MIS 500 – BUSINESS INTELLIGENCE

Fall - 2018/2019

Dr. Issam ElMoughrabi

PREPARED BY

Abdulrahman AlQallaf – 1601102047

Contents

The Rise of the Information Age

We are at a turning point in our history, a period characterized by a shift from an economy that is based on industry, to an economy that is primarily based on information. This increased digitization of our daily life coupled with an increased access to both computation power and storage capacity is leading humanity into the "Information Age". And to illustrate the tremendous growth in the global size of data, the following two charts show the "Annual Size of the Global Datasphere" along with the units of measurements for reference[1].

Figure 1 – Annual Size of the Global Datasphere

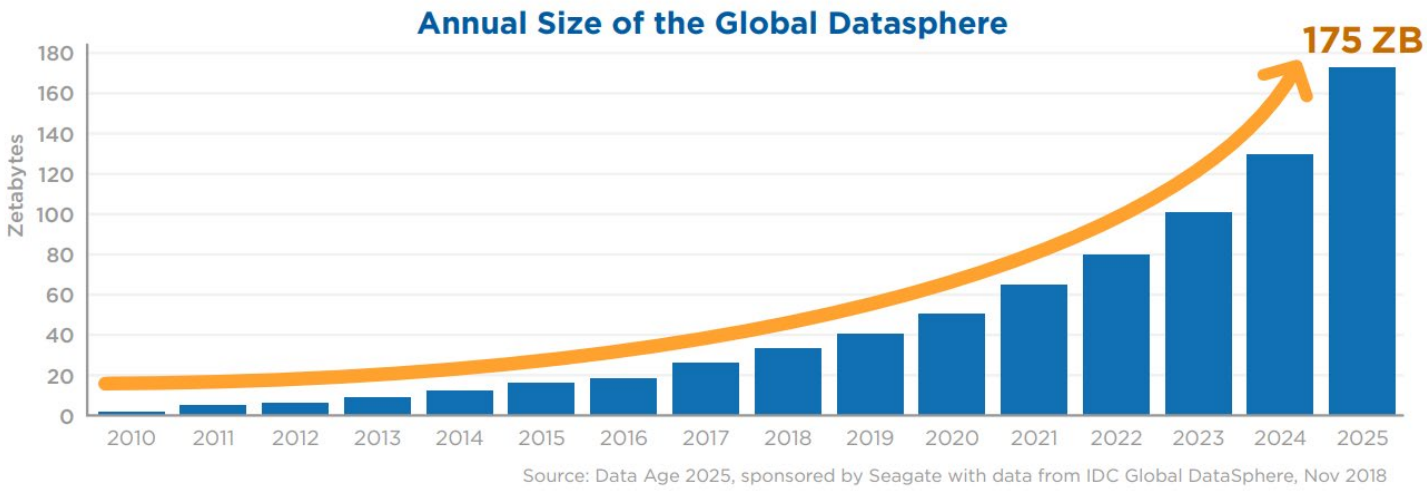


Table 1: Data Measurement Units

Unit	Abbreviation	Decimal Value	Binary Value	Decimal Size
bit	b	0 or 1	0 or 1	1/8 of a byte
byte	B	8 bits	8 bits	1 byte
kilobyte	KB	$1,000^1$ bytes	$1,024^1$ bytes	1,000 bytes
megabyte	MB	$1,000^2$ bytes	$1,024^2$ bytes	1,000,000 bytes
gigabyte	GB	$1,000^3$ bytes	$1,024^3$ bytes	1,000,000,000 bytes
terabyte	TB	$1,000^4$ bytes	$1,024^4$ bytes	1,000,000,000,000 bytes
petabyte	PB	$1,000^5$ bytes	$1,024^5$ bytes	1,000,000,000,000,000 bytes
exabyte	EB	$1,000^6$ bytes	$1,024^6$ bytes	1,000,000,000,000,000,000 bytes
zettabyte	ZB	$1,000^7$ bytes	$1,024^7$ bytes	1,000,000,000,000,000,000,000 bytes
yottabyte	YB	$1,000^8$ bytes	$1,024^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

