators (the forecast was an automated ensemble of forecasts from six different individual models – Holter-Winters, ARIMA, conventional time series regression model, two LASSO models, and a random forest model);
- Simulation model of bankruptcy of an enterprise on its local network;
- Simulation model of public investment subsidies of an enterprise on its local network;
- Automated benchmarking of a single enterprise against similar enterprises;
- Automated sustainability index for a single enterprise.

The project will allow policy makers and common users of the models to forecast the economic situation of the different sectors of Estonia's economy. Policy makers will be able to produce smarter and faster real time economic policies. Moreover, the firms will observe a reduction of their administrative burden during the submission of the corporate reports.

Another Big Data-driven project of CITIS is called "How Estonian enterprises are connected to each other and to the world economy? Application of Estonian Tax and Customs Board Transaction data". The work uses data from Estonia's Commercial Registry dataset of firms' annual reports from 2016, VAT declarations and firm-product destination-market level trade dataset of the full population of exporting firms in Estonia. The work

analyzes the relationship between Estonian enterprises considering their level of productivity and their position in the value chain as well. The project shows the relative position of the firm in the value chain, the distance of the firm to the end user and the distance to the resource. The work illustrates how 29% of Estonian firms are exporters and 80% of enterprises are up to two steps from becoming exporters. The same result also applies to imports.

The results of the study suggest that exporting enterprises and firms sufficiently close to them are more productive than other firms that operate far from exporting sectors. The findings also show that a high level of added value depends on the distance between the firm and the consumer. Productivity is higher when the final consumer is far from them. In its last strategic plan, the Estonian government emphasized the importance of increasing exports and, in particular, high-tech exports. The results of the project will provide Estonian policy makers with precise suggestions as to how the country could increase its rank among export countries.

CZ1 - Internal bank data analysis – *Ceska Sporitelna, a. s.*

Data sources: ATMs, internet banking, applications for smart-phones, historical transactions.

Big Data analysis implementation

The bank, generally, collects a huge amount of data about their clients from realized transactions like ATM withdrawals and electronic transfers, which involves Big Data analysis. The banks can use these data across more verticals:

- customer spending behavior
- customer segmentation
- product cross-selling
- threat prevention

Customer spending behavior is based on searching spending patterns such as the day of the month, the month of the year etc. The bank can view your salary and how you change your behavior after you receive it, how much you send to a saving account, and utility providers. An analysis of these patterns can be used for risk management, for loan screening, mortgage evaluation or for offering you insurance products.

Customer segmentation is focused on creating specific profiles such as people who spend frequently, investors with risk aversion, customers who contract loans frequently and live with constant debt, customers who pay all debt on the last day, etc. This knowledge about customers helps to predict when customers will pay off the loans and develop detailed strategies.

The banks can use your personal data to do cross-selling and offer you other products based on your spending profile and ability to repay loans. It's much easier to sell a new product to current customers than to new customers.

Threat prevention is process which can alert a bank if something unusual has happened. The bank knows your profile and if you e.g. never withdraw all cash from your account, it is reported to the bank as a possible threat.

The Ceska Sporitelna bank uses information from their ATMs (currently 1561), branches (currently 634) and digital channels (currently 1,6 millions users). All data is collected in its own transaction and analytical systems.

The main reason why The Ceska Sporitelna bank uses data analysis is the daily processing of all transactions, interest rates etc.

Another area that has a growing tendency in data processing complexity is banking and financial market regulation. The banks process large amounts of data for regulators and do a lot of calculations. The last but no less important area is the processing of data for providing quality services to clients, especially in the area of solving their life situations, when the data helps offer the client the right service at the right moment.

The Ceska Sporitelna bank uses a variety of standard technologies which are very powerful in terms of the data processing volume. The bank has recently prototyped large data processing in technologies that are more suited to this area, such as Hadoop, Keboola, BigQuery, SnowFlake, Tableau, GoodData, etc. In addition, these technologies will enable it to enrich its own data with data from other third-party services.

CZ2 - Geodata analysis – *O2 operator*

Data sources: Anonymized signaling data from O2 network.

