International Journal of Corporate Social Responsibility and Management
Volume1, Issue 1 2025

An Overview of Data Management and Big Data's Effectiveness

Rainer A. Sommer

Associate Professor

George Mason University, Founder's Hall, Arlington, USA,
Department of Enterprise Engineering and Public Policy

Research Article

Date of Submission: 14-05-2025 **Date of Acceptance: 10-05-2025** **Date of Publication: 11-07-2025**

ABSTRACT

Businesses are there to provide a clientele with products and services. Businesses, whether in the public or private sectors, are made up of management structures, procedures, and technology that all work toward the same goal of creating, maintaining, and enhancing value for customers. Utilizing business intelligence and data analytical techniques that primarily focus on the analysis of structured and unstructured data, this latter idea—adding value—has pushed organizations to delve deeper into their data in the hopes of creating an additional competitive advantage with which to improve market impact or position.

Keywords: *Business intelligence, data analytics, structured data, unstructured data, data lake*

1. Trusted Data and Data Type Definitions

To put it simply, there are many different definitions of business intelligence, and they often vary depending on the field being discussed and the kind of data being examined. Structured and unstructured data are two different categories into which data may be divided. The majority of businesses utilize structured data to represent conventional business data, such sales, accounting, and human resource data, that are trusted and used to power daily operations. This data gives management the capacity to function continuously and is stored in relational databases that are very reliable and safe (Williams and Williams, 2003). The following traits are often assumed to be present in reliable data. • Authority: it is current and acknowledged as the reference copy of the pertinent information. • Authenticity: it can be traced back to its original source and is what it claims to be. • Reliability: it is a complete and accurate depiction of the pertinent information, transaction, or business procedure. • Integrity: it is comprehensive, unchanged, and maintains the chain of custody and context. • Usability: it is reachable, locateable, retrievable, presentable, and interpretable. On the other hand, unstructured data is essential to the operation of the business. It is often recorded as a byproduct of routine corporate operations and kept in non-relational data repositories, sometimes known as "Data Lakes." This information, which is often unreliable as well, is produced by monitoring website clicks and gathering user or customer opinions from online sources or documents such blogs, forums, social media platforms, bulletin boards, and phone conversations, to mention a few. The most crucial thing to keep in mind is that a large portion of this data is unstructured and often just receives a time stamp and a generic category qualifying ID called Meta-Data prior to being stored in data lakes. Although a large portion of this data may never be utilized, it is stored in the data lake in case it is needed for analysis in the future. In this regard, businesses aim to increase the effectiveness of their intra- and inter-organizational business operations (Sommer, 2017). Intra-organizational business activities are primarily focused on business processes that effectively support the unfettered flow of goods and information across internal organizational silos (or stovepipes). As a result, improving locally managed business

processes is highly prioritized. Inter-organizational business activities are more complex, since this model requires the enterprise to extend its internal processes beyond its traditional organizational boundaries to include other claimants. This model is often termed as an "Extended Enterprise" because it attempts to leverage integrated internal (Intra-organizational) efficiencies into a value chain that includes external suppliers and customers. This model would allow customer and suppliers to take advantage of the efficiencies created within the host organization to achieve an aggregate competitive advantage2. The extended enterprise is a model that expands the basic enterprise to include customers, suppliers, partners, and other corporate claimants. By definition, it also extends the traditional enterprise model to include new business models such as e-Commerce, Supply Chain Integration and Management, and Customer Relationship Management. Regardless of what types of business model an organization develops, the ultimate efficiency of the enterprise is governed by a complex balance of process flow, data access and integrity, regulatory and policy control, and business intelligence. Hence the proper management of data becomes a strategic competitive advantage (Schniederjans and Schniederjans, 2014)

2. The "Managing" of Big Data Management

Big Data Management is the organization, administration and governance of large volumes of both structured and unstructured data. Hence, the organization must address the management of data with the same zeal and caution as any other strategic asset that provides a long-term competitive advantage. Generally, a data management plan must at a minimum address the development, execution and supervision of various plans (i.e., backup/recovery, disaster, access, etc.), policies, programs and practices that control, protect, deliver and enhance the value of data and information assets. To aid in this process, the Data Management Association (DAMA) has provided the following framework to help guide the implementation of such a management structure[1].

- Data Governance,
- Data Architecture Management,
- Data Development,
- Database Operations Management,
- Data Security Management,
- Reference & Master Data Management,
- Data Warehousing & Business Intelligence Management,
- Document & Content Management, Metadata Management, and
- Data Quality Management.

The DAMA framework divides implementation and execution responsibilities among both management and IT functions and whose ultimate goal is to allow companies to locate valuable information in large sets of structured, unstructured, and semi-structured data from a variety of sources, including traditional relational data as well as forum posts, images, blogs, texts, websites, clicks, cookies, e-mail, pop-ups, spam, and social media scans.