[1] Source: http://www.itutility.net/data-measurement-abbreviations-refresher/

Business Intelligence – An Overview

In general, unorganized raw data is not useful for people to consume and benefit from. Therefore, to properly manage data and to maximize the benefit extracted from it, what we need to have is an information system that provides users with the following:

- Accurate information (free from errors)
- That is timely (available when needed)
- And relevant (useful for what is required)

This will allow us to move from:

1. Raw data (fit for computer application consumption)
2. To readable information (fit for human consumption)
3. To knowledge (developing an understanding of existing information to extrapolate / arrive at new information)

There are many businesses today that rely heavily on information in order to operate, such as technology companies (e.g., Google and Facebook), Telecoms, and Banking. In order for these businesses to maximize their benefit from data, they need what is normally referred to as "Business intelligence" (or BI for short). As a high-level simplification, BI entails the following three steps:

1. Collect relevant data and transform it into information
2. Analyze information and come up with new knowledge
3. Deliver information and knowledge to end users

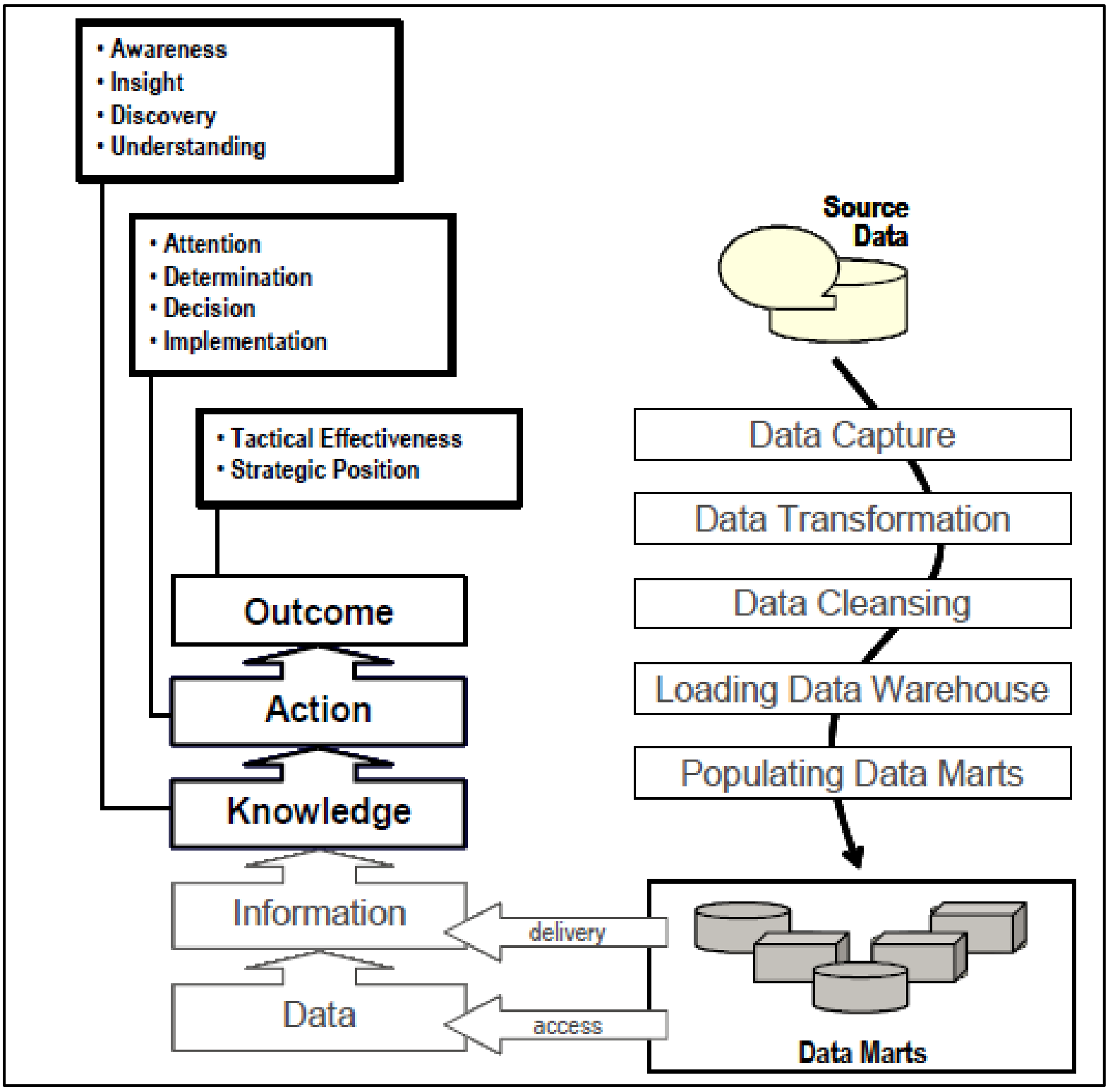
In detail, Business intelligence is a term used to describe the infrastructure for warehousing, integrating, reporting, and analyzing data that comes from the business environment and from external sources. This foundational infrastructure collects, stores, cleans, and makes relevant information available to users all over the organization.