Big Data analysis implementation

The O2 Big Data Team are collecting anonymized signaling data from O2 network and use it for product called O2 Geodata. Signaling data contains a lot of information about calls, positions etc. O2 Geodata product is focused on customer privacy and offers a high level data output as service. This means that it necessary to work with anonymized data. There are 5 steps as to how O2 Geodata works:

- data collecting from O2 network
- data anonymization
- data processing
- recalculation to the whole Czech population
- outputs and statistics
- visualization

Data is collected from all O2 network which covers the Czech Republic. This means around 52 000 access points. An access point is called BTS (Base Transceiver Station) and it is a transceiver and receiver of the GSM signal. In the city of Brno, which is the second largest city in the Czech Republic, you can find thousands of BTSs and their sufficient coverage for geolocation.

The next step is data anonymization because customers privacy is very import for the O2 operator and all collected signaling data cannot be assigned to a specific customer. This anonymization process will erase the details about a customer; like, e.g., what his/her phone number is. Data processing is realized in the private and secure O2 system with an application of sophisticated processing methods. O2 operator covers around 40% of the Czech population and the results have to be recalculated to all the Czech population. This means that all geodata output statistics will be linearly extended to customers using different operators.

The output of O2 Geodata product contains basic anonymized statistics and geographic information, such as an interactive web presentation and an XLS table with predefined templates. The customers can import this output data to their own data analysis workflow and can create their own visualization.

Outputs of O2 Geodata are usually used for:

- transport planning
- tourism analysis
- retail shops
- short-term events like music festivals or concerts

CZ3 – Smart agriculture – *CleverFarm*

Data sources: Satellites, weather information, sensors located in the soil.

Big Data analysis implementation

The modern farming attitude is focused on growing process optimization and profit optimization as well. It's quite easy to get satellite pictures, information about historical and current weather, or even information about weather prediction. All these data is enriched with data from soil sensors and can be analyzed with machine learning. CleverFarm solution contains different types of sensors:

- weather station
- leaf moisture sensor
- field sensor
- soil sensor
- grain hall sensor
- silo sensor
- animal sensor

The Weatherstation provides an overview of meteorological conditions like temperature, rainfall, atmospheric pressure, humidity. CleverFarm solution allows to save historical data about these values and can help to plan on a daily basis or even for seasonal planning.

The Leaf moisture sensor measures wet leaves and you can see the drying process over time in a chart. This sensor is usually used for determining when leaves need to be sprayed. The sensor can also predict the risk of diseases and pests.

The Field sensor measures the temperature and the humidity of vegetation as well as monitors their microclimate conditions. This information is used to predict the occurrence of diseases or various pests. The monitoring of vegetation conditions can also be useful when applying plant protection products; it is able to determine their efficiency.

The Soil sensor measures both the temperature and the water potential of the soil. In addition to measuring the soil water potential, the sensor is also suitable for irrigation management.

The Grain Hall sensor is equipped with five temperature sensors and is the perfect sensor to put inside your grain storage hall. An increased temperature within the grain profile as compared to the surrounding area indicates higher humidity, which can be the main detrimental factor in the sale value of the harvest. It also reliably detects the occurrence of pests. The sensor can also be used to monitor conditions in silage pits where it can withstand temperatures of around 60 °C.

The Silo sensor can be used for grain or other harvests stored within siloes. The silo measures temperatures in 7-10 evenly spaced segments (depending on the height of the silo, up to 25m). This sensor reliably monitors indoor conditions as well as air-flow within the silo.

The Animal sensor measures the temperature and humidity to ensure and monitor animal welfare in stables and coops. The data is used to control ventilation, heating, and cooling. These sensors allows to monitor the stability of animals' housing conditions.

The output of this Big Data analysis can be focused on profit prediction and it can predict the occurrence of diseases and pests. All this data is stored in one place using even the basic version of the CleverFarm app, and it's possible to access it anytime, on any device, and anywhere you have an internet connection. The CleverFarm sensors monitor the temperature and humidity of crops and storage areas so customer can access the conditions information anytime.

A part of CleverFarms app is Land Registry, which manages farmland and always keeps updated data about landowners and leasing agreements,

pulled straight from publicly available catastral data. This feature is available only for the Czech Republic. [10]

III. Concluding remarks

This paper provides a cross-country comparative analysis of several Big Data use cases from Germany, Estonia, and the Czech Republic. Even though all three countries are members of the European Union, according to the Digital Economy and Society Index (DESI[11]), they differ from each other in terms of digital awareness and skills, integration of digital technology by businesses, digital public services, and research and development in the ICT sector. Thus, Estonia – one of Europe's pioneers in the field of digitalization and digital innovation – ranks eighth, Germany only twelfth and the Czech Republic, nineteenth[12].

In spite of the fact how the Digital Economy and Society Index ranks the three countries, we find that – at a sectoral level – companies across all countries exhibit a relatively high level of digitalization which makes possible to implement Big Data solutions and technologies in a very successful way.

Firstly, in the case of Germany, the five use cases provide valuable insights into the digitalization skills of companies across various sectors, like transport and logistics, medicine, consumer goods, media, and e-commerce. Even if the analyzed companies are very different from each other and have completely different business models, they all succeed to improve their activities and boost their profits by using new and innovative solutions about how to put their data to work. The use cases analyze different Big Data analytics solutions implemented by the Hamburg Port Authority (HPA), the National Center for Tumor Diseases (NCT), Beiersdorf AG, ProSiebenSat.1 Media AG and, respectively, OTTO.

Secondly, the Estonian use cases shed some light on the level at which Big Data analytics arrived so far in two sectors, which usually – when it comes to digital innovation and digitalization – enjoy much of the public attention. The first one refers to the financial services sector and provides an analysis of a Big Data solution at TransferWise, a FinTech enterprise. The second one presents a Big Data and machine learning solution that has been implemented at the Center of IT Impact studies (CITIS).

10 Source: https://www.cleverfarm.org/sensors

11 The Digital Economy and Society Index (DESI) is a composite index that summarizes relevant indicators on Europe's digital performance and tracks the evolution of EU member states in digital competitiveness.

12 European Commission, *The Digital Economy and Society Index (DESI)*, available at: https://ec.europa.eu/digital-single-market/en/desi

Thirdly, for the Czech Republic we described three different use cases from different fields. The first use case is focused on the banking sector, where we described how generally banks can use data analysis and how the Ceska Sporitelna uses its own data collected from ATMs, branches and electronic banking. The second use case is focused on geo data and on how the Czech O2 operator uses this signaling data for monetization. The third use case is focused on smart farming and on the ways Big Data analysis can be used for better growing and higher profits.

Irrespective of a country or a sector, the selected case studies clearly demonstrate the potential of using Big Data analysis and digital technologies, and the significant impact of the Big Data phenomenon on businesses and society as a whole. However, according to the DESI index countries under focus are significantly lagging behind other European competitors in terms of digital experience, abilities, and awareness and need to act now if they want to keep up with the increasingly competitive environment of the digital age[13].

The use cases have been selected so that they cover different levels and perspectives; in this way, the impact can be generalized. In all these perspectives, the potential for using Big Data analysis as a direction for an improvement of business activities and boosting profits is shown to be very high.

This comparative analysis shows the potential for the economies of the European Union member countries to change due to new disruptive technologies. In general, EU countries show different levels of digital awareness and skills as well as a different potential of exploring the impact of digitalization. Governments should develop and enforce more instruments in order to be able to cope with the ever increasing demand for digitalization and to support companies in becoming more competitive on the international markets.

13 PAC, Holistic Customer Experience in the Digital Age – A Trend Study for Germany, France and the UK, available at: http://blogs.adobe.com/digitaleurope/files/2015/09/StudyHolisticCXAdobe1.pdf

References

Abernathy D., *Using Geodata and Geolocation in the Social Sciences: Mapping our Connected World* (2016)

Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte. Leitfaden*, (2012)

Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele* (2015)

Bitkom, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., *Germany – Excellence in Big Data* (2016)

Christensen, Clayton M., *The Innovator's Dilemma; When New Technologies Cause Great Firms to Fail*, Brighton, Mass. (1997)

ICT revue, Big Data in Czech companies (2015) Available at: https://ictrevue.ihned.cz/c3-64232170-0ICT00_d-64232170-vyuziti-velkych-dat-v-ceskych-spolecnostech

European Commission, Directorate-General of Communications Networks, Content and Technology (October 2018), *International Digital Economy and Society Index 2018*

European Commission, *The Digital Economy and Society Index (DESI)*, retrieved from https://ec.europa.eu/digital-single-market/en/desi

PAC, *Holistic Customer Experience in the Digital Age – A Trend Study for Germany, France and the UK*, retrieved from http://blogs.adobe.com/digitaleurope/files/2015/09/StudyHolisticCXAdobe1.pdf

Tan P., Steinbach M., Karpatne A., Kumar V., *Introduction to Data Mining* (2018)

Thomas R., McSharry P., *Big Data Revolution: What farmers, doctors and insurance agents teach us about discovering big data patterns.* 1st Edition (2015)