While each of these concepts merits its own investigation as to its overall contribution to organizational efficiency, increasingly the focus has turned to big data analysis for finding the next level of competitive advantage. Not surprisingly, due to the complexity of their vast business relationships, companies that have embraced inter-organizational business models have often taken the lead in leveraging large scale data analysis to support their rapidly changing value chains. Yet here-in lies the problem; even though "Big Data", "Data Mining", "Analytics", and "Business Intelligence" have been part of our professional and academic lexicon for many years, their underlying purpose and use is not very well understood.

3. Business Intelligence and Discovery Analytics: Two Distinct Views

Simply stated, business intelligence has many meanings, and they often depend on the domain in which are being described, and what type of data is being analyzed (Herschel and Jones, 2005), At its core there are two

very distinct views about "Big Data" that must be explained (Figure 1).

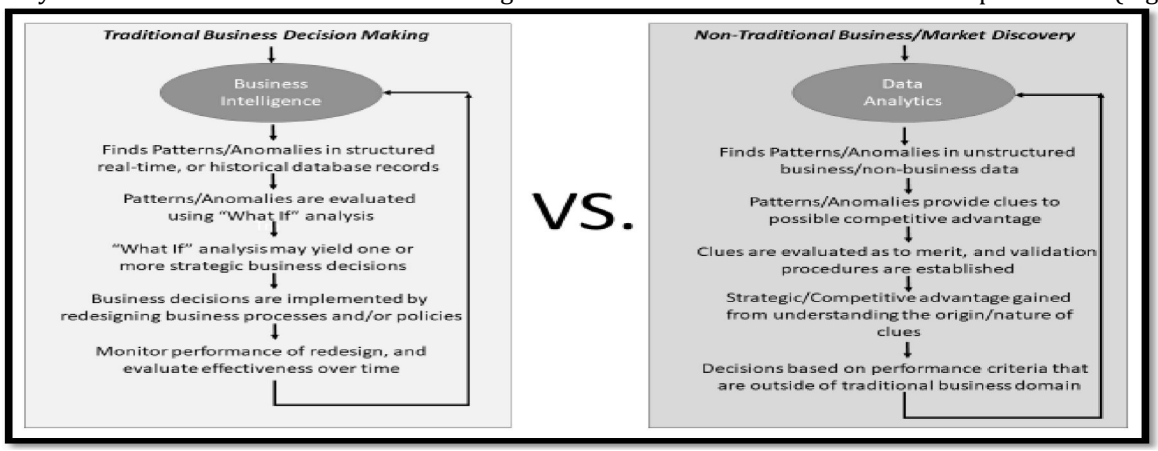


Figure 1

- Traditional Business Intelligence requires an organization to "know more about what they currently know".
- With Data Discovery organizations are looking for "things" they don't know, and may not have the questions "yet;" e.g., the discovery exercises generate "clues"

Since data can be organized into two distinct categories - Structured Data and Unstructured Data, in most enterprises structured data mostly represents traditional business data that is trusted and used to drive day-to-day operations; such as sales, accounting and human resource data. This data resides in very robust and secure relational databases and provides management the ability to operate on a continuing basis. As mentioned previously, unstructured data is ancillary to running the enterprise. It is a by-product of normal business transactions, and stored in the data-lake. Being unstructured, the data is generated from ancillary sources such as web-site clicks, bulletin boards, telephone calls, blogs, or forums – just to name a few. Since the data is not necessarily critical to daily business operations, much of it may never get used, but it is kept in the data lake for future use, or for specialized projects (Figure 2).

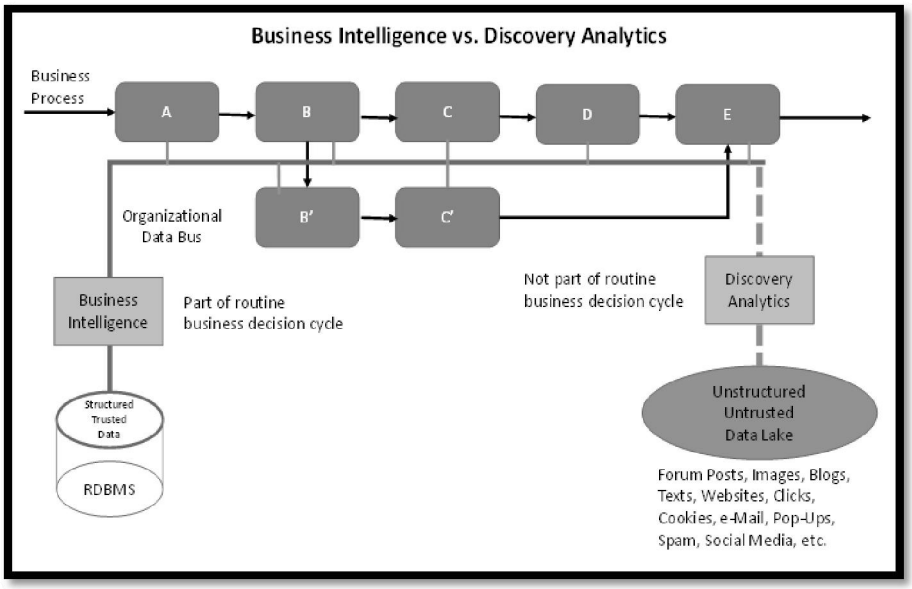


Figure 2

4. Finding a Competitive Advantage

4.1. Business Intelligence (BI)

From a historical business perspective, the analysis of structured data in making business decisions is a wellresearched and known quantity. Thousands of tools and numerous methodologies exist to help businesses

extract, transform and load (ETL) structured relational data into formats and data packages that can be distributed among any number of systems to develop an enterprise wide business solution. This is also the domain that is highly integrated around the concept of Business Intelligence (BI). Business Intelligence leverages the formal and well-defined ETL process to allow managers to pose and run "what-if" decision-making scenarios by extrapolating current trusted structured data forward into proposed future business cycles (Ranjan, 2009). Regardless of whether the future direction has to do with demand, profitability, outsourcing, mergers and acquisition, or any combination of other variables; the most relevant input data for the scenario usually comes from a trusted structured source. Hence BI concepts (and related tools, optimizers, rendering application platforms, etc.) are designed and optimized around a traditional and well understood relational database warehouse model (Chen, et al, 2012).

4.2. Discovery Analytics (DA)

Conversely, unstructured untrusted data sitting in a data lake can't necessarily be extracted, transformed, and loaded (ETL) into a relational table format. That doesn't mean it's impossible, but considering how much unstructured data can be generated on an hourly, daily, and monthly basis, it would be quite expensive to attempt such a project on a large scale. The better solution (mostly from a time and money perspective) would be to take manageable portions of the data lake and analyze that data for anomalies that could help support a business-related hypothesis, or point to a pattern of behavior that could possibly provide a competitive advantage in current or future business cycles. This then is the domain of Discovery Analytics (DA).

With Discovery Analytics, an organization is essentially placing an informed bet that there is a high probability that the expense associated with collecting and storing vast amounts of unstructured and untrusted data will yield a competitive advantage at some point in time. Essentially DA is used to mine vast portions of the data lake for randomly occurring patterns. The odds of finding a pattern that can actually provide a competitive advantage are relatively low – UNLESS, the data lake has the capacity to capture the smallest and most trivial of data. The reasoning is simple. The more data there is in the lake, the better the odds of finding a random pattern that could pay off and take the organization to the next level of efficiency and profitability (Agiu, et al, 2014)
Unlike structured data management paradigms that rely on traditional ETL methods to align critical elements of the business process, unstructured data is rather difficult to align with traditional data management architectures. To make the domain even more complicated, many of the terms and constructs used to define structured data management may have a different meaning when describing the dynamics of managing unstructured data. Therefore, business functions that rely on some form of unstructured data may first have to be mapped to a procedural model that defines the goal of the exercise (Figure 3).

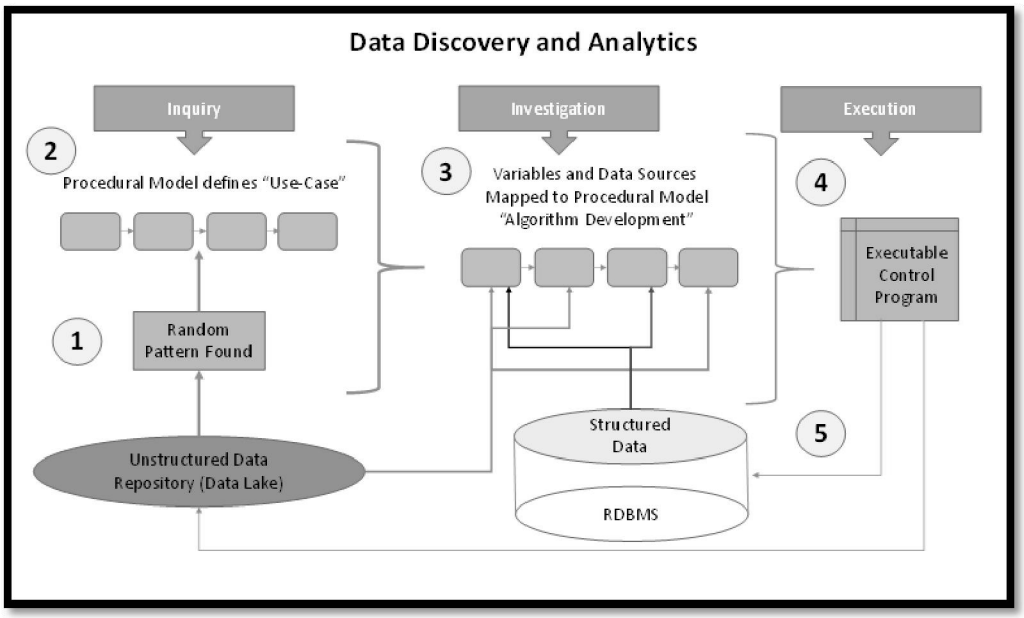


Figure 3

This is often referred to as building the "use-case" for the problem. For example,in the "Inquiry" phase a random heat-map pattern of cell phone usage generated from unstructured data (such a specific type ofimage

upload /download activity) may prompta law enforcement team to seek further investigation into human traffickingor gambling activities. Based on this evidence an “Investigation” phase may be called for, and theteam will build a use-case procedural model to help bound the project. The model will outline the specific steps and procedures the project team thinks will be required to successfully complete this part of the investigation. This use-casewill then serve as the basis for developing the “Execution” phase where a control program (an executable algorithm) is developed, and which defines the variables that must be included in supporting the data requirements of the procedural model. The procedural model may define the problem in any number of dimensions (i.e., cultural, ethnic, socio-economic, and local consumer preference, etc.).The variables to support these dimensions may consist of both structured and unstructured data, and their relationship(s) to each other are defined within the structure of the control program. The output of the execution phase control program may help to pin-point human trafficking and organized crime activities within a very precisely defined geographic boundary.

5. Big Data Technical Challenges

Needless to say, the amount of data generated in today’s global business environments are staggering. Inter- and intra-organizational processes generate vast amounts of structured and unstructured data. While structured trusted data can be culled and managed in relational databases, unstructured data is by definition “unruly” in its inability to be easily classified and categorized, and in the sheer volume in which it is generated. When collecting clicks, cookies, phrases, documents, images, texts, emails, spam, posts, etc. that are related to an organization’ s dealings with the world at large, the store and maintenance costs of data are enormous. That cost is often justified with the premise that the data “may” at some point in time reveal a game changing competitive advantage. Until that time comes, storage and management costs are steadily climbing. Not surprisingly, many senior managers are questioning the continued use of data analytics because they see relatively few returns from the investment. Research suggests that they are correct in their assessments, and points to the fact that many companies don’t know how to exploit the data already embedded in their core business systems.

6. Conclusion

The main issue with business intelligence and data analytics/discovery is that there are no hard and fast rules, particularly when it comes to the latter technology. It is still up to managers' expertise and judgment to decide which DA-generated patterns are promising (and so call for further research and investment) and others should be disregarded. "The art" of good management is not something that can be easily taught or developed, as is sometimes the case with enterprises. Warren Buffet would probably focus right away on a pattern that makes perfect sense to his superior investing instincts when presented with a sequence of DA outputs, while I would look at the same pattern and think, "So what?" Thus, the capacity to decipher the DA output and proactively create a successful business solution in response is the real competitive edge. One may argue that relatively few individuals possess the wisdom to profit from patterns created by DA. A "goldmine" solution from a DA platform may be offered, but it will stay elusive until its potential is acknowledged.

7. References

i. Mateescu, V., Agiu, D., & Muntean, J. (2014) Overview of Business Intelligence. Journal of Database Systems, 3, 23–36.

ii. The Human Aspect of CGMA® Report Business Analytics and Decision Making (2016). Chartered Management Accountants Institute. iii. T. Elliott (2003) Putting a Business Intelligence Strategy into Action! A Useful Manual for Business

White Paper on Intelligence Standardization, businessobjects.com IV. Knowledge management and business intelligence: the significance of integration (Herschel, R.Y., & Jones, N.E., 2005). Knowledge Management Journal, 9, 45–55.

V. In their 2012 book Business Intelligence and Analytics: From Big Data to Big Impact, Hsinchun, C., Chiang, R., and Storey, V.C. Quarterly for MIS, 36, 1165–1188. vi. Business Intelligence: Concepts, Components, Techniques, and Benefits, Ranjan, J. (2009). Applied and Theoretical Information Technology Journal, 9, 60–70. vii. In their 2014 book Business Analytics Principles, Concepts, and Applications: What, Why, and How, Schniederjans, M. J., Schniederjans, D. G., and Starkey, C. M. Pearson Education, Inc., Upper Saddle River, NJ. ix. R.A. Sommer, Public Sector Transformation: Comprehending Vertical and Horizontal Integration, Journal of

Business & Management International (IJBM), 5, 82–86. ix. In their 2003 book The Business Value of Business Intelligence, Williams, S., and Williams, N. Decision Path Consulting, Business Intelligence Journal,