As you can see from the following illustration[2], the first part of any BI infrastructure is the data warehouse. It collects and stores current and historical data from various sources and makes it available for users to access as needed (data is not altered once collected). Companies often build enterprise-wide data warehouses, where a central data warehouse serves the entire organization.

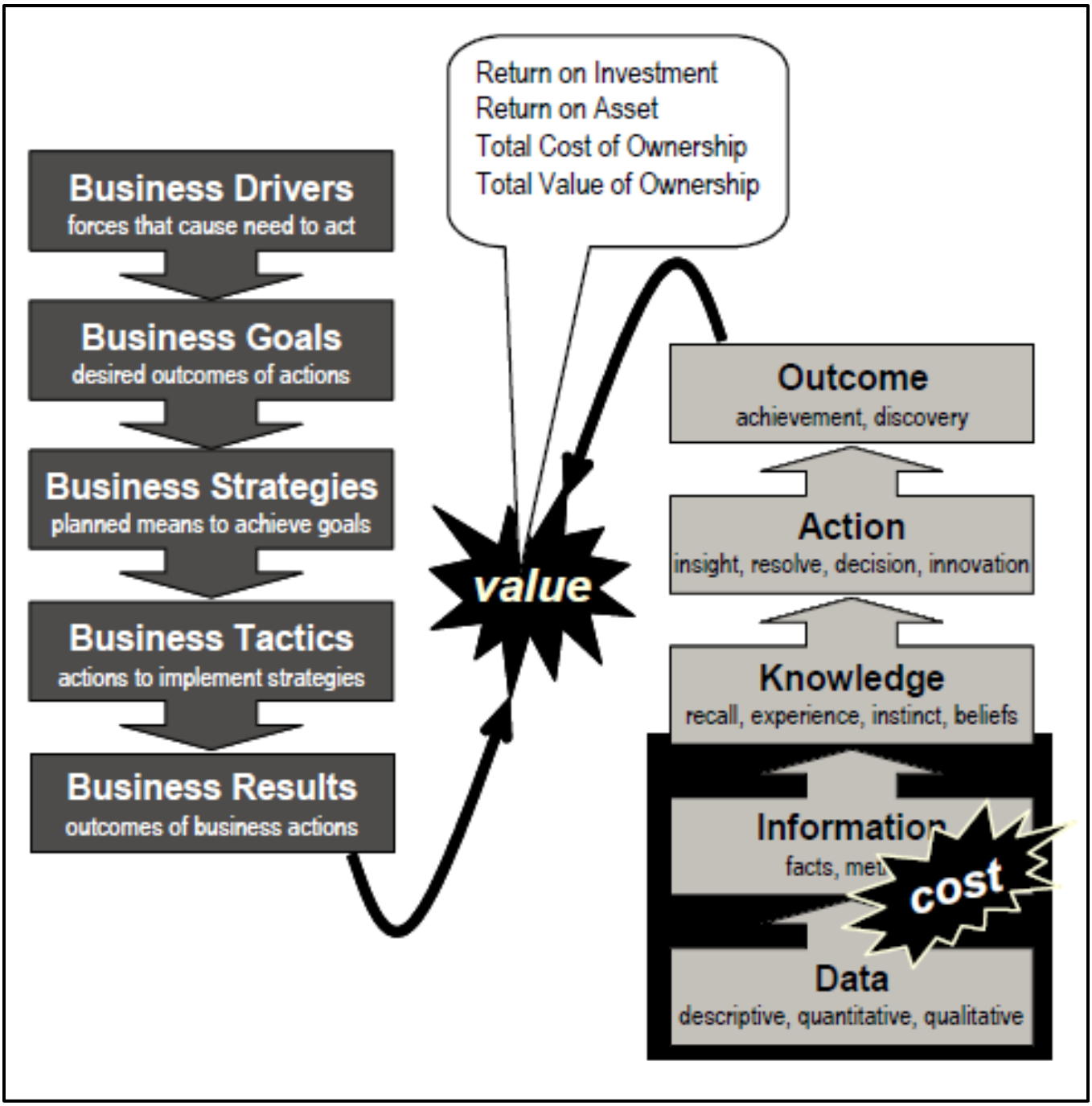
Once data is collected and organized in a data warehouse, a data mart (which is a subset of a data warehouse) can be created. A data mart serves as a summarized or highly focused portion of the organization's data for a specific population of users.

