Unveiling the Silver Lining of Dark Data for Organizations

Shivangi Priya[1]
Research Assistant,
School of Commerce and management,
Lingaya's Vidyapeeth

Samriti Mahajan[2]
Associate Professor,
School of Commerce and Management,
Lingaya's Vidyapeeth

ABSTRACT

Purpose: Dark data is a type of data that can be extracted from various types of databases and computer network operations. These can be taken out but are not used in any way by an organization to gain profit or benefit from them. We are surrounded by a flood of data. IDC estimates that between 2013 and 2020, the quantity of new data generated annually would increase tenfold, from 4.4 trillion gigabytes to 44 trillion gigabytes. We can succeed only if we can navigate this data tsunami. Companies that succeed at making data-driven decisions are, on average, 5% more productive and 6% more profitable than their rivals, claim MIT researchers. According to a study by IDC, users of Dark data and analytics who employ a variety of data sources, analytical tools, and metrics are five times more likely to surpass project goals than those who don't.

As businesses and management face increasing competition, the use of large or big data is increasing as it frequently contains information that cannot be found in any other format, dark data is valuable. As a result, businesses continue to invest in the costs associated with gathering and retaining dark data in the hopes of one day using it to access important information.

In contrast, many organizations are unaware that data collection is taking place in their organization. It's data that is always adjoining, omnishambles and unrevealed. Nevertheless, a big chunk of the collected data is still not properly examined. But with the increase in competition, there are certain entities that have perceived the value of such data. They have started working towards inventing new methodologies, and procedures and developing the operating system so that the data can be taken out and used in the organization's favor, which benefits generating revenue and adding more and more customers.

The objective of this study is to increase knowledge and consciousness among businesses and the general public about the concept of "dark data" and how it can be leveraged for future revenue and profits. The study will also propose methods for organizations to use dark data cost-effectively, and how it can be a game-changer for them. Furthermore, this study will explore the

potential possibilities that arise once the data is extracted and applied, leading to rapid growth for organizations in the future and along with that how to address the challenges and risks associated with working on dark data.

Keywords: Dark data, Network operations, consciousness, Disorganized data , Security violation, Cost-efficient, leveraged

INTRODUCTION

Gartner Research Inc. developed the concept of Dark Data, also known as Big Data, and described everything as "Big data" is a high-volume, velocity, or variety of knowledge assets that need expense, innovative aspects of the data processor for understanding and managing and decision - making processes.

In 2001, Gartner analyst Doug Laney struck the term "three Vs. back" in 2001. The mass engrossing of the three Vs back is variety : businesses are extracting incredible insights from text, locations, or log files. Elevator logs assist anticipate vacant real estate, shoplifters tweet about stolen products right next to the business, and emails indicate effective project communication habits. Gartner coined the term "Dark data" to describe the underutilized wealth of information that predominantly resides within organizations.

Big data research across disciplines is primarily concerned with the analysis and use of actual data. Social networking sites, businesses, and governments are all interested in understanding the implications of big data in social behavior, corporate operations, and the role of government in citizens' lives (Mayer-Schonberger and Cukier 2013; Glass and Callahan 2014). Big data in scientific and technical research makes vast volumes of data available to researchers, allowing them to pursue new pathways of exploration and discovery. Excellent work on the benefits and implications of big data in scientific research is evident in neurology (Choudhury et al. 2014), biology (Leonelli 2013,2014), & astronomy (Edwards and Gaber 2014), to name a few scientific disciplines.

The universe of big data, of which we have only discussed a few examples, is primarily interested in "physical data," that is, data that is readily available to the user (e.g., researchers, public servants, businessmen, and businesswomen).

In contrast to "physical data," "dark data" is data that is kept on servers and is available for use, but users are ignorant of its presence. Although the specialist literature has been intimately concerned with analyzing "tangible data," it has mostly ignored "dark data." This article focuses on the inverse side of data, the one we term dark since it contains potentially rich sources of trustworthy information but is forgotten on storage servers where it cannot be used. The essential goal of metadata, then, is to structure, inform, and recognize data by using pertinent details about them. When such administration function then criteria are not being accompanied as is admittedly the situation for many HPC facilities, data becomes black, opaque, and undetected by

researchers. This, we believe, is a critical issue for data management in HPC facilities that must be addressed. The availability of dark data is especially concerning when these facilities are linked with public institutions rather than the commercial sector.

Our example is the High-Performance Computing Center Stuttgart (HLRS), which is situated in the center of the University of Stuttgart and is one of the leading high-performance computing facilities in Germany and worldwide, and also one of the biggest of its type in Germany. The information obtained by an HPC facility, like the HLRS, is important in scientific and engineering processes, thus the cost of production, as well as the social and environmental advantages, are notably different from other kinds of big data. Consider the expenses of teaching an engineer, then add your costs of creating scientific data, and then compare these fees towards the expenditures of gathering data from firms like Facebook or Google. In the latter situation, rather than real investigators creating data, robots are most often engaged in harvesting data from either the user's login to their own accounts.

The Rise in Data-Driven Marketing -

The rising usage and attractiveness of data-driven advertising were expected to boost the growth of both the darker industries at that time over the projected period. Data-driven branding has attracted attention as technology has increased. Individual-targeted and improved marketing tactics may be created by exploiting and analyzing client data, and therefore the data-driven industry has grown. Furthermore, since rivalry in data-driven advertising has escalated over the years, employers are constantly digging into dark data to discover new and untapped insights about clients in order to remain ahead of the competition.

With the emergence of new technologies like machine learning and big data analytics, leveraging dark machine learning for marketing is predicted to be one of the main differentiators in the forthcoming brand marketing competition.

As a result, increased competition in data-driven marketing is predicted to motivate marketers to employ sophisticated technologies such as dark analytics, which is expected to fuel the growth of the data analytics market over the forecast period.

Why Focus on Data Security Increased?

The rising focus of enterprises on information security has resulted in higher demand for dark data analytics, which would be expected to boost the growth of the dark analytics market over the forecast period. According to the Govt. Of India data, the overall number of cybercrime incidents in India nearly tripled between 2019 and 2020, hitting 1.16 billion in 2020.

A data breach including patients' private details was discovered in Mississippi's Coastal Family Medical Center in May 2021. (CFHC). Identities, addresses, security numbers, health insurance data, and health care and treatment information were all compromised. Such data breach

instances have been on the rise in recent years, raising data security worries among businesses.

When unexplored dark data is detected and managed, it leads to an increase in the group's security, which safeguards undiscovered sensitive data from hackers, and is one of the major factors anticipated to propel the use of deep analysis in the next few years.

The cost of data storage taking only the storage expenses into account: According to IDC's economic growth forecast for information, the cost of storing one terabyte (1TB) of data is approximately £200 per month. If this data remains unmanaged, the forecast suggests it could potentially grow to 10TB by 2025, resulting in a monthly storage cost of around £2000.Approximately 60% of this data may be of little or no use to the organization. Doing nothing now essentially commits a corporation to a future cost of £1,200 per month without no return. These costs to the organization from storage alone might be dominated, if the data is later put through a data breach, with the emerging fines and renown harm to the organization likely to have a negative impact on the business. This example mentioned is mainly for a company with up to 1TB of data today; but what about companies with tens or even hundreds of Terabytes of data? Compliance is an important aspect of the discussion. McKinsey & Company, a management consultant, found an issue in 2018 and alluded to such a potential solution.

Why Is There Dark Data ?

The causes of dark data are multifactorial and gauge the diapason of exploration and reporting exertion. At one minimum, in a largely competitive exploration Terrain, there exists a perception amongst experimenters that delineation attention to sweats that have been unprofitable in demonstrating an anticipated outgrowth can work against their career pretensions and chances of unborn backing. In a data-driven world, changing this perception goes to the heart of Exploration culture, and involves feting and celebrating those who have pursued well-planned and designed avenues of exploration, indeed if those results aren't “ positive". ”

At the other extreme, competition for space in top-tier peer-reviewed journals has meant that null thesis Calligraphies have faced a high bar for acceptance and contend against papers with positive results, which could be seen to have advanced marketable value in terms of attracting citation, subscriptions, and reprints. This results in repeated Guests of handwriting rejection. numerous journal editors believe, despite substantiation to the negative, that null thesis papers are less likely to be cited in unborn papers, with citation being used as a crude index for exploration of Applicability and impact. In addition, the “novelty” of a study can be a consideration for journals.

That is, in making publication opinions, journals frequently assess whether the exploration is “new, true and does anyone watch?” Negative, inconclusive, and confirmatory results may not meet journals’ prospects for new and unique exploration. Since the analysis, jotting, and handwriting drafting processes are time-consuming for time-poor experimenters, numerous choose to concentrate their sweat on exploration that they perceive has a lesser chance of

publication success. Changing this perception requires close commerce with major journals, their editorial brigades, and the establishment of devoted space for well-designed studies that affect negative and inconclusive issues.

Objective of Study

1. To Investigate unstructured data in FMCG companies which is crucial for future business success, as data holds the key to unlocking new opportunities and achieving growth

2.To investigate the feasibility of identifying a body or governmental organization that can offer cost-free support to FMCG companies for accessing Dark Data.

3.To investigate various types and sources of dark data and highlight ways in which organizations can leverage this untapped resource to gain valuable insights and improve decision-making.

4. To address the challenges and risks associated with working on dark data and propose strategies for mitigating these risks.

Need Of the study

The purpose of this study is to raise public knowledge of the idea of dark data.This study focuses on the Unorganized data generated accidentally in company and by users. Dark data, as by the Gartner, is "the 'information asset' that businesses analyze, acquire and engage during ordinary commercial activities but infrequently recruit for any grounds."

There are several specialized use cases. One of the most significant will be the creation and development of new and more productive corporate business strategies. According to IBM, 80% of all the data is 'dark ', whereas" 35% of which is relevant, and just 0.5% gets evaluated! Because data also isn't actively saved, a vast amount of information goes undiscovered or unused. Processing it becomes burdensome since data is dispersed, unstructured, and incomplete. Discovering ideas from dark data is like to discovering a "needle in a haystack," but the issue is,

Can this needle be neglected? Especially if it incorporates intelligent insights into company strategy via data threads? Some of the reasons FMCG companies fail to capitalize on this plethora of dark data include:

1. Preferences: Teams in sales and marketing have a myopic view of data which only provides instant satisfaction, such as customer transactions, rather than data related to what a consumer attempted but did not convert into a buy and why.

2. Separation between departments: Often, one department has data that another can proactively use, but owing to a lack of collaboration, this wealth of data remains underused.

3. Constraints imposed by technology: Organizations that lack robust platforms for data collecting and analysis miss out on possibilities to gain insights.

Review Of Literature

In a study titled "Extracting Databases from Dark Data with DeepDive," Stanford university researchers discuss how the system DeepDive can extract structured databases from dark data more correctly and rapidly than humans. DeepDive is distinguished by its ability to generate extraordinarily realistic databases with minimal human engineering effort. DeepDive seems to have been able to extract information with a precision that matches or exceeds that of human experts in a surprising number of applications. DeepDive's excellent quality is made possible by a one-of-a-kind architecture based on probabilistic reasoning and a specific engineering development cycle; it is supported by various technical advancements for fast statistics training and sampling. It was used to extract high-quality data in a variety of fields, including genetics, insurance, Web newspaper ads, materials science, paleontology, and others.

Heidorn starts with the topic of dark data and showcases how small-scale astronomical research initiatives produce untapped data as a result of limited funding allocated towards managing research data responsibilities (Heidorn et al. 2018). To be sure, Heidorn et al. admit in their research that dark data may also be available in digital format, bringing it closer to human interests. Heidorn categorizes scientific initiatives on a spectrum ranging from substantial to smaller-scale endeavors, with the latter involving only a handful of researchers, while the former may encompass dozens or even hundreds of individual (Heidorn 2008, 281).The long tail of science varies from the head of data in that "data is more difficult to come by and far less commonly reused or retained" in the long tail of research (Heidorn 2008, 281). Instead, cranium data is stated to be uniform and well-maintained, with central curation chores in place and unrestricted access to all approved researchers.

CommVault produced a report titled "Turning Dark Data into Smart Data," which was informed by Gartner Research Inc. It discusses the CommVaultSimpana Mail and File Analytics, which assist in having a control on data growth- By assisting organizations in cleaning up old and current knowledge by shifting it to reduced-cost storage and deleting something that can only be erased. Keep only what is necessary for business, compliance, or proof, reducing storage expenses by 70%. Assist with compliance and eDiscovery: To decrease risk, organize and impose retention and justified erasure. Give you command over access privileges and security procedures. Simpana may be used to enforce knowledge governance principles and support a well-defined data strategy. Apply, audit, and use data categories to improve unstructured data insight, management, and security.

The "Big Data" special issue of Nature (2008) gives an excellent summary of current triumphs and problems in managing enormous data collections. However, most scientists labor on the right side of the graph on smaller research projects that typically generate modest data sets.

Because of the length of the tail, while the data volumes are minor when seen individually, they constitute a considerable amount of a country's scientific output when evaluated collectively. In truth, the phrase "big science" is somewhat deceptive. Many of the smaller research initiatives at the tail are really conceptually interconnected endeavors that are funded through a dispersed financing mechanism.

Another meaning of black data that has previously been used is unreported data from "failed" research (Goetz, 2007). In this context, failed alludes to the fact that only good outcomes are often publicized, rather than faulty research. Investigations that properly indicate that the treatment condition has no impact are genuine discoveries, but they are less likely to be published. As a result, the data gets black, and subsequent meta-analyses of the literature present a biased image of the real scientific results.

Another research was done by 1Njeru Mwiti Kevin,Felister Munyi Wanyaga, David Kibaara, Wilkister Atieno Dinda, James Kariuki Ngatia; in April 2016 he aim of this research was to provide insights to organizations on how business analytic tools and software can be applied in lighting up previously unknown or ignored data. This is done through an in-depth analysis of secondary data and practitioner reports to provide an understanding of the various concepts and tools essential in identifying meaningful patterns and trends into an organization’s data.

Researcher Disha Chhabra Puri in June 2021 concluded the study by saying ,success of many businesses depends on their ability to create their customers and retain their Consumers.Her study identifies the level of influence of various factors on the purchase of FMCG products to the respondents of this study. The branding of FMCG has become an integral part of the livelihood of consumers. Consumers are literally confronted with hundreds of brands on a daily basis.

Research Methodology

This study is done with secondary data sources such as websites, published papers, journals, magazines, and books, among other sources, to gather information.

METHODOLOGY FOR COLLECTING AND USING DARK DATA

Data is critical for any organization since it contains the key to successfully managing the firm, attracting new consumers, and increasing growth. Which is why big data has become a huge business. Dark data is not only a subset of big data. It is the largest piece of the pie and has enormous promise for those who can manage it . However, the most important thing to understand about dark data is that it does not have to remain dark. Whenever dark data is leveraged to acquire ideas, it becomes practical and no more remains dark.

In many circumstances, companies are simply unaware of the presence of dark data. So, to begin, hence a need to enhance consciousness of the presence and potential that may be derived from dark data. Following that, the architecture that will allow dark predictive modeling must be

implemented. The recommended method is to build a Data Lake architecture, which will transport terabytes of data from numerous places. This new storage will consolidate all data into a single interconnected system, making it easy to retrieve and ensure that it is never ignored anymore.

1. Give a plan with no obvious costs in which we are compensated by the savings we achieve for our visitors - Obtaining executive access to every aspect of the warehouse, including all servers, hard drives, and other installed equipment

2. Search for data-nSearch and identify all available data sources. Look at the operations, bias, people, and processes.

3. Roster data - dissect and classify all data that is used by linked data sources, including the data stored in relational databases, logs, textbook data, multimedia data, IoT aqueducts, IoT metadata, auditing data, and any other data that are stored.

4. Sequestration and security- All legitimacy concerns must be linked in this step, and security and sequestration concerns must be evaluated for each dataset.

5. Calculate the cost- Based on the requirements of the business, decide which questions must be addressed first. Look for datasets that will let you answer these questions

6. Determine the value- Grounded on the business needs determine which questions are the most important to be answered first. Find datasets that will help you answer these questions.

7. Move the data - Store all of the utmost data in the centralized Data Lake.

8. Expand the data- In this step, the thing is to find if there is fresh important data that's tasted or collected but not stored. examples include some detector data, intermediate data, fresh more detailed log data, or data that's present but isn't digitized. This conduct will bear fresh trouble, so some estimate of the value of this data related to the price of getting it will be demanded.

9. Link the data together- The data that was gathered came from many processes and sources and is typically not connected.We must keep in mind that information is not limited to data but also to connections. Because it connects two or more distinct business corridors, this information can frequently be crucial for business processes and models.

10. Link to external data- Link the data with external data sources like rainfall conditions, geolocations, stock exchange, news, and large public and open data sets like DBpedia or Wikipedia**.**

Analysis while embarking on their data-driven journey for insights, retail and FMCG firms should consider the following steps:

1. Locate sources of data: The core of dark analytics is identifying data sources and developing a systematic digitization strategy for this raw data for better analysis in the present and future. Identifying data sources is crucial as businesses typically possess substantial amounts of data in various formats, including organized, semi-structured, or unorganized forms.

2. Set priorities for data sources and sets: When gathering and analyzing data, it may seem like a good idea to gather as much information as possible, but this can be expensive and time-consuming. Therefore, it's important to determine which data sources and sets are most important and prioritize them accordingly.

3. Prepare for fresh data: Dark data analytics is a continuous process that generates data indefinitely. As a result, it is critical to design a strategy for data collecting and analysis that leverages digitalization, operational alignment, and appropriate platforms.

4. Being accountable: It is critical to be accountable for data, both in terms of integrity to preserve quality and security to avoid compromising sensitive information.

5. Check with external sources: External data can be quite useful in decision-making. It not only introduces new reference points, but it also highlights disparities between the past and present.

6. Outcome strategy: The value of data is determined by what can be accomplished with it.

The technique of analyzing databases to obtain useful insights is known as data analysis. It provides answers to a lot of things like patterns, answers of the questions and helps in making judgments which are based on certain authentication.The FMCG business jumped on the bandwagon in order to simplify its production, distribution, and guest service operations.

The FMCG sector, also known as the Consumer Packaged Goods (CPG) industry, is the world's largest industry. Prior to the Covid-19 outbreak, the MENA (Middle East and North Africa) area had significant demands for consumer products.

Why Is Data Analytics Necessary in FMCG?

Whether it is stocks, supply - chain management, or client experience, data analytics could assist the FMCG business in making the correct decisions by using predictive intelligence in real-time. We'll concentrate completely on how big data analytics firms help businesses cope with customers' requirements and comprehend the target audience.

1. Evaluating Potential Customers

A competitive business exerts undeniable demand. With startups, goods, and organizations entering the market on a daily basis, it's difficult to maintain loyal clients in the FMCG industry.

Data analytics may assist you in better understanding of the consumers who are your genuine audience.Who are your supreme consumers ? What data an organization will require about

them?Data analytics can provide answers to such problems. It aids in the identification of potential clients for a long-term engagement with the company.

2. Improved Positioning

In today's business landscape, marketing and advertising play a crucial role in attracting customers and showcasing brand offerings. To achieve this, companies need to identify the most effective marketing and advertising methods that will yield the best results. For instance, creating a company website and optimizing it for search engines is essential to attract website visitors. Similarly, TV commercials are effective for companies whose target audience spends more time watching television. With the help of data analytics, businesses can gain crucial insights to determine the most suitable advertising schedules for their projects

3. Forecast Requirements Of the customers

Identifying their requirements and offering the products/services is one technique to create consumer loyalty. If a consumer wants to go digital and make online purchases, an eCommerce website will assist in keeping the customer rather than losing them to a rival that provides this service.FMGC big data utilizes data from numerous sources and analyzes it in real-time to forecast what customers desire using predictive analytics.

4. Improvements in Planning

Businesses can use data analytics to accurately track market trends and capitalize on opportunities before their competitors. By analyzing consumer behavior and preferences, brands can identify what products are in demand and make necessary modifications to increase sales. Additionally, utilizing data-driven insights can aid in the development of innovative products that can be patented, further enhancing the brand's reputation. Creating an appealing shopping environment can also increase customer engagement and spending.

What is the significance of Dark Data in marketing and sales?

Retailers are emphasizing the collection and integration of customer data at every point of interaction. While some firms have mountains of data saved in their data warehouses, masses of data are still lost between processes and consumer interactions or touch-points. Dark Data sources such as network transactions or dispersed databases can be extremely valuable to the marketing/sales department.

The widespread presence of Dark Data has significant implications for various aspects of marketing. Those who are capable of gathering and analyzing this data before others will have a crucial advantage in terms of growth and innovation. Compared to traditional firms, marketers who specialize in data analysis view marketing through a unique lens. In today's competitive market, machine learning, big data analytics, and cloud access have become essential tools for success. Dark Data analysis allows us to gain a better understanding of our customers, monitor

our service and customer experience, and provide the hyper-personalization that today's consumers demand. Furthermore, it can help us manage our brand reputation, measure the impact of our communication and marketing efforts, and engage directly with current or potential customers through precision targeting.

In an increasingly data-driven world, developing a solid data strategy may help your organization achieve a competitive advantage. Furthermore, using Dark Data allows you to gain insights into client journeys, recognise consumer unhappiness sooner, and handle customer problems more quickly.

Challenges and risks associated with working with dark data

1.Lack of source

When data lacks provenance, it is difficult to establish its credibility and trustworthiness, making it unusable for analysis. Data scientists depend on data sources that have a clear origin and history to ensure the accuracy and reproducibility of their analyses. The raw, unstructured form of data is essential for establishing its provenance, but it may not always be accessible, leading to the creation of dark data. However, as discussed in Part 2, data lakes can help preserve this unstructured data and facilitate the curation of its provenance, making it more usable for analysis.

2.Deprived paperwork

When data sources lack sufficient and accurate metadata, the data becomes difficult to access and analyze through queries, making it a type of dark data. Inaccurate or low-quality metadata can also render good data inaccessible through metadata searches. Furthermore, inconsistent metadata can lead to the fragmentation of a category, causing variations in the label metadata and making it harder to analyze the data effectively. As such, proper and consistent metadata management is essential for maintaining the accessibility and usability of data.

The potential dangers associated with dark data are as follows :-

1.Quality of data

Dark data has a significant impact on the quality of data used for analysis, hindering the ability to extract valuable insights. The difficulty in accessing and locating vital information, as well as verifying its origin and accuracy, creates obstacles in making informed, data-driven decisions. This impact on data quality arises from several factors —

a) Availability of data- Access to crucial information that might enhance evaluation is lost when data that is chaotic or in another media type, such as photos, video,or audio cannot be retrieved.

b) Precise data - The accuracy of the input data ascertains how accurate the data analysis will be. Retrieval of information with high subjective value follows accurate analysis. Dark data therefore significantly affects the accuracy of the information retrieved and the caliber of the information created by that analysis.

c) Data Traceability - If the origin of data cannot be traced, it may not be included in analysis, resulting in lower data quality. This can ultimately lead to inaccurate decision-making based on the data.

Data security

Sensitive information is typically present in collected data. Proprietary information, trade secrets, and customer and employee personal data, including financial and health records, are all considered critical information. Organizations are exposed to data security measures when they are not aware that the required data is present in their data warehouse types. Hackers who are the first to find this private information are frequently the cause of data security breaches. Significant accountability and corrective work are needed for this.

Conclusions and recommendations -

One of the most fundamental characteristics of major corporations in today's rapidly changing information-driven world is how well they understand their market, consumers, and rivals. Organizations are accumulating increasingly large volumes of data - 'big data' - as part of the 'digital age' revolution. Therefore, Dark Data will have a good impression in the future when keeping data and milking it to its exceptional potential will acquire significance among companies.There are already signals of businesses redesigning to the transfer and legislation and jurisdiction of conserving data are also being adopted.Genuine evaluation of Dark data would become very essential all over the world and we need software that makes searching and retrieving data from such archives simple. We need to educate a new generation of academic output curators who are trained in suitable computer technologies and have a passion for science and scientific sociology. Most importantly, we require new students' learning and rewards to provide the next generation of scientists with the information they require to make educated decisions about the broader use of their data and the broader effect of their research.

Acknowledgments

I would like to thank Dr.Samriti Mahajan at Lingayas Vidyapeeth,Faridabad for helping to gather the information on Dark data and to make valiant efforts to correct and improve earlier versions of this text.

References

1. Alan Dayley, "Innovation Insight: File Analysis Innovation Delivers an Understanding of Unstructured Dark Data", Gartner Research G00250750, pp. 6-12, 2013.

2. “Data age 2025” – International Data Corporation (IDC)

3. Datta, Shoumen Palit Austin. "Emergence of Digital Twins-Is this the march of reason?." Journal of Innovation Management 5, no. 3 (2017): 14-33.

4. "Managing data lakes in the big data era: What's a data lake and why has it become popular in the data management ecosystem?" Fang, Huang. 2015 IEEE International Conference on, pp. 820-824. IEEE, 2015.

5. From Data to Knowledge: Dark Data, Mar. 2018, [online] Available: www.ibm.com/developerworks/library/ba-data-becomes-knowledge-3/index.html.

6. “GDPR compliance after May 2018: A continuing challenge” – McKinsey & Company

7. //hbr.org/2013/12/small-businesses-need-big-data-too Commercial Momentum-Industrial Internet of Things Evolution." InTech 2017, no. 2 (2017).

8. Halevy, Alon Y., Flip Korn, Natalya Fridman Noy, Christopher Olston, Neoklis Polyzotis, Sudip Roy, and Steven Euijong Whang. "Managing Google's data lake: an overview of the Goods system." IEEE Data Eng. Bull. 39, no. 3 (2016): 5-14.

9. Kasey Panetta, “Gartner Top 10 Strategic Technology Trends for 2018”, Gartner, (2017)

10. Lasi, Heiner, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. "Industry 4.0." Business & Information Systems Engineering 6, no. 4 (2014): 239-242.

11. Mohammad Fikry Abdullah and Dr. Kamsuriah Ahmad, "The Mapping Process of Unstructured Data to Structured Data", 3rd International Conference on Research and Innovation in Information Systems (ICRIIS’13), 2013

12. Miloslavskaya, Natalia, and Alexander Tolstoy. "Big Data, Fast Data and Data Lake Concepts." Procedia Computer Science 88 (2016): 300-305.

13. Rosen, Roland, Georg von Wichert, George Lo, and Kurt D. Bettenhausen. "About the importance of autonomy and digital twins for the future of manufacturing." IFAC-PapersOnLine 48, no. 3 (2015): 567-572.

14. Trajanov, Dimitar, Riste Stojanov, Milos Jovanovik, Vladimir Zdraveski, Petar Ristoski, Marjan Georgiev, and Sonja Filiposka. "Semantic sky: a platform for cloud service integration based on semantic web technologies." In Proceedings of the 8th International Conference on Semantic Systems, pp. 109-116. ACM, 2012.

15. What Is Data and Why Is It Important?, Sept. 2018, [online] Available: www.import.io/post/what-is-data-and-why-is-it-important/.

16. Wan, Jiafu, Shenglong Tang, Zhaogang Shu, Di Li, Shiyong Wang, Muhammad Imran, and Athanasios V. Vasilakos. "Software-defined industrial internet of things in the context of industry 4.0.