

Aligning IS/IT with Business Allows Organizations to Utilize Dark Data

LiyanaShirin Akbar, Khulood Al-Mutahr, Mohammed Nazeh

Abstract: Any data that is left unexplored by an organization is an opportunity lost and a potential security risk says Ganesh Moorthy (2018). This paper discusses about the importance of aligning information system and information technology with business and how that helps organizations to utilize dark data efficiently. Moreover, the types of strategic alignment models and how organizations should adapt those models are also briefly described. The concept of dark data, types of dark data and how organizations can make use of it are further explained in this paper. The impact of dark data and tools to extract dark data is also discussed in this paper. The insights and discussions that are stated in this paper would definitely benefit organizations to understand the importance of aligning the business with IS/IT and make good use of darkdata.

Keywords: Information System, Information Technolog Dark Data, Strategic Alignmen tFramework

I. INTRODUCTION

As defined by Margaret Rouse (2017), data is information that has been translated into binary digital form that is efficient for processing and analyzing. On the other hand, the information collected, processed and stored by the organizations during the regular business activities basically fail to utilize the data for other purposes are termed as Dark Data(Gartner, 2017). Dark Data includes all types of data that are not analyzed for any business intelligence or to help in business decision making. All organizations need proper information system (IS) and information technology (IT) to harness the data collected therefore the IS and IT should be strategically aligned with the business.

First of all, information system is a set of components for collecting, storing, and processing data to provide information, knowledge, and digital products. According to Vladimir (2016) all organizations in this era highly rely on information systems to carry out and manage their operations, interact with customers and suppliers, and compete in the marketplace. However, according to Wikipedia (2018) Information Technology is the use of computers to store, retrieve, transmit, and manipulate data or information, often in the context of a business. On the other hand, alignment is the degree to which the needs, demands, goals, objectives, and structures of two components are consistent, which in this case the components are business and IS/IT.

Revised Manuscript Received on December 28, 2018.

LiyanaShirin Akbar, Limkokwing University of Creative Technology, Inovasi 1-1, JalanTeknokrat 1/1, Cyberjaya, 63000 Cyberjaya, Selangor, Malaysia.

Khulood Al-Mutahr, Limkokwing University of Creative Technology, Inovasi 1-1, JalanTeknokrat 1/1, Cyberjaya, 63000 Cyberjaya, Selangor, Malaysia.

Mohammed Nazeh, Lecturer, Limkokwing University of Creative Technology, Inovasi 1-1, JalanTeknokrat 1/1, Cyberjaya, 63000 Cyberjaya, Selangor, Malaysia.

There are many strategic alignment frameworks and there have also been numerous studies on strategic alignment made to successfully align business with IS and IT. Llanos Cuenca, Angel Ortiz, and Andres Boza (2010), suggests that including enterprise engineering techniques into designing the strategic alignment framework has made the alignment on a organization a successful one. The framework that was developed by Llanos et al. (2010) is as shown below in figure 1.

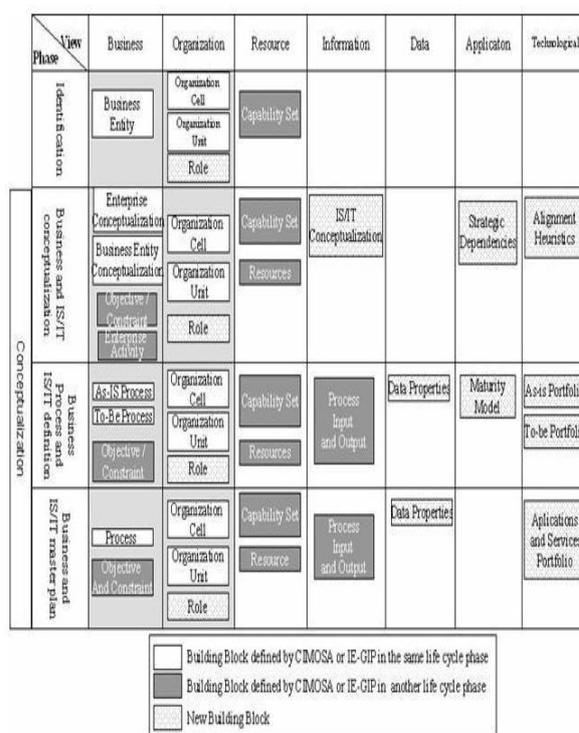


Fig 1: Strategic Alignment Framework (Llanos et al., 2010)