The next step would be to analyze the data available to come up with actionable knowledge in order to achieve business outcomes. This is where data mining and analytics come in (also referred to as data science).

[2] *Source: The Data Warehousing Institute (TDWI) – Business Intelligence Executive Briefing*

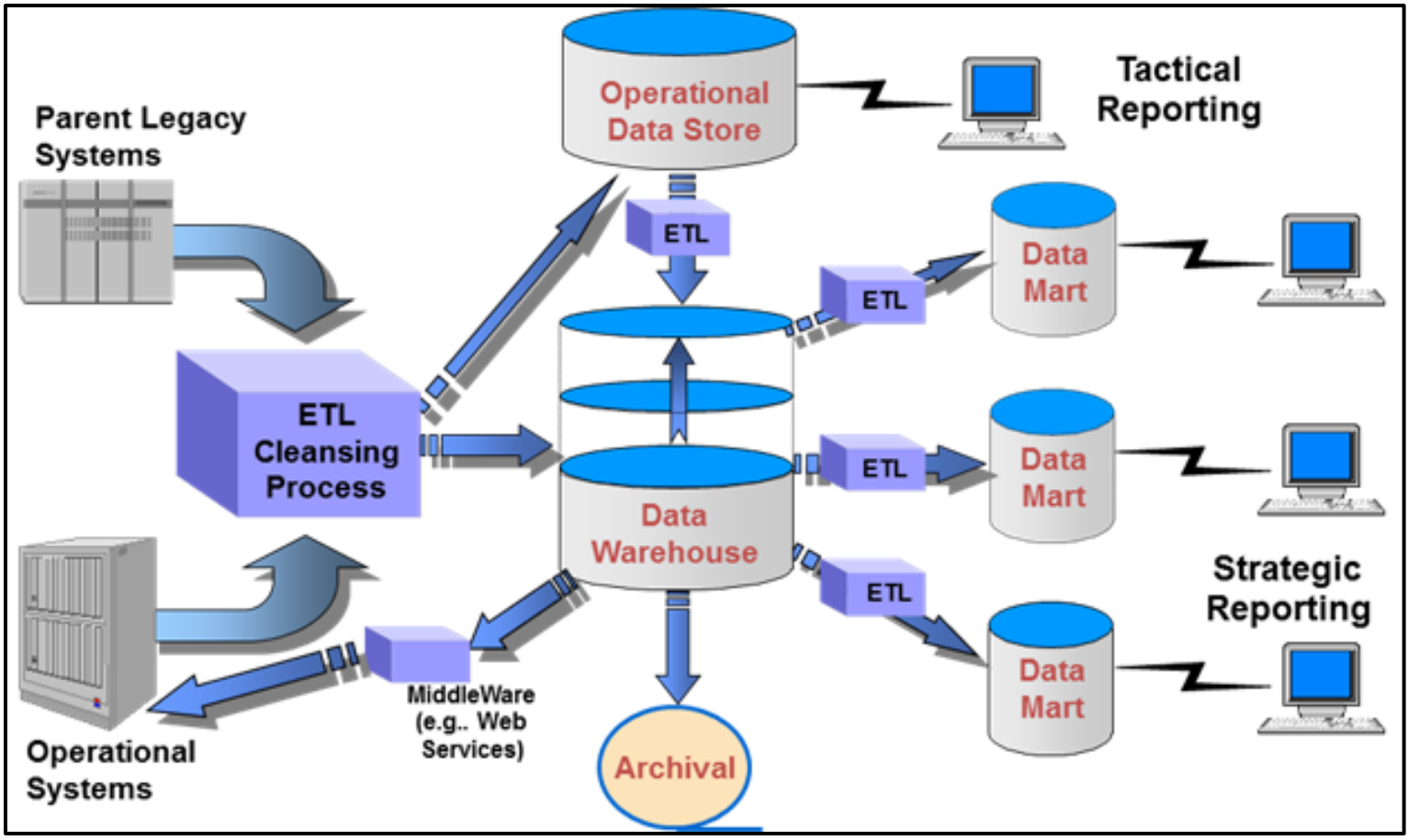

• Awareness
• Insight
• Discovery
• Understanding
• Attention
• Determination
• Decision
• Implementation
• Tactical Effectiveness
• Strategic Position
Outcome
Action
Knowledge
Information
Data
delivery
access
Source Data
Data Capture
Data Transformation
Data Cleansing
Loading Data Warehouse
Populating Data Marts
Data Marts

All of these core BI activities (collection, analysis, and delivery) increase the cost of operating a business (normally referred to in the work environment as "cost centers"). What makes this additional cost worthwhile is that, if done properly, it will bring in much more value to the business than it costs. The added value of BI activities cannot be properly quantified unless they are paired with the resulting action and outcome that is achieved from following through on the knowledge produced by BI (see this illustration for how BI activities can generate value to the business[3]).



In the following three sections, I will provide more details on each core part of the BI flow (collection, analysis, and delivery).

[3] *Source: The Data Warehousing Institute (TDWI) – Business Intelligence Executive Briefing*



4

A data warehouse (DW for short) is a system used for reporting and data analysis, it is considered a core component of business intelligence. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place that are used for creating analytical reports for workers throughout the enterprise[5].

Characteristics of a DW are as follows[6]:

1. Subject-oriented
   1.1. Operational data is organized by specific processes or tasks and is maintained by separate systems
   1.2. Data warehouses are organized by subject area and are populated from various operational systems
2. Integrated
   2.1. Data is gathered from the operational systems and is combined according to business requirements / rules and cleansed
3. Non-Volatile
   3.1. Data warehouse data is not directly updated by the end-user
   3.2. Even for "real time" DW's – updates occur through an ETL process – not directly by the user

[4] Source: https://www.ewsolutions.com/foundations-data-warehousing/
[5] Definition from Wikipedia (https://en.wikipedia.org/wