Some Challenges and Lessons-Learnt from the Practice of Business Analytics

Manel Brichni
Luxembourg Institute of Science and Technology (LIST), 5, Avenue des hauts-fourneaux L-4362 Esch-sur-Alzette, Luxembourg

Dimitrios Kampas
Luxembourg Institute of Science and Technology (LIST), 5, Avenue des hauts-fourneaux L-4362 Esch-sur-Alzette, Luxembourg

Jorge Sanz
Business School and School of Computing, National University of Singapore (NUS), Singapore.

ABSTRACT

Business Analytics is a main driver of economic value in enterprises. It requires a continuous process that involves understanding, adapting and assembling solutions based on data footprints from within organizations and often, also from interactions between organizations and their main external stakeholders. In actual practice, building such solutions differs from one business situation to another. Each context brings distinct challenges and calls for specific needs. The collection of these contexts and experiences contributes to a better understanding of the Business Analytics domain. In line with the above considerations, the end objective of this paper is to offer a better understanding of the practice of Business Analytics and key aspects of this interdisciplinary domain. Specifically, a case-driven approach and a pragmatic research-oriented view is followed based on three real-case scenarios from Investment Fund, Manufacturing and Wealth Management. Varied infrastructure, heterogeneous functionality, continuous adaptation of analytics capabilities, cost of deployment and data quality and quantity are among the topics discussed.

Keywords: Business Analytics, Industrial Applications of Analytics, Data-driven Decision Making, Data Analytics, Natural Language Processing, Machine Learning, Case-driven Approach.

Received 14 July 2018 | Revised 28 October 2018 | Accepted 10 November 2018.

1. INTRODUCTION

Business Analytics is a relatively new field of practice with the potential of transforming organizations and enhance their business value through digitalization in organizations (Sanz & Chow, 2016). It is an interdisciplinary field that helps discover and validate business insights from structured and unstructured data, by carefully understanding the provenance of information in enterprises or their Lines-of-Business (LoBs) and using methods from areas of statistics and optimization, such as machine learning, data mining,

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

forecasting, et cetera. In addition, Business Analytics includes the deployment of these data-driven insights into the core of operations in the enterprise, thus making the analysis of data an integral part of new processes and practices.

Creating business value with data is a continuous process that requires to effectively consider the *design* and *deployment* of analytics solutions (Sanz J. , 2019) (Johnson, 2015). *Design* involves the process of developing and assembling analytics solutions, whereas *deployment* requires the consumption of the said solutions in particular business environments or operational contexts. These contexts may call for implementing different business processes where an analytics solution is made into a tangible contribution as it gets adopted by specific LoBs. As part of a continuous operation, the said Business Analytics solution will evolve and adapt. Such a practice requires a significant effort that may vary from one context to another where each environment presents its own challenges, requires specific capabilities and contributes to the understanding of the Business Analytics domain. For example, after a data-centric general model for predictive maintenance is discovered for a certain type of an extruder engine, the deployment of the said solution will suffer significant adaptation from plant to plant where the equipment may be used. Different utilization patterns, aging, variations of the actual version of the equipment and even talent across engineering teams will dictate how the maintenance process is deployed. As many practitioners have documented over the last five years, analyzing data footprints - also recently called "data science"- is relatively simple for designing analytics solutions (Sanz J. , 2019). On the contrary, the challenge lies in the effectiveness of the actual deployments. Demonstrating value and return-on-investment may take a long cycle of trial-and-error attempts until the right Business Analytics solution becomes available.

The aim of this paper is to study the challenges and harvest some lessons-learnt from different industries and LoBs in order to foster a practical understanding of Business Analytics. To this end, three real-case scenarios in the Investment Fund, Manufacturing and Wealth Management areas are considered. The respective enterprises have been digitizing essential internal functions laterally to their value chain, by ceaselessly injecting innovative technologies and growing data-centric services (Wright, 2015) (Kagermann, Wahlster, & Helbig, 2013) (Li, Jiang, Yang, & Cuzzocrea, 2015).

Considering the domain of Business Analytics, many relevant works have been published under different names (Brichni, Kampas, & Sanz, 2018) and for various domains discussing data, infrastructure, databases, cloud architectures and services, sensors and their interconnectivity and so on. Several case studies in the literature discuss some challenges from the practice of analytics (Lécué, Tallevi-Diotallevi, Hayes, Tucker, & Bicer, 2014) (Giannakos, Chorianopoulos, & Chrisochoides, 2015) (Lonn, Aguilar, & Teasley, 2013) (Bakharia, Kitto, Pardo, Gasevic, & Dawson, 2016) (Kennedy, 2014) (Davenport & Harris, 2017). A recent book on the challenges of Business Analytics from T. H. Davenport et J. G. Harris (Davenport & Harris, 2017) discusses the management and organizational culture challenges confronted by analytics in its different stages of evolution in enterprises and illustrated them through multiple practical scenarios. In many cases (Parilla, 2019) (Lécué, Tallevi-Diotallevi, Hayes, Tucker, & Bicer, 2014) (Giannakos, Chorianopoulos, & Chrisochoides, 2015) (Lonn, Aguilar, & Teasley, 2013) (Bakharia, Kitto, Pardo, Gasevic, & Dawson, 2016) (Kennedy, 2014), the insights are limited to a technical or usability assessment of specific analytics solutions, whereas the aim of this paper is to harvest lessons-learnt in order to contribute to the understanding of the Business Analytics domain from multiple facets. In (Hartzband & Jacobs, 2016), the

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

authors provide some recommendations related to data quality (i.e. accuracy, reliability, and completeness), the infrastructure suitability (i.e. servers, storage, and network) and the required expertise. However, some of the pointed recommendations, such as the analytics infrastructure, are not scrutinized and the liaison to the considered use case is opaque.

