

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

Liyana Shirin Akbar, Khulood Al-Mutahr, Mohammed Nazeem

Received 14 September 2018 ▪ Revised 23 October 2018 ▪ Accepted 24 November 2018

**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 dark data.

**Keywords:** Information System, Information Technology, Dark Data, Strategic Alignment Framework.

## BACKGROUND

Education is a core of the public life, including cultural, which unites and cements the society, indicates the future development and ensures this development enables the society to have protected its competitive place in the global division of labor that maintains social stability. The system of education is a complex social instrument of solving social problems, modernization and optimization of public life, integration of the entire community, the continuity of the educational process throughout life, providing competitive advantage and appropriate future for each subject of the educational space.

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 that organizations collect, process and store during regular business activities, but generally fail to make use of for other purposes are known as Dark Data, stated 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.

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.

---

<sup>1,2,3</sup>Limkokwing University of Creative Technology, Inovasi 1-1, Jalan Teknokrat 1/1, Cyberjaya, Selangor, Malaysia.

Phase \ View	Business	Organization	Resource	Information	Data	Application	Technological
Identification	Business Entity	Organization Cell Organization Unit Role	Capability Set				
Conceptualization	Enterprise Conceptualization	Organization Cell	Capability Set	IS/IT Conceptualization		Strategic Dependencies	Alignment Heuristics
	Business Entity Conceptualization	Organization Unit	Resources				
	Objective / Constraint Enterprise Activity	Role					
	Business Process and IS/IT definition	As-IS Process To-Be Process	Organization Cell Organization Unit Role	Capability Set Resources	Process Input and Output	Data Properties	Maturity Model
Business and IS/IT master plan	Process Objective And Constraint	Organization Cell Organization Unit Role	Capability Set Resource	Process Input and Output	Data Properties		Applications and Services Portfolio

Building Block defined by CIMOSA or IE-GIP in the same life cycle phase  
 Building Block defined by CIMOSA or IE-GIP in another life cycle phase  
 New Building Block

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

Conversely, Henderson and Venkatraman (1993) developed a Strategic Alignment Model (SAM) which recommends that organizations 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 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).

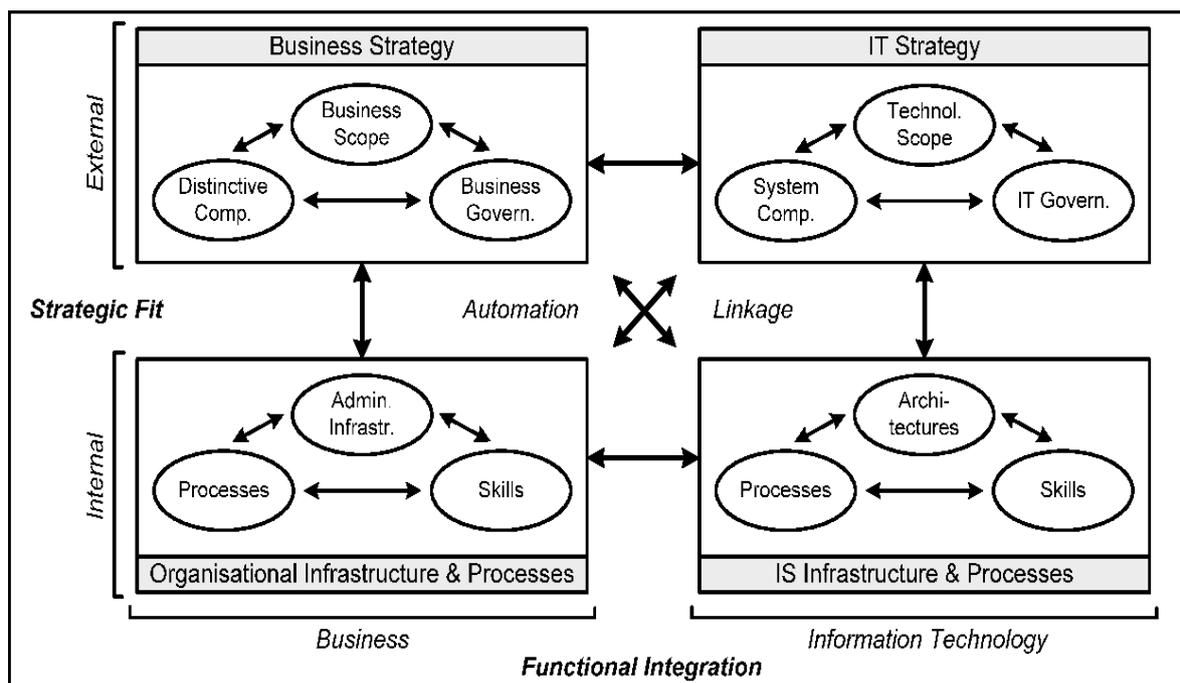


Figure 2: Strategic Alignment Model (Henderson and Venkatraman, 1993)

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 a bottom-up prescriptive model that can be used to evaluate and improve an organizational maturity in aligning IS/IT and business. The summary of the SAMM model is as shown in figure 3.

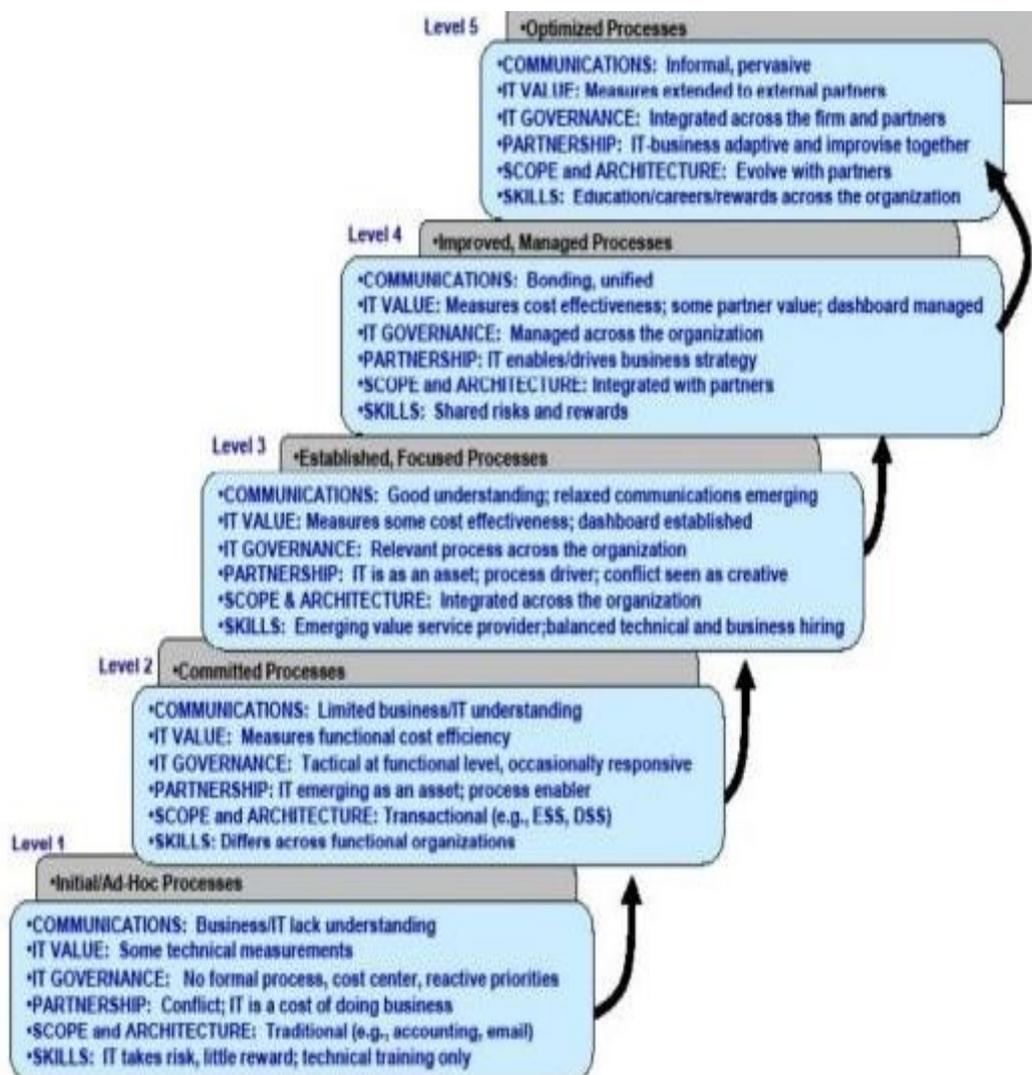


Figure 3: Strategic Alignment Maturity Model (Luftman, 2007)

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 is also found in log files and data archives stored within large organization class data storage locations and it includes all data objects and types that have yet to be analyzed for any business or competitive intelligence or aid in business decision making. In most cases, dark data is difficult to analyze and is stored in locations where analysis is difficult. The overall process can be costly and also can include data objects that have not been seized by the enterprise or data that are external to the organization, such as data stored by partners or customers (Shane Ryan, 2014).

## 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.

## 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.

## 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 business allows information system (IS) to influence the dynamic business environment. Organizations have a keen interest in this alignment because it enables the business to increase its value and improve its performance.

As stated by El-Masri et.al (2015) in research, investigating IS/IT and business alignment is still a challenge since the concept continues to evolve and transform as the technology advances. 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) explains that according to a recent examination of the alignment literature, 65 out of 184 articles attempted to create new ways to measure IS/Business alignment. 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.

As stated by Peppard et al. (2014), strategic management research must move towards a micro focus which highlights the day-to-day activities, contexts, processes and content that relate to strategic outcomes rather than moving towards macro focus. There is no one-size-fits-all strategic alignment framework that all organizations can adopt 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 from big 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).

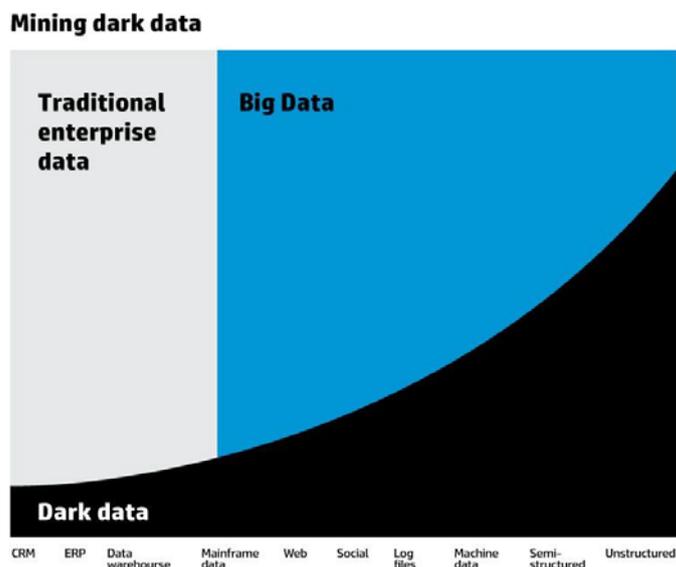


Figure 4: A wealth of information lies below the surface of traditional enterprise data

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"(kdnuggets, 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 collection of data that is no longer accessible because it has been stored on devices that have become obsolete. 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, data that is being collected, but is difficult to access at the right time and place also becomes a part of dark data. Finally, dark data is also data that is collected and available, but is not yet been productized, or fully applied.

## IMPACT

As reported by (kdnuggets, 1997) the main impact of dark data is on the quality of data used for analysis to extract valuable information. Dark data makes it difficult to access and find vital information, confirm its origins, and promptly obtain essential information to make good, data-driven decisions. The impact on quality stems from the following factors:

- 1) **Data accessibility:** loss of access to essential information that would improve analysis is a result of the inability to access data that is unstructured or in a different media format, such as images, audio, or video.
- 2) **Data accuracy:** The accuracy of a data analysis depends on the accuracy of the input data. Accurate analysis leads to the extraction of qualitatively more worthy information. Hence, dark data has a significant impact on the accuracy of the extracted information and the quality of the information produced by that analysis.
- 3) **Data audit ability:** The inability to trace the provenance of data can lead to its omission from analysis, thereby affecting the data quality. This, in turn, can lead to faulty data-driven decision making.

## 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. According to (Vinay R. Rao, 2018) organizations extracting dark data incur an expense and spend considerable engineering effort, but there are many benefits to doing this. Dark data is valuable because it often holds information that is not available in any other format. Adding that, dark data extraction is less expensive and uses less engineering effort when using these techniques and tools. With access to better data sources and more information, the quality of analytics improves dramatically. Also, extracting dark data leaves organizations less exposed to risks and liability in securing sensitive information. In addition to the dark data itself, dark data extraction technologies are extremely valuable with the many open source dark data extraction tools.

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 is a system to extract value from dark data "the great mass of data buried in text, tables, figures, and images, which lacks structure and so" is essentially cannot be processed by existing software that helps bring dark data to light by creating structured data (SQL tables) from unstructured information (text documents) and integrating such data with an existing structured database. It is a new type of data management system that enables one to tackle extraction, integration, and prediction problems in a single system, which allows users to rapidly construct sophisticated end-to-end data pipelines, such as dark data BI (Business Intelligence) systems.

By allowing users to build their system end-to-end, DeepDive allows users to focus on the portion of their system that most directly improves the quality of their application. DeepDive helps bring dark data to light by creating structured data from unstructured information and integrating such data with an existing structured database. With the data in a database, one can use a variety of standard tools that consume structured data; e.g., visualization tools like Tableau or analytics tools like Excel. DeepDive differs from traditional systems in several ways:

- DeepDive asks the developer to think about features and not the algorithms.
- DeepDive systems can achieve high quality: PaleoDeepDive has higher quality than human volunteers in extracting complex knowledge in scientific domains in entity relation extraction competitions.
- DeepDive is aware that data is often noisy and imprecise: names are misspelled, natural language is ambiguous, and humans make mistakes. For example, if DeepDive produces a fact with probability 0.9, the fact is 90% likely to be true.
- DeepDive is able to use large amounts of data from a variety of sources. Applications built using DeepDive have extracted data from millions of documents, web pages, PDFs, tables, and figures.
- DeepDive allows developers to use their knowledge of a given domain to improve the quality of the results by writing simple rules that inform the inference (learning) process.
- DeepDive is able to use the data to learn "distantly". In contrast, most machine learning systems require tedious training for each prediction.
- DeepDive's secret is a scalable, high-performance inference and learning engine. The techniques pioneered in this project are part of commercial and open source tools including MADlib, Impala, a product from Oracle, and low-level techniques, such as Hogwild!. They have also been included in Microsoft's Adam and other major web companies.

Secondly, Snorkel is another tool that was also developed by the Stanford University, (2017). Snorkel accelerates dark data extraction by developing tools to create datasets to help train learning algorithms for dark data extraction. 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.

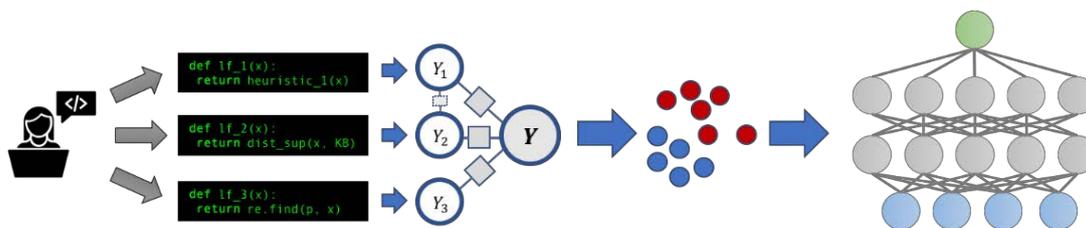


Figure 5: Snorkeling Model (Stanford University, 2017)

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.