Conversely, the Strategic Alignment Model (SAM) – developed by Henderson and Venkatraman (1993)– recommends the organizations that need to integrate business and IT components at three main levels and they are as follows:

- Strategies (external integration or intellectual alignment)
- Infrastructures (internal integration or operational alignment)
- Strategies and infrastructures (cross-domain integration)

Based on several case study researches, Jennifer E. Gerow et.al states that Strategic Alignment Model (SAM) accurately reflects alignment concepts used in modern

Aligning IS/IT with Business Allows Organizations to Utilize Dark Data

businesses and it is also necessary for organizations to adapt and change in response to the technological advances. Figure 2 shown below is the SAM model that was developed by Henderson and Venkatraman (1993).

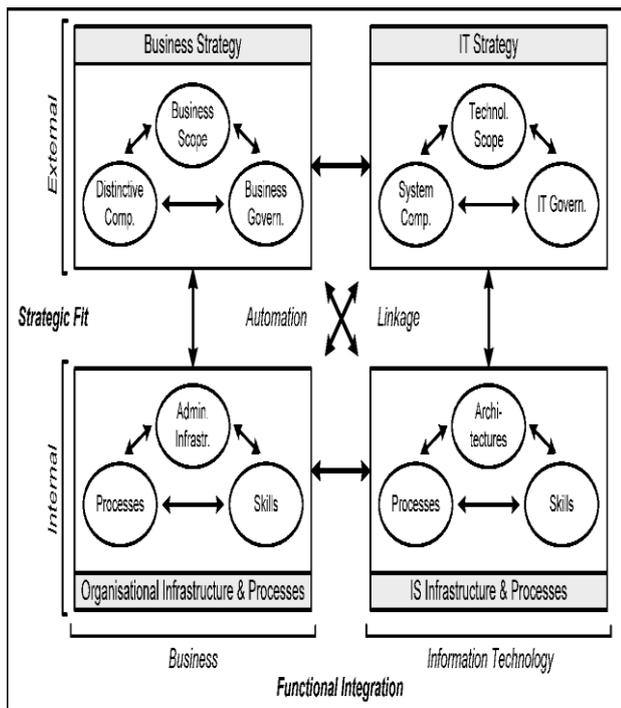


Fig 2: Strategic Alignment Model (Henderson and Venkatraman, 1993)

Moreover, Strategic Alignment Maturity Model (SAMM) is also one of the main alignment models developed by Luftman (El-Masri et.al, 2015). The SAMM model is denoted as a bottom-up prescriptive model that evaluates and improves as a mature organization that aligns wit IS/IT and other businesses. The summary of the SAMM model is as shown in figure 3.

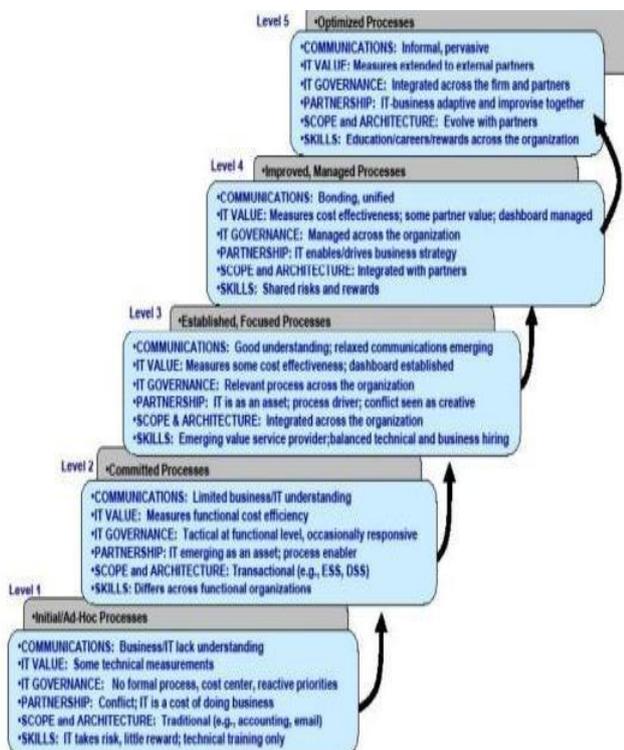


Fig 3: Strategic Alignment Maturity Model (Luftman, 2007)

Therefore, organizations are given a wide range of strategic alignment frameworks that have been proven to be successful for them to incorporate into the business. By doing this, organizations are able to successfully identify, utilize and benefit from the dark data that has been accumulated throughout the years.

According to Willian et al. (2018) dark data is actually a term that is used especially to refer to operational data that is left unanalyzed in an organization. Kaushik Pal (2015) states that organizations gather huge volumes of data which, they believe will help improve their products and services. Data that are collected by organizations are data on users' usage of products, internal statistics about software development processes, and even website visits. However, large portions of the collected data are never analyzed and such data is known as dark data. Dark data is a subset of big digital information that is not being used, but it constitutes the biggest portion of the total volume of big data collected by organizations in a year.

Furthermore, dark data can also be located in online database including the logged files and archived data that are stacked in large organizations or store rooms. It also contains those data objects and types that are yet to be assessed for any kind of business or competitive intelligence or helping aid in business related decision makings. Dark data is mostly found to be difficult for analysis as they are stored in those locations that are difficult to access and analyze. Therefore, this difficulty in its analysis makes the overall process more costly and sometimes it include the data objects that were never seized by the enterprise or were external to the organization, such as with partners or customers (Shane Ryan, 2014).

II. PROBLEM STATEMENT

Organizations are not completely utilizing the data that is being collected and the accumulation of these unused data which is called dark data could be a potential security risk and an opportunity lost for an organization. According to Shane Ryan (2014), companies will continue to waste 80% of customer data that has been collected. Dark data is a threat for organizations and it silently multiplies beyond the control of the organization says Andy Berry (2018). On top of that, according to IDG (2016), dark data is growing at a rate of 62% per year and by the year 2022 IDG has forecasted that 93% of all data will be unstructured.

Mckinsey (2011) says that the biggest challenge for organizations today is how to handle big data. This is because these organizations have access to a wealth of information which they do not know how to gain value from as it is sitting in its raw form in either a semi-structured or unstructured state and as a result, organizations are not even sure it is worth keeping (Chris Eaton and Paul C. Zikopoulos, 2011). In addition, there are further arguments on organizations facing massive amounts of data and organizations not knowing how to manage this data are overwhelmed by it.



III. OBJECTIVE

The main objective of this research is to emphasize on the importance of strategically aligning business with IS and IT in organizations in order to identify and utilize the dark data to make better business decisions and improve the organizations performance.

IV. LITERATURE REVIEW

According to Thierry Nautin (2014) agility is one of the main measures of an organization's success. When strategy, goals, and meaningful purpose support one another it gives an organization a huge advantage and this is called achieving real alignment. This enables the organization to have a clearer sense of what to do at any given time and the organization will be able to focus less on deciding what to do and focus more on just doing it (Thierry Nautin, 2014). Furthermore, the alignment of IS/IT and businesses enable the information system (IS) to make an effective impact on the dynamic business environment. Organizations having keen interest in such alignments enable the business to increase their value and performance.

As stated by El-Masri et.al (2015), it is still a challenge to investigate the IS/IT and business alignments as such concepts continues to evolve and transform with advances in technology. Even though, the alignment of business and IS/IT increases the profitability of the organization, it could also result in wasted resources, poor data management and failed IT initiatives that leads to poor financial and organizational outcomes if it is not strategically aligned (Kearns and Lederer, 2003). Lindsay Wise (2016) has stated that in order to create an effective and lasting data management strategy, people, processes and technology must be incorporated strategically.

Gerow et al. (2014) explained that recent literature evaluations concluded that 65 out of 184 articles tried to create new measures to compute IS/Business alignments. Based on several researches, it is found that organizations that are aligned strategically are able to leverage IT to support overall business objectives and exploit opportunities in the market in a more frequent manner. This enables the organizations to create a sustainable competitive advantage and achieve great profits (Jennifer E.Gerow et.al, 2014). However, Jennifer E. Gerow et.al (2014) concludes that the relationship between alignment and organization performance has been inconsistent due to the usage of static frameworks that do not respond to the technological and environmental changes.

Strategic management research should focus more on highlighting the day-to-day activities, contexts, and processes relating to the strategic outcomes (Peppard et al. 2014). There is no one-size-fits-all strategic alignment framework that will be effective for all organizations, says El-Masri et al. (2015). For this reason there are many frameworks, models, and strategies that are being researched and developed in order to help all organizations to have a dynamic strategy for aligning information system and information technology.

Moreover, Quostar (2018) has stated that utilizing a strategic IT-business alignment model helps organizations to improve the business's overall performance and most

importantly allows organizations to beneficially make use of all types of data. Dark data which is also known as dusty data is a large amount of data that is being accumulated in organizations. As explained by Techopedia (n.d) dark data is a type of unorganized, unlabeled and untapped data that is found in data repositories and has not been analyzed. Dark data is mostly neglected by business and IT administrators in terms of its value and having not proper understanding on how dark data differs frombig data.

Figure 4 visually explains the difference between dark data and big data. Dark data is actually a subset of Big data that is not used or analyzed for any purpose (Ben Austin, 2014). A wealth of information lies below the surface of traditional enterprise data and getting to it requires cutting edge technologies (Hp/Syncsort, n.d).

Mining dark data

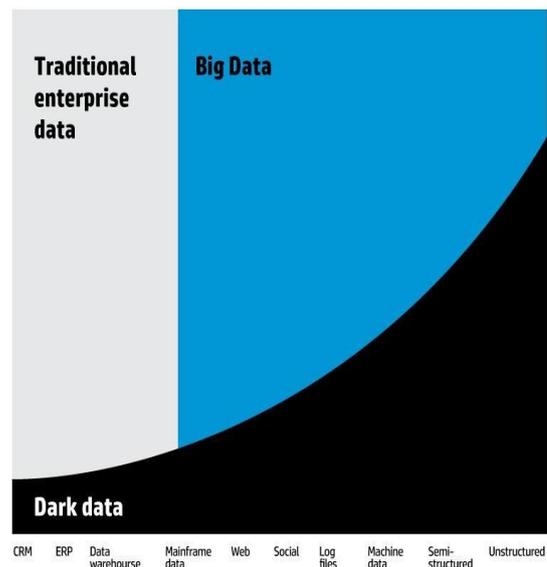


Fig : 4 A wealth of information lies below the surface of traditional enterprise data

Vinay R. Rao (2018) said that "Data becomes inaccessible and unusable for many reasons, but the principal reason is that big data is, well, big. Not just big, but mind- bogglingly enormous. Based on several social media statistics it is shown that in 2017, every minute on average, Twitter users sent a half million tweets, and 4 million Facebook users clicked Like"(kd nuggets, 1997). It is surprising because at the time of data collection, organizations assume that the data is going to provide value and they invest a lot on data collection so both monetarily and otherwise, data should be considered important.

One of the reasons why there is so much of dark data accumulated in organizations is because of the data collected that is difficult to access after being stored on obsolete devices. There are three main types of dark data that can be found in organizations. Firstly, data that is currently not being collected also adds to dark data. Secondly, it becomes difficult to access the collected at the right time and location leading to be enlisted as dark data. Finally, the collected and available dark data is not fully productive or applicable.



V. IMPACT

Kdnuggets (1997) reported that dark data has a crucial impact on the data quality that is analyzed for extraction of valuable information. Dark data create obstacles in accessing the essential information, confirming its origin, and obtaining the necessary information for effective decision making. The impact of dark data is evident from the following factors:

- 1) **Data accessibility:** Access to essential information that results in improving the analysis is lost which further leads to the inability to access unstructured data, such as images, audios, or videos.
- 2) **Data accuracy:** Data analysis accuracy is depended on the input data. Analysis when done with utmost accuracy leads to the qualitatively information extraction. Hence, dark data has also significant influence in data accuracy.
- 3) **Data audit ability:** The failure in tracing data can omit it from the analysis part, which would eventually affect the data quality leading to faulty data-driven decision making.

VI. DATA-EXTRACTION

Data extraction is the act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage (data migration). The import into the intermediate extracting system is thus usually followed by data transformation and possibly the addition of metadata prior to export to another stage in the data workflow. Institutions that are indulged in extracting dark data invest more for considerable engineering effort to gain more benefits. Dark data often carried the valuable information that is not accessible any other format. Moreover, extraction of the dark data is less expensive and implements less engineering effort when such techniques and tools are used. The quality of analytics progresses dramatically with increase in the access of better data sources and information. Also, dark data extraction process also results in organizations of being less exposed to risks and liability while acquiring the sensitive as well as valuable information. Due to the open source dark data extraction tools, the dark data extraction turns out to be very valuable. technologies are extremely valuable with the many.

There are several tools used to extract dark data successfully. Firstly, DeepDive is an open source tool that was developed by Stanford University (2017) and it was commercially supported by Lattice Data. Development is no longer active with Apple's acquisition of Lattice Data in 2017. DeepDive system is utilized to access the valuable information from dark data that is buried in varied forms lacking structures, i.e. texts, tables, figures, or images, which cannot be managed by the existing software to convert the dark data to light version. This conversion is possible by creating structured data (SQL tables) from the unstructured information (text documents), which is later integrated into the existed structure. This new form of data management system enables to tackle the issues related to the extraction, integration, and prediction in a single system and allows the users to create rapid and sophisticated end-to-end data

pipelines, such as dark data Business Intelligence (BI) systems.

This DeepDive system allows the users to access the main portion of the system that improves the application quality directly during the system building process. DeepDive extracts dark data to light by constructing a structured data from the unstructured information and integrating it with the existing structured database. an individual can now implement the standard tools that feed on the structured form of data; e.g., visualization tools like Tableau or analytics tools like Excel. DeepDive is different from the traditional systems in varied ways:

- It instructs about thinking features instead of algorithms to the developer.
- Paleo Deep Dive consists of higher quality data compared to human volunteers for the extraction of the complex knowledge in scientific domains enlisted under the relation extraction competitions.
- DeepDive data is noisy and imprecise as sometimes it misspells names, speaks ambiguous languages, and make mistakes like. For example, in case a fact of DeepDive data is produced with 0.9 probability, the fact may be 90% true.
- DeepDive utilizes large data collected from varied sources, which are applications using DeepDive data extracted from millions of documents, web pages, PDFs, tables, and figures.
- Developers are able to use their understanding about domains with the help of DeepDive for improving the quality of the outcomes inferred from the inference (learning) process.
- DeepDive allows the machine learning systems to use the data for learning purpose with the involvement of tedious training procedures for each prediction.
- DeepDive's secret techniques are scalable, high in performance inference, and learning engine. It has become a part of the commercial and open source tools including MADlib, Impala, a product from Oracle, Hogwild, Microsoft's Adam, and other major web companies.

Secondly, Snorkel is another tool that was also developed by the Stanford University, (2017). Snorkel fastens the dark data extraction time by developing tools that create datasets with an intention to help in trainings to learn the dark data extraction algorithms. According to HazyResearch (2018) Snorkel is a system for rapidly creating, modeling, and managing training data. The goal is to make routine Dark Data and other prediction tasks dramatically easier. At its core, Snorkel focuses on a key bottleneck in the development of machine learning systems: the lack of large training datasets. In Snorkel, a user implicitly creates large training sets by writing simple programs that label data, instead of performing manual feature engineering or tedious hand-labeling of individual data items.

Today's state of the art machine learning models require massive labeled training sets which usually do not exist for

real-world applications. Instead, Snorkel is based around the new data programming paradigm, in which the developer focuses on writing a set of labeling functions, which are just scripts that programmatically label data. Surprisingly, by modeling a noisy training set creation process in this way, we can take potentially low-quality labeling functions from the user, and use these to train high-quality end models.

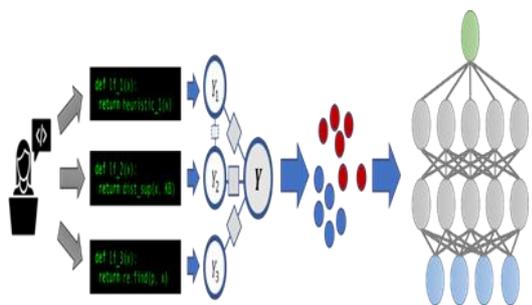


Fig 5: - Snorkeling Model (Stanford University, 2017)

Finally, Dark Vision services is a technology demonstration leveraging Cloud Functions and services by an application that processes videos to discover what is inside of them (Frederic Lavigne, 2017). By analyzing individual frames and audio from videos with IBM Watson Visual Recognition and Natural Language Understanding, Dark Vision builds a summary with a set of tags, famous people or landmarks detected in the video by using a simple architecture for extracting frames and audio from a video. The figure 6 below shows how Dark Vision extracts information.

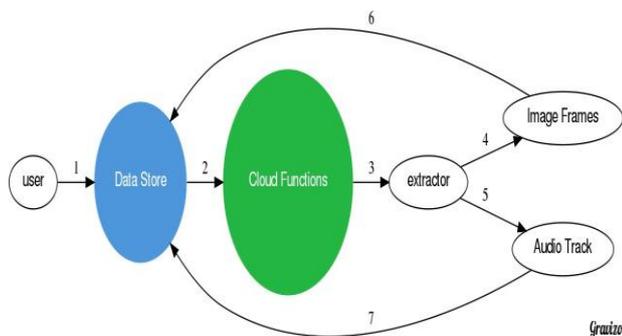


Fig 6: Dark Vision Services - Extracting Information

VII. DISCUSSION

This paper was motivated by a desire to help organizations to utilize and benefit from dark data, which is only possible if the information system and information technology are aligned strategically with the business. There were many models and strategies that are made available for organizations to follow in order to align the organization with IS/IT. When this alignment is strong, dynamic and accurate, organizations are able to identify, analyze, and utilize the collection of dark data. The dark data that is abundant in organizations will be put to good use and this would definitely improve the organizations performance.

Regardless of the type of dark data organizations collect, or how it is stored, the key to keeping data out of the dark is to ensure that the organizations have a means of translating it from one form to another and ingesting it easily into whichever analytics platform you use, says Dayley to Gartner (2017). In order for the organization to even

translate the collected data from one form to another and ingest it, it has to have a strategically aligned IS/IT with its business. Steven Astorino (2016) states that in this current information-driven era, all the employees in an organization share a common need and that is delivering the right information at the right time, in context.

VIII. CONCLUSION

In conclusion, the alignment of business with information system and information technology helps organization to successfully make use of dark data that has been accumulated. The process of identifying and storing data is highly dependent on the IS and IT system in the organization. Therefore the alignment needs to be properly set in order for the organization to benefit from not only dark data but any type of data. Based on the researches made in this paper, it is shown that in order for organizations to even use any of the tools to extract dark data it would need information system and information technology to be aligned with the business.

REFERENCES

- Ganesh Moorthy, 2018, "Dark Data: The two sides of the same coin" [Online]. Available at: <http://analytics-magazine.org/dark-data-two-sides-coin/> [Accessed on: 5th December 2018]
- Margaret Rouse, 2017, "Data" [Online]. Available at: <https://searchdatamanagement.techtarget.com/definition/data> [Accessed on: 5th December 2018]
- Gartner, 2017. "How to tackle dark data" [Online]. Available at: <https://www.gartner.com/smarterwithgartner/how-to-tackle-dark-data/> [Accessed on: 5th December 2018]
- Vladimir, 2018. "Information System" [Online]. Available at: <https://www.britannica.com/topic/information-system> [Accessed on: 6th December 2018]
- El-Masri, Mazen; Orozco, Jorge; Tarhini, Ali; and Tarhini, Takwa. "The Impact of IS- Business Alignment Practices on Organizational Choice of IS-Business Alignment Strategies" (2015). PACIS 2015 Proceedings, Paper 215.
- Wikipedia (n.d). "Information Technology" [Online]. Available at: https://en.wikipedia.org/wiki/Information_technology. [Accessed on: 7th December 2018]
- Llanos Cuenca, Angel Ortiz, and Andres Boza 2010, 'Business and IS/IT Strategic Alignment Framework' Research Centre on Production Management and Engineering, p. 24-31
- Venkatraman and Henderson, 1993. "Strategic alignment: Leveraging information technology for transforming organizations" IBM Systems. Accessed on: 7th December 2018
- Jennifer E. Gerow, Jason Bennett Thatcher and Varun Grover, "Six types of IT-business strategic alignment: an investigation of the constructs and their measurement" (2014).
- Luftman, Jerry, "An Update on Business-IT Alignment: 'A Line' Has been Drawn", MIS Quarterly Executive, Vol.6, Issue 3, 165-177, 2007
- Willian Dimitrov et al, 2018. "Types of dark data and hidden cybersecurity risks" [Online]. Available at: https://www.researchgate.net/publication/329119026_Types_of_dark_data_and_hidden_cybersecurity_risks [Accessed on: 7th December 2018]
- Shane Ryan, 2014. "Illuminating Dark Data," Accessed on: 5th December 2018.
- Andy Berry, 2018. "Shed light on your dark data before GDPR comes into force" [Online]. Available at: <https://www.cio.com/article/3269009/big-data/shed-light-on-your-dark-data-before-gdpr-comes-into-force.html> [Accessed on: 8th December 2018]
- Kaushik Pal, 2018. "Importance of dark data and big data" [Online]. Available at: <https://www.kdnuggets.com/2015/11/importance-dark-data-big-data-world.html>. [Accessed on: 8th December 2018]