In (Hartzband & Jacobs, 2016), the authors present the challenges of deploying analytics in the healthcare domain to support operations and clinical practices, where they emphasize on the importance of data quality and integrity, suitable infrastructure as well as training and engagement. In the discussed use case, data is mainly structured and semi-structured following standardized formats. Therefore, those lessons might not be representative for other applications where data can be unstructured too. In (Lécué, Tallevi-Diotallevi, Hayes, Tucker, & Bicer, 2014) the authors discuss the challenges transpire from the application of smart traffic analytics on semantic web technologies, where challenges arise mainly due to the unstructured format of data and related technology needed. Indeed, recently, unstructured data have generated an increasing interest among enterprises and analysts (IBM, 2018), which led to new business challenges. In (Davenport, Analytics 3.0, 2013) (Giannakos, Chorianopoulos, & Chrisochoides, 2015), the authors discuss lessons-learnt from clickstream interactions in a video-assisted course.

One of the main domains that have been widely addressed by literature is education (Giannakos, Chorianopoulos, & Chrisochoides, 2015) (Lonn, Aguilar, & Teasley, 2013) (Bakharia, Kitto, Pardo, Gasevic, & Dawson, 2016) (Kennedy, 2014). (Lonn, Aguilar, & Teasley, 2013) discusses challenges related to the importance of exploiting several analytics techniques to better decision-making in education. Analytics for educational and learning purposes have shown important results and emphasized meaningful challenges and lessons-learnt. Nevertheless, the heterogeneity of the *Advanced Manufacturing* and *Education* domains make the conclusions emerging in the latter studies insufficient for understanding the former. For example, the usability gaps that students face (Lonn, Aguilar, & Teasley, 2013) should not be an issue in an industrial context using or developing analytics.

In conclusion, unveiling the main challenges observed in the deployment of Business Analytics in various application environments will pave the knowledge of more challenges and also potential generalizations from individual experiences in different domains. The remainder of this paper is structured as follows. Section 2 describes the research question and approach. Sections 3, 4 and 5 describe the three use case scenarios. Then, the identified lessons-learnt are detailed in Section 6. The paper ends with a conclusion and some perspectives in Section 7.

2. RESEARCH QUESTION AND APPROACH

The purpose of this paper is to identify some relevant business needs, challenges and opportunities with Business Analytics. Thus, the research question addressed in this paper is:

– What are the lessons-learnt from the practice of Business Analytics?

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

To proceed, an interdisciplinary approach is adopted, where multiple case-scenarios are considered from the Investment Fund, Manufacturing and Wealth Management industries. These industry segments deal with different problems and objectives and thus, face different challenges in the adoption of Business Analytics.

As a result, we categorize the lessons-learnt according to the challenges that the considered application domains are facing. First, we discuss technical challenges to the application of Business Analytics, namely infrastructure and its application to the problem at hand. Second, we refer to other organizational and business challenges relevant to the disruption that analytics brings along to typical enterprise LoBs.

The strength of the study is that we address the Business Analytics from the perspective of real needs in enterprises. This helps to identify realistic and valuable opportunities, but also contribute to the state-of-the-art with a cross-disciplinary approach for harvesting multiple insights from the use of Business Analytics in different settings.

In fact, we selected the aforementioned scenarios under the following two main criteria. First, the nature and characteristics of data are taken into consideration. Fund documentation, specifically in marketing collateral for Investment Funds, is essentially made of unstructured data; on the other hand, Manufacturing data is typically structured; and lastly, Wealth Management has both structured and unstructured data. In addition, the *velocity of data* differs among the three use-cases. For example, in Investment Funds and Wealth Management, data is historical, whereas, in most Manufacturing floor applications, data is mostly streamed in real-time. Second, the analytics domain is addressed from two other angles, namely, processing for the Investment Funds and Wealth Management scenarios, and infrastructure for the Manufacturing scenario. Thus, the combination of these use-cases helps identify and understand the deep challenges that typical companies face in making the "Business Analytics promise" real for their organizations.

3. SCENARIO 1: INVESTMENT FUND DOCUMENTATION ANALYSIS

Digitalization has been increasingly taking a prominent position in the life of companies (Loebbecke & Picot, 2015) (Kokin, 2013) (Sanz & Pang Yan, 2016), and Investment Fund institutions are not the exception (Wright, 2015). In fact, throughout its operating process, the Investment Fund industry heavily relies on documentation as an important source to facilitate information flow between several players (Salusi, 2013) (DII, 2017) (LuxembourgforFinance & Digital Lëtzebuerg, 2016). However, the ever-changing regulations and compliance controls make documentation delivery even more challenging, time-consuming and error-prone. Consequently, counting on new technology to support *compliance* in the Fund industry is very appealing. In this section, we focus on the Investment Fund documentation analysis and particularly, how text analytics serve this domain for value creation.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

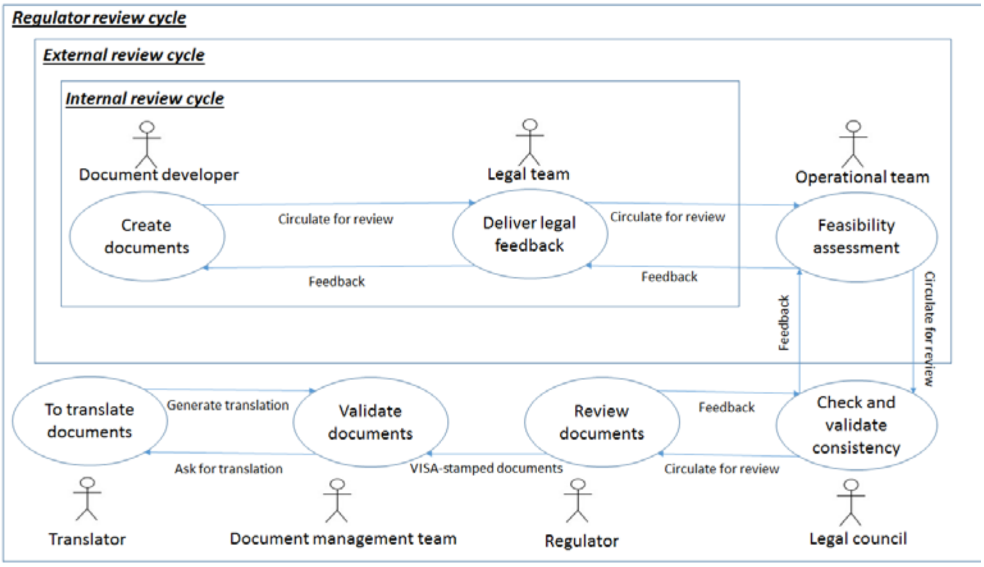

Figure 1: A simplified operating model for Fund documentation management

Figure 1 demonstrates a simplified operational model for Fund documentation. As depicted, each document follows three review cycles: Internal, external and regulatory. All involve several players, mainly, document developers, legal reviewers, external legal counsel, regulators, document managers, and translators. Throughout this process, the developer produces a first version of the fund documentation and the legal team reviews the outcome. Then, an operational team carries out a feasibility assessment analyzing the potential of success under the market conditions or any risks linked to the Investment Fund as it is described in the underlying documentation. Next, the external legal counsel reviews and performs a consistency check across documents. After validation, the last version is submitted to the regulator for a stamped approval. Finally, the internal management team generates the documents translated versions for cross-border distribution to various countries and legislations. Furthermore, during a fund's lifetime, a fund manager may need to consider several updates that must be reflected across the respective documentation and the translated versions. However, the updated and translated documents do not pass through the regulator which may, upon an investor complaint, raise legal penalties.

Apparently, the multiple cycles, the numerous documents and the various players involved in the production and maintenance of a fund's documentation make the entire process tedious and error-prone. These issues often lead to regulatory consistency breaches. For example, since excerpts of text should repeat among different documents, a manual copy-paste operation may end up to an erroneously dropped or positioned narrative. Practice shows that incomplete or missing data (e.g. the risk of the Investment Fund) is quite usual. Such flaws may imply legal violations calling for fines from several thousands of euros fixed penalty to a varying daily escalating penalty of potentially millions of euros.

In this paper, two types of Fund documents are considered: *Prospectus* and *UCITS KIID*[1]. First, a Prospectus is a long legal document that is a master-source of information referring to financial (e.g investment risk, investment objective), legal (e.g. investments

[1] www2.deloitte.com

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

restrictions), and other aspects (e.g. cutoff time) relevant to an investment fund. Second, UCITS KIID contains concentrated the significant pieces of information, according to the regulator's requirements, articulated into a plain language for informing a potential investor about a Fund. To proceed, a real-case scenario was addressed in collaboration with an SME working on Fund documentation management and compliance. The scenario aims to assist asset managers in the distribution of consistent information across documents. For confidentiality reasons, the name of the company is not mentioned in this paper, and the fictitious name FDP will be used instead.

FDP is a European company aiming at providing digital solutions into an Investment Fund's documentation issuing and maintenance value chain. The intended solutions are oriented to reduce the time spent on documentation management but also alleviate the risk of flows through automation embedded into the regulatory compliance processes.

4. SCENARIO 2: MANUFACTURING DATA ANALYSIS AND INFRASTRUCTURE

With the need to strengthen Manufacturing companies' competitive position, the Manufacturing industry has been continuously evolving to encompass digitalization competences into different LoBs (Brichni, Dupuy-Chessa, Gzara, Mandran, & Jeannet, 2015) (Kagermann, Wahlster, & Helbig, 2013). Currently, Manufacturing is driven and enabled by networking and internet. In most companies, various sensors implanted in different machines are able to communicate among them and with humans in order to gather insights that could not be discovered while being isolated. To this end, data from the entire production value chain (Rüßmann, et al., 2015), including products, customers, and order data, is networked in multiple systems to help capacity-planning, production logistics and quality control (Brichni, Dupuy-Chessa, Gzara, Mandran, & Jeannet, 2017) (KPMG AG, 2014).

In the steelmaking industry, very complex manufacturing processes are continuously executed (Jiu-sun, Xiang-guan, Chuan-hou, & Shi-hua, 2008), such as predictive control. The latter depends on several inputs, but also requires a lot of techniques and systems to manage and optimize the overall plant's efficiency and stability. In fact, the predictive control allows operators to manage certain production parameters such as oxygen enrichment, blast moisture, cold blast temperature, and others (Tunckaya & Koklukaya, 2016).

For example, as part of the predictive control and in order to seamlessly produce metal at a pre-established quality, keeping the mix close to a target temperature is mandatory. However, blast furnace temperature prediction is a nonlinear and highly complex problem to model and analyze. New forms of process control could be identified and applied in actual production by exploring and deploying suitable analytics solutions.

In this context, in order to identify and understand the challenges encounter a typical Manufacturing company with regards to data analytics and infrastructure, in the following, a real case scenario addressing the predictive maintenance is considered. Specifically, a real case scenario with a leading provider in the steelmaking industry was studied. The scenario explored Business Analytics techniques as well as the needed infrastructure for predictive control so that suitable actions can be quickly applied in

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

production. For confidentiality reasons, the name of the company remains concealed, and instead, a fictitious name “TPC” will be used.

Several factors led TPC to proceed with its predictive control using analytics techniques and infrastructure convenient for data ingestion, storage, processing, and visualization. First, the genres of data from operations are becoming more and more complex. In the current setup, TPC records 3000-5000 data points from its sensors every minute for a typical plant, such as temperature, pressure, heat loss, chemical analysis and so on. In addition, data describing the chemical composition of raw materials is considered, as well as the operator-set control parameters. Second, the existing storage solutions, mainly RDBMS, are inefficient to deal with such a huge volume of data from the manufacturing floor. Third, processing and exploring production data exploiting elementary analytics techniques (i.e. manual reporting, descriptive statistics) fall short to deal with a high number of complex parameters. For example, basic visualization tools that TPC uses are inadequate to visualize results in real-time while also taking into account the changes that may happen inside the plant or by operators, sometimes occurring every millisecond. Finally, data security has always been essential in Manufacturing and it will be continuously increasing alongside with new needs, such as temperature prediction, where data ought to be extracted in real time, assessed by operators, shared with data analysts and validated by business managers and others. Such a process involves several data exchanges, which requires secure transfer and integration plans. Therefore TPC embraced sophisticated analytics and infrastructure solutions to facilitate data processing for predictive control applications.

This use case demonstrates how data was collected, stored, processed and visualized for predictive control resulting in new revenue streams. The roles participating in this work have different profiles in the company and distinct skill-sets ranging from executives and researchers to engineers; each contributed to analytics, process control or business management. The findings are presented in the next sections.

5. SCENARIO 3: WEALTH MANAGEMENT ANALYSIS

The Finance industry has always been driven by data (Sanz J. , 2016). Finance executives understood the potential value of data in their decision-making processes and have been exploring and incorporating cutting-edge technologies into various finance fields (e.g. capital markets, financial services). Historically, finance organizations have broadly incorporated systems of business intelligence for obtaining insights from operational data stored in relational databases (Li, Jiang, Yang, & Cuzzocrea, 2015). These systems are rapidly becoming insufficient due to their limited analytical capacity and the lack of sufficient mechanisms for big data handling.

Recently, finance organizations have been rethinking and restructuring the internal processes and operations due to a number of different factors: Increasing competition, economic instability, pressing need for growth and revenue-making and lately, also increasing regulatory burden after the financial crisis of 2008. Particularly in Europe, the corresponding regulatory bodies introduced regulations to impose transparency on investment firm-client relationship. Investment firms should thus comply with a number of principles and should undertake a number of reasonable steps to prove that they match clients’ investments goals.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Particularly, the Markets in Financial Instruments Directive[2] (MiFID II) requires investment firms to consider the goals, the financial status and the level of sophistication and understanding of clients before providing any financial advice. In this context, investment firms measure the risk tolerance or aversion of the clients and justify the provided advice. In the MiFID II, the aforementioned aspects are comprised under the *suitability* and *appropriateness* assessment impositions.

The suitability assessment lies in three main pillars: (i) Client's sophistication, (ii) client's financial status, and (iii) investment objectives. The genre of information and data that an investment firm needs to collect and process are summarized as follows: (i) The type of services consumed, (ii) the type of transactions and frequencies, (iii) the investment instruments preferred, (iv) the ability of a client to bear losses, and (v) other demographic and personal information (e.g. profession, education, income, possession of assets, and so on). The previously-said information should be considered in the frame of the client's investment needs and time horizon.

An investment advisory firm may achieve clarity on their retail clients and be able to a thorough investment advice, under the requirements of the European Securities and Markets Authority (ESMA), by aggregating and analyzing multi-dimensional data of historical transactional and personal data of clients. The purpose is to capture the sophistication as well as the risk and investment profiles of clients. This type of analysis is feasible nowadays through the development of the analytical methods and big data technologies.

In this scenario, it was essential to process the transactional data of the bank's clients of a forty months period. We focused on clients that they were not contractually engaged with the bank to receive *formal* advice; the so-called *execution-only* mandate. The objective of this analysis was to identify the execution-only clients who invested in using the bank's trading platform and adopted the bank's investment strategy through personal channels. The previously said analysis is important under the MiFID II directive to alleviate the liabilities of the bank to provide advice that is not reported and may be inappropriate for a particular client with respect to the corresponding guidelines.

To proceed, this scenario considers the Wealth Management organization of a large financial institution in Europe. A number of analytics features were considered to better probe clients' characteristics, with a focus on the investment profile, investment activity and advice appetite. Among the introduced attributes, the asset classes and the investment products bias, the frequency of trading, and the portfolio diversification are taken into account. By applying a number of complex statistical and machine learning methods for a period of six months and after scrutinizing the extracted results and tuning the statistical parameters, it was possible to calculate the probability that a client under execution-only engagement has received advice. The analysis required the comparison with clients who are engaged under other mandates for identifying behavioral similarities. The range of techniques that were used comprised from classification to clustering and other probabilistic models on a dataset of many millions of transactions, a few thousands of clients and many dozens of different features.

In the previous sections, three scenarios were described and considered to harvest important lessons-learnt from the practice of analytics, as detailed in the next section.

[2] https://www.esma.europa.eu/policy-rules/mifid-ii-and-mifir

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

6. LESSONS-LEARNT

The scenarios described above allowed to harvest two types of lessons-learnt, technical and business as follows.

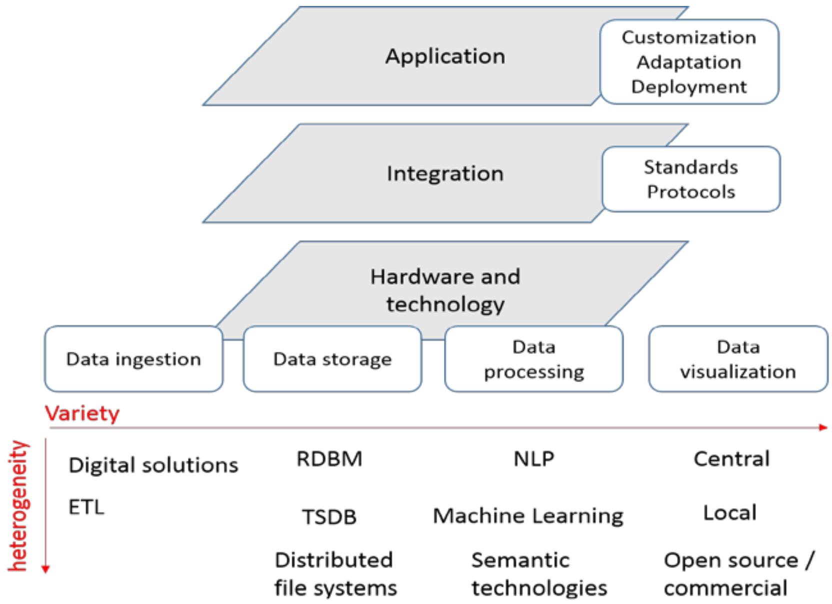

Figure 2 Business Analytics infrastructure

Based on the findings from the case scenarios, a Business Analytics infrastructure can be composed of three layers (Figure 2). First, the hardware layer deals with the most suitable capabilities in terms of techniques, methods, and tools. Second, the integration layer provides a standardized communication between technology and its application. Third, the application layer addresses the way to customize, adapt and deploy Business Analytics solutions in the application environment. Each of these layers contributes to the lessons-learnt from the practice of Business Analytics in the aforementioned domains.

Hardware and technology

(i) Varied infrastructure and heterogeneous functionalities

As depicted in Figure 2, all the considered scenarios require various infrastructure components: Data ingestion, storage, processing, and visualization. Each faces different challenges and requires specific analytics capabilities, which requires assembling a set of heterogeneous functionalities and techniques.

For example in Manufacturing, a high number of sensors' data must be stored in an efficient way to facilitate querying. Usually, companies deploy traditional RDBMS for recording high-frequency sensors data sourced from their plants and equipment and hence, have been suffering low-performance issues such as latency of reporting and processing, and lack of scalability. Indeed, because of the limitation in storage capacity, TPC was forced to periodically erase important minute-based sensor data from its databases after a few days. Similar issues occurred at its customers' end as the data from the equipment was also fed into relational databases that come with limited throughput and scalability. Since in TPC, sensors data is timestamped, monitored, down-sampled, and aggregated over time, we recommended the deployment of time series databases

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

(TSDB), which offer parallel query performance, high concurrency, high volume random reads and writes, et cetera. Similarly, TPC needs two levels of real-time visualization capabilities, one used locally and one used on its customers' premises. To this end, we evaluated several commercial and open source tools ending up deploying a hybrid one.

In the Fund scenario, as none of the documents come in a handy machine-readable format, data processing involves heterogeneous techniques. First, Natural Language Processing (NLP) aims at recognizing, and extracting human language snippets (Zeroual & Lakhouaja, 2018) (Afzal, et al., 2018) (Mishra, & Sahoo, 2016). However, problem-solving with NLP and various Machine Learning (ML) methods would require a repetitive trial-and-error process, which is costly (Roberts, et al., 2012) (Zheng, et al., 2014). Therefore, semantic technologies and a knowledge base were required to effectively capture the semantic information of the text. This is essential for deciphering ambiguous (e.g. risk) and identical finance concepts (i.e. bonds and fixed income) that a machine may fall short to grasp. As a result, a combination of techniques that lie into various domains is required in Fund documentation analysis, namely NLP, ML, and semantic technologies.

Integration

(ii) Need for integration and standardization capabilities

In order to manage interoperability, standardization is a key factor for communication among systems across industries. However, standards can be a barrier to new technologies, because industry bodies lack digital standards and norms and are too slow to develop novel ones for industry 4.0 (Obitko & Jirkovský, 2015). As a result, promoting communication among infrastructure systems and operators in the application domain through standards and protocols require an overhead of requirements and systems identification.

In TPC, new initiatives to standardize a unique specification for devices are adopted that enable security of data gathering, for example, the Object Linking and Embedding for Process Control Unified Architecture (OPC-UA) and the Classical OPC developed by OPC Foundation[3]. Both allow ensuring the communication among machines through IoT spans.

However, the overall initiative became a challenging task since not only sensors devices were considered, but also other sources, such as Manufacturing Execution Systems (MES) and Enterprise Resource Planning (ERP) systems including also data from customers, suppliers, schedules, shifts, et cetera. Consequently, even with an advanced solution for connecting devices, the configuration efforts are enormous in terms of implementation and maintenance costs as well as administration and security configuration.

This has been also part of the Fund documentation and Wealth Management analysis where regulatory compliance requires an integrated solution that fits with sets of standards, guidelines and best practices. In fact, due to the ever-evolving nature of regulations, performing and integrating compliance standards within a regulatory framework is an ongoing process. Consequently, the applicable regulations demand a continuous and effective integrating throughout the analysis process in the application environment. For instance, the set of analytics techniques such as NLP and ML need to

[3] https://opcfoundation.org/

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

be integrated and developed to tap out and maybe transform a piece of information in a form that complies with regulations to accommodate compliance controls. This is also the case of data storage where multiple policies and standards require companies to store data for a specified period of time, such as the EU General Data Protection Regulation[4] (GDPR).

As a result, these aspects indicate the way such solutions may be effectively developed and integrated into the application environment by putting standards and policies into actions.

Application

(iii) Significant amount of customization of software in analytics applications

Usually, companies rarely find infrastructure pre-packaged solutions to cover their particular needs and requirements. They might present some connectivity and configuration issues or restrictions in functionalities or other deployment limitations. In this context, analytics-enabled solutions require considerable customization to resolve compatibility issues when used and deployed in an integrated environment.

For example, in TPC, Manufacturing Business Analytics solutions are expected to behave in production as both central and local hubs for data ingestion, storage, processing, and visualization. However, the format of data when extracted and its format when processed and visualized are usually different. This is due to the fact that in many contexts, data ingestion and storage solutions require a various set of standards and interfaces such as OPC UA and OPC HDA that offer limited connectivity to external processing and visualization solutions. On the one hand, this prevents the company from taking advantage of the various functionalities that might be offered by open source solutions, for instance, and dealing with commonly used formats, such as XML, JSON, et cetera. On the other hand, this makes more difficult the deployment of a consistent and fully connected local and central infrastructure in production. Therefore, data ingestion, storage processing, and visualization systems also need to be adequately customized to cater to the needs of the whole infrastructure.

In addition, predictive maintenance in Manufacturing requires complex predictive models to assist forecasting the behavior inside the plant and guarantee runtime for time-critical processes. However, deploying these models into production and gathering results in real-time requires many customization tasks in both the model and the infrastructure. To this end, some solutions that simplify the deployment of predictive models into production exist either in commercial or open source tools, such as Predictive Model Markup Language[5] (PMML). But, even with such facilities, it is still required to deal with some specific modeling requirements that are challenging to handle.

A similar challenge is observed in Fund where the document processing service is offered to several European markets. Consequently, multiple customizations related to the language should be performed to target English, French and German speaking Fund domiciles, as main European languages. Particularly, many multilingual information extraction models were customized and implemented to match the heterogeneous language structures.

[4] https://www.eugdpr.org/
[5] http://dmg.org/pmml/v4-3/GeneralStructure.html

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Consequently, as discussed above, an efficient and effective analytics infrastructure would need to be adequately customized to cater to the business needs by means of embedding the functionalities of retrieving, storing, processing and visualizing data.

(iv) Continuous adaptation of analytics capabilities

Adaptation is an ongoing effort to improve analytics-enabled capabilities to comply with new business and technical requirements in the light of their efficiency and effectiveness.

For example, in the Investment Fund industry, managing an ever-changing regulation has been always a challenging task. Consequently, since regulatory requirements are continuously evolving, the documentation evolves accordingly. For example, starting 2019, the UCITS KIID will be replaced by PRIIPS KID[6] that, among other changes, will be longer - three-page document instead of two - including more graphics. Also, due to the globalization of Investment Fund services, new languages will be considered. Consequently, the deployed models to tackle the UCITS KIID analysis will need to be redesigned. As previously described, the effectiveness of NLP models relies heavily on the quality and the volume of the training input. Therefore, making swifts on the length, the structure, and the articulation may decrease the performance.

This is also the case in Manufacturing where analytics techniques need to be continuously adapted to new business requirements arising from either customers' needs or new manufacturing standards. For instance, it is mandatory to enhance data storage systems with self-adaptation capabilities, so that they can meet the requirements of new storage services in terms of scalability, security, and availability.

(v) Cost of deployment (local, cloud, hybrid)

Usually, one of the major business endeavors is the cost reduction. Particularly, in Business Analytics, the deployment platform can be chosen based on the speed of processing, investment cost, ease of deployment, et cetera.

Indeed, in Manufacturing, automatically (re)deploying models into production requires enormous configuration efforts that are costly in terms of implementation, time-to-deploy, and maintenance. For instance, because temperature prediction models are essentially developed in particular environments, they may be unstable for production-level deployment, they may be incompatible with existing production components and they may not be easily transferable from one software environment to another in real production lines. As discussed above, there exist some solutions that facilitate the deployment of ML models into production, such as PMML.

Generally, analytics-enabled solutions may be deployed depending on the desired way to access and share resources and to configure and manage the infrastructure.

For example, in Wealth Management, due to the sensitivity of financial data, in-house solutions still have been preferred by financial institutions to store and process large amounts of transactional data. On the contrary, in Fund documentation, most of the documents, such as KIIDs and Prospectuses, are publically accessible. Consequently, a cloud solution may simplify the storage and processing of the Fund documentation.

In Manufacturing, edge computing may be an efficient solution to perform data processing near to the data source, so that it reduces communication bandwidth and security threats. TPC, for example, expects a mass deployment of real-time dashboards

[6] Commission de Surveillance du Secteur Financier (CSSF) - http://www.cssf.lu

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

on several dozens of edge devices and nodes which are in proximity to the industrial equipment on the manufacturing floor. Thus, a robust solution to address the needs of the company must not have only the capacity to be integrated within visualization solutions but also to tap into the analytics ecosystem to retrieve data.

In all cases, local, cloud or hybrid solutions deployment may be an expensive investment that depends on the requirements of each company.

(vi) Insufficient quantity and quality of data

Cutting-edge analytical methods require data maturity to return accurate and meaningful results to substantially impact the decision-making process of an enterprise. In reality, some of the business-significant data sources are not digital (e.g. manually-generated paper documents), not adequately recorded or may lack quality (e.g. missing or erroneous data). Therefore, the retrieval, transformation, and aggregation with the purpose of applying sophisticated analytical techniques is a complex, tedious and challenging task. In fact, the Artificial Intelligent tools' effectiveness relies on the data quality. Low-quality inputs may result in poor results which add zero value to the business - the "garbage in, garbage out" principle.

For example, in the Wealth Management area, the data that reflects the Wealth Manager – Client relationship is usually poorly mastered or not in digital form. Getting a better understanding of the clients' behavior would be more effective when both the transactional and the Wealth Manager's advisory data were collectively analyzed.

Furthermore, complex predictive cognitive methods (e.g. Deep Neural Networks) require plenty of data for parameter estimation and effective generalization. In the Investment Fund documentation use-case, the "learning curve" of the multi-parameter Machine Learning and Natural Language Processing models for information extraction increases with the amount of training (Zhu, Vondrick, Ramanan, & Fowlkes, 2012). Each machine learning model is dedicated to extracting a particular type of information and its training is carried out on a specific annotation oriented to the *data point* of interest. For instance, for extracting the Fund's *asset classes* and the Fund's *cut-off* time, two different models need to be designed and trained on two different annotations with a significant number of samples each. The overhead for the manual annotations needed is cumbersome for enterprises that may want to incorporate sophisticated text analytics technologies.

Data quality may be also an issue in Manufacturing. For example, TPC still has been facing a number of challenges with respect to the in-house data, such as incomplete, noisy, or biased datasets. For instance, furnaces might not have enough historical and comparable data due to changes in raw material composition, process control methods, and aging equipment. In other cases, due to the operating conditions and very high temperature of the hot metal (>1000C), the temperature cannot be always recorded using a sensor. Consequently, the manual recording of hot metal temperature is performed only a few times per cast, which impacts data collection and may make accurate data-driven models more difficult to build.

(vii) Lack of commitment to taking risks with new solutions

The implementation of data analytics solutions based on *black-box* tools is reasonably straightforward to achieve due to the development of sophisticated tool libraries. Nevertheless, the return-on-investment is a function of the effectiveness of the models which may reach an adequate level over time under the tuning of the provided outcomes against the actual events. This is a long-lasting effort that requires a commitment to

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

continue with any proposed solution. Instead, business personnel may lose interest and consequently, the solution development life-cycle will not be long-lived. In many cases, companies cannot even support the time-investment needed for the efficiency improvement or may find it too disruptive for their business.

For example, in the Investment Fund documentation scenario, the provided solution comprised a large number of integrated Natural Language Processing and Machine Learning models to tackle the extraction of different types of information. The improvement of the models required a lot of feedback for performance improvement from clients and the company's personnel after the solution was put into place. The maintenance of the entire system and frequent communication needed with technical teams added burdens to the company for a period of one year.

Although the provided solution in the Wealth Management scenario was assessed and approved by the bank under a number of exhaustive controls that exhibited an outstanding performance, the deployment of the provided model was left pending. The reason resides in the reluctance of some organizations to replace revenue-making processes and related systems by adopting new modern solutions. Delays in incorporating new state-of-the-art systems entail some risks, such as replacing ongoing operations and tools. Moreover, the understanding of *black-boxes* is important for reporting to regulators and for justifying machine-based decisions to end-users.

(viii) Lack of compelling business cases and strategic plans

Business Analytics use-cases have been already in place in the majority of organizations and companies. Executives recognize the need for data-driven decision making to support growth and cost reduction in their business. Unfortunately, analytics solutions may not be well-designed to be incorporated into LoBs. Often, these interventions stay at the level of an exploration or as internal research. There is not a strategic framework for driving business cases and there is no clarity on how Business Analytics may be used or not used without incurring upfront investments and carrying out some extensive experimentation.

For example, in the Investment Fund documentation scenario, a crisp business plan on how a solution could support the business through a clear return-on-investment was missing. Therefore, an assessment of the complexity of the many dozens of datapoints of interest requested by the business was performed with a return-on-investment to be met within a reasonable time-horizon. A viable business-model was selected according to the technical specifications and complexity after a sound and intense investigation of the market needs and capacity of FDP.

On the other hand, the developed solution in the Wealth Management scenario was not part of a strategic plan of the bank. This is concluded from the fact that, although the artificial intelligence system was able to process millions of data entries and produce results approved for their correctness, the deployment of the solution is still under discussion. This use-case was mainly an experimentation scenario in the bank carried out to assess what modern technologies may bring to their business but without having clarity on how to use and incorporate any outcomes into their operations, legal framework and requirements of reporting.

7. CONCLUSION AND PERSPECTIVES

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

In this paper, real case-scenarios from the Investment Fund industry, Manufacturing and Wealth Management were presented with the goal of understanding the actual vicissitudes of Business Analytics encountered in industrial applications. Different objectives and challenges have been addressed in connection to Business Analytics practice. As a result, several lessons-learnt were identified from business and technical points of view, such as the need for a varied infrastructure and heterogeneous functionality, the changing nature of each individual domain, the heterogeneous structure of data and its semantics, among other factors.

Despite the focus given to the case-driven approach for conducting this research, some limitations stem from the need not to disseminate confidential information. Confidentiality is a significant barrier to fully benefit from experiences that different companies and other enterprises may harvest. Some useful information may not always be shared or detailed, as it is also the case of those processes that enterprises have been following in the description of the cases, as well as related data.

An important contribution of this work resides in the general nature of the findings reported here, thus having potential application to other LoBs and industry segments. Therefore, the outcomes from this are definitely adaptable to other companies or organizations with similar contexts.

In terms of future perspectives, the aim will be to investigate how to deal with these challenges in terms of the human, technological and organizational capabilities. To this end, the short-term goal will be to provide a general approach for assisting those Business Analytics projects addressed in this paper to effectively deal with the most common challenges. In the long term, the objective will be to expand the approach to other domains beyond the present three cases where the application of Business Analytics is still complex and challenging.

REFERENCES

[1] Afzal, N., Mallipeddi, V. P., Sohn, S., Liu, H., Chaudhry, R., Scott, C. G., . . . Arruda-Olson, A. M. (2018). Natural language processing of clinical notes for identification of critical limb ischemia. *International Journal of Medical Informatics, 111*, 83-89.

[2] Ang, L.-M., & Seng, K. P. (2016). Big Sensor Data Applications in Urban Environments. *Big Data Research*, 1-12.

[3] Bakharia, A., Kitto, K., Pardo, A., Gasevic, D., & Dawson, S. (2016). Recipe for Success—Lessons Learnt from Using xAPI within the Connected Learning Analytics Toolkit. *6th International Learning Analytics and Knowledge (LAK) Conference*, (pp. 378-382).

[4] Brichni, M., Dupuy-Chessa, S., Gzara, L., Mandran, N., & Jeannet, C. (2015). Business intelligence for Business Intelligence: A case study at STMicroelectronics. *IEEE Ninth International Conference on Research Challenges in Information Science (RCIS)*, (pp. 239-249).

[5] Brichni, M., Dupuy-Chessa, S., Gzara, L., Mandran, N., & Jeannet, C. (2017). BI4BI: A continuous evaluation system for Business Intelligence systems. *Expert Systems With Applications, 76*, 97-112.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[6] Brichni, M., Kampas, D., & Sanz, J. (2018). Some Challenges and Lessons-Learnt from the Practice of Business Analytics. *Society of Interdisciplinary Business Research.* Hong Kong.

[7] Brichni, M., Kampas, D., & Sanz, J. (2018). Some Challenges and Lessons-Learnt from the Practice of Business Analytics. *Society of Interdisciplinary Business Research.*

[8] Dataiku. (2016). *Data Science for Banking & Insurance.* New York City: Dataiku.

[9] Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, pp. 1-13.

[10] Davenport, T. H., & Harris, J. G. (2017). *Competing on Analytics: The New Science of Winning.* Boston: Harvard Business Review Press.

[11] DII, D. I. (2017). *Who Are the Actors in an Investment Chain?* Retrieved April 1, 2018, from https://www.followingthemoney.org/chapter/what-is-an-investment-chain/who-are-the-actors-in-an-investment-chain/

[12] Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Making Sense of Video Analytics: Lessons Learned from Clickstream Interactions, Attitudes, and Learning Outcome in a Video-Assisted Course. *International Review of Research in Open and Distributed Learning, 16*(1), 260-283.

[13] Hartzband, D., & Jacobs, F. (2016). Deployment of Analytics into the Healthcare Safety Net: Lessons Learned. *Online Journal of Public Health Informatics, 5*(3), 1-13.

[14] Hwang, J., Huang, Y.-W., Vukovic, M., & Jermyn, J. (2015). Cloud Transformation Analytics Services A Case Study of Cloud Fitness Validation for Server Migration. *IEEE International Conference on Services Computing*, (pp. 387-394).

[15] IBM. (2018). 3 Big Advancements in Unstructured Data Analytics. IBM Journal.

[16] Jiu-sun, Z., Xiang-guan, L., Chuan-hou, G., & Shi-hua, L. (2008). Subspace method for identification and control of blast furnace ironmaking process. *American Control Conference*, (pp. 2481-2485).

[17] Johnson, B. (2015). *Designing and Deploying Data and Analytics-Enabled Business Capabilities.* Department of Management Science & Engineering. Executive summary - Stanford University.

[18] Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0.* Final report of the Industrie 4.0 Working Group.

[19] Kennedy, P. J. (2014). Redesign of Data Analytics Major: Challenges and Lessons Learned. *5th World Conference on Educational Sciences, 116,* pp. 1373-1377.

[20] Kokin, S. a. (2013). The technological business intelligence capabilities and business intelligence system flexibility- the underlying relationship. *Digital Information Management (ICDIM), 2013 Eighth International Conference*, (pp. 114-118).

[21] KPMG AG, W. (2014). *Building the Factory of the Future.* KPMG.

[22] Lécué, F., Tallevi-Diotallevi, S., Hayes, J., Tucker, R., & Bicer, V. (2014). Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City. *Web Semantics: Science, Services and Agents on the World Wide Web, 27,* 26-33.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[23] Li, K.-C., Jiang, H., Yang, L. T., & Cuzzocrea, A. (2015). *Big Data: Algorithms, Analytics, and Applications* (1 ed., Vol. Chapter 17). Chapman and Hall/CRC.

[24] Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems, 24*, 149-157.

[25] Lonn, S., Aguilar, S., & Teasley, S. D. (2013). Issues, Challenges, and Lessons Learned When Scaling up a Learning Analytics Intervention. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, (pp. 235-239).

[26] LuxembourgforFinance, & Digital Lëtzebuerg. (2016). *Digital Luxembourg FinTech Working Group.* Luxembourg: LuxembourgforFinance and Digital Lëtzebuerg.

[27] Mishra,, B. K., & Sahoo, R. (2016). A Hybrid Knowledge Mining Approach to Develop a System Framework for Odia Language Text Processing. *International Conference on Processing of Materials, Minerals and Energy, 5*, pp. 1335–1340.

[28] Obitko, M., & Jirkovský, V. (2015). Big Data Semantics in Industry 4.0. *7th International Conference HoloMAS, 8062*, pp. 217–229.

[29] Official Journal of the European Union. (2014). *Markets in Financial Instruments (MiFID II) - Directive 2014/65/EU.* Brussels.

[30] Parilla, E. S. (2019). Exploratory Study on Lean Manufacturing Tools and Techniques in Selected Manufacturing Firms in Metro. *Review of Integrative Business and Economics Research, 8(s1)*, pp. 15-24.

[31] Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J. a., & Harnisch, M. (2015). *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries.* The Boston Consulting Group.

[32] Roberts, K., Rink, B., Harabagiu, S. M., Scheuermann, R. H., Toomay, S., Browning, T., . . . Peshock, R. (2012). A machine learning approach for identifying anatomical locations of actionable findings in radiology reports. *AMIA Annual Symposium Proceedings, 2012*, pp. 779–788.

[33] Salusi, M. (2013). *Luxembourg: The Global Fund Centre.* Luxembourg: Technical report - ALFI Association of the Luxembourg Fund Industry.

[34] Sanz, J. (2016). The Transformation of Financial Competences in Enterprises — The case of SAP in Asia-Pacific-Japan. In *Book on Finance Transformation through Digital Acceleration - Case-studies worldwide.*

[35] Sanz, J. (to appear 2018). Business Analytics — Evolution and New Opportunities in Enterprises. *World Scientific Publishing.*

[36] Sanz, J., & Chow, J. (2016). Big Asymmetries in the Era of Big Data - Where do the United States, the European Union and Asia stand? *Risk Dialogue Series, Swiss Re Center for Global Dialog.*

[37] Sanz, J., & Pang Yan, J. (2016). *Big Data and Business Analytics: Accelerating Digital Transformation in Enterprises and Industries [Powerpoint slides].* Luxembourg: European Data Science Conference (EDSC) Lecture.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[38] Tunckaya, Y., & Koklukaya, E. (2016). Comparative Performance Evaluation of Blast Furnace Flame Temperature Prediction using Artificial Intelligence and Statistical Methods. *Turkish Journal of Electrical Engineering and Computer Sciences, 24*, 1163-1175.

[39] Wright, J. (2015). Recognos Reference Data Center to Transform Critical Mutual Fund Data into Actionable Insight. *Recognos Financial Launches Utility*, 160-167.

[40] Zeroual, I., & Lakhouaja, A. (2018). Data science in light of natural language processing : An overview. *Procedia Computer Science*, *127*, 82-91.

[41] Zheng, C., Rashid, N., Wu, Y.-L., Koblick, R. a., Levy, G. D., & Cheetham, T. C. (2014). Using natural language processing and machine learning to identify gout flares from electronic clinical notes. *Arthritis Care AND Research*, *66*, pp. 1740–1748.

[42] Zhu, X., Vondrick, C., Ramanan, D., & Fowlkes, C. C. (2012). Do We Need More Training Data or Better Models for Object Detection? *BMVC*.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)