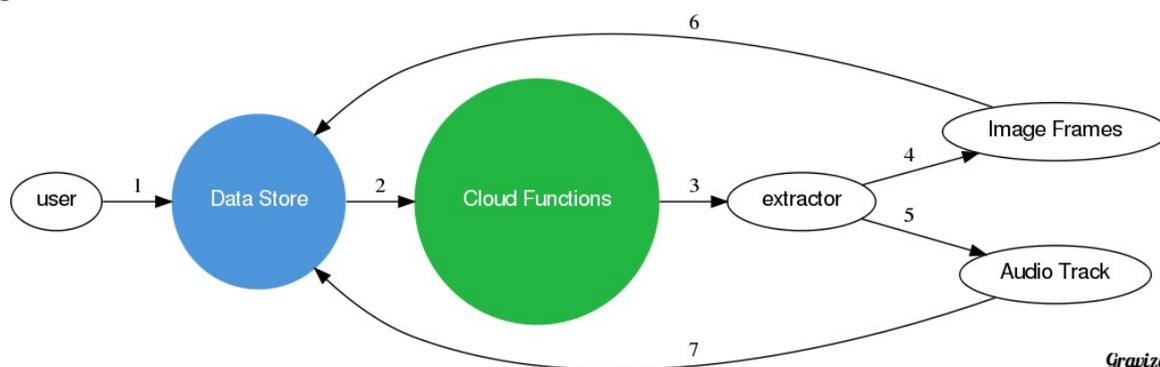


Figure 6: Dark Vision Services - Extracting Information

## 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.

## 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

- [1] Ganesh, M. (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]
- [2] Margaret, R. (2017). "Data" [Online]. Available at: <https://search.data.management.techtarget.com/definition/data> [Accessed on: 5th December 2018]

- [3] 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]
- [4] Vladimir, 2018. "Information System" [Online]. Available at: <https://www.britannica.com/topic/information-system> [Accessed on: 6th December 2018]
- [5] El-Masri, M., Orozco, J., Tarhini, A., & Tarhini, T. (2015). The Impact of IS-Business Alignment Practices on Organizational Choice of IS-Business Alignment Strategies. *PACIS*.
- [6] Wikipedia (n.d). "Information Technology" [Online]. Available at: [https://en.wikipedia.org/wiki/Information\\_technology](https://en.wikipedia.org/wiki/Information_technology). [Accessed on: 7th December 2018]
- [7] Cuenca, L., Ortiz, A., & Boza, A. (2010). Business and IS/IT strategic alignment framework. *Doctoral Conference on Computing, Electrical and Industrial Systems*, 24-31.
- [8] Henderson, J. C., & Venkatraman, H. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM systems journal*, 38(2.3), 472-484.
- [9] Gerow, J. E., Thatcher, J. B., & Grover, V. (2015). Six types of IT-business strategic alignment: an investigation of the constructs and their measurement. *European Journal of Information Systems*, 24(5), 465-491.
- [10] Luftman, J., & Kempaiah, R. (2007). An update on business-IT alignment: "A line" has been drawn. *MIS Quarterly Executive*, 6(3), 165-177.
- [11] Willian, D. (2018). Types of dark data and hidden cyber security risks" [Online]. Available at: [https://www.researchgate.net/publication/329119026\\_Types\\_of\\_dark\\_data\\_and\\_hidden\\_cybers\\_eurity\\_risks](https://www.researchgate.net/publication/329119026_Types_of_dark_data_and_hidden_cybers_eurity_risks) [Accessed on: 7th December 2018]
- [12] Shane, R. (2014). Illuminating Dark Data, 2018.
- [13] Andy, B. (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]
- [14] Kaushik, P. (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]
- [15] Lyndsay, W. (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]
- [16] IDG, 2016. "2016 Data and Analytic Research" [Online]. Available at: <https://www.idg.com/tools-for-marketers/tech-2016-data-analytics-research/> [Accessed on: 9th December 2018]
- [17] MCKINSEY, (2011). Big Data: The Next Frontier for Innovation, Competition and Creativity, MCKINSEY Global Institute, New York.
- [18] Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.
- [19] 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.nashx>. [Accessed on: 9th December 2018]
- [20] Shweta, G., & Sachinmishra, A. (2017). Modeling of solar/wind hybrid energy system using mtalab simulink. *International journal of pharmacy research & technology*, 7(2), 38-41.
- [21] Peppard, J., Galliers, R. D., & Thorogood, A. (2014). Information systems strategy as practice: Micro strategy and strategizing for IS. *J. Strategic Inf. Sys.*, 23(1), 1-10.
- [22] 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]
- [23] Techopedia, n.d. Dark Data" [Online]. Available at: <https://www.techopedia.com/definition/29373/dark-data> [Accessed on: 9th December 2018]
- [24] Ben, A. (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.R. (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/chismre/papers/modiv923-zhangA.pdf> [Accessed on: 10th December 2018]
- [28] Pradeep, M., Abinya, R., Sathya Anandhi, S., & Soundarya, S. (2017). Dynamic Smart Alert Service for Women Safety System. *International Journal of Communication and Computer Technologies*, 5(2), 58-66.