Aligning IS/IT with Business Allows Organizations to Utilize Dark Data

16. Lyndsay Wise, 2016. "Aligning Strategy with Data Management". [Online]. Available at: <https://www.cio.com/article/3082553/leadership-management/aligning-strategy-with-data-management.html> [Accessed on: 8th December 2018]
17. IDG, 2016. "2016 Data and Analytics Research" [Online]. Available at: <https://www.idg.com/tools-for-marketers/tech-2016-data-analytics-research/> [Accessed on: 9th December 2018]
18. MCKINSEY, "Big Data: The Next Frontier For Innovation, Competition and Creativity," MCKINSEY Global Institute, New York, 2011.
19. Paul Zikopoulos, Chris Eaton, Diros Dirk, Tom Deutsch, and George Lapis, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, 1st ed., Steve Sit, Ed. New York, United States: MCGRAW-HILL, 2012.
20. Thierry Nautin, 2014. "The Aligned Organization" [Online]. Available at: <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/Operations/Our%20Insights/The%20lean%20management%20enterprise/The%20aligned%20organization.aspx>. [Accessed on: 9th December 2018]
21. Kearns and Lederer, 2003. "A Resource-Based View of Strategic IT Alignment: How Knowledge Sharing Creates Competitive Advantage" [Online]. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-5915.02289> [Accessed on: 9th December 2018]
22. Peppard, Joe, Robert D. Galliers, and Alan Thorogood. "Information systems strategy as practice: Micro strategy and strategizing for IS." *Journal of Strategic Information Systems* (23:1), 2014, pp. 1-10.
23. Quostar, 2018. "Why business leaders must align IT strategy with business strategy" [Online]. Available at: <https://www.quostar.com/blog/why-align-business-strategy-it-strategy/> [Accessed on: 9th December 2018]
24. Ben Austin, 2014. "Dark Data: What is it and Why Should I Care?" [Online]. Available at: <https://www.r1soft.com/blog/dark-data-what-is-it-and-why-should-i-care> [Accessed on: 9th December 2018]
25. Hp/Syncsort, n.d. "4 Ways to Use Dark Data" [Online]. Available at: <http://blog.syncsort.com/2017/05/big-data/4-dark-data-examples-use-cases/> [Accessed on: 10th December 2018]
26. Vinay R. Rao, 2018. "From Data to Knowledge" [Online]. Available at: <https://www.ibm.com/developerworks/library/ba-data-becomes-knowledge-1/index.html>. [Accessed on: 10th December 2018]
27. Stanford University, 2017. "Extracting Databases from Dark Data with DeepDive" by Michael Cafarella et al. [Online]. Available at: <https://cs.stanford.edu/people/chrisre/papers/modiv923-zhangA.pdf> [Accessed on: 10th December 2018]
28. Hazy Research, 2018. "Snorkel: Fast Training Set Generation for Information Extraction" [Online]. Available at: https://hazyresearch.github.io/snorkel/pdfs/snorkel_demo.pdf [Accessed on: 10th December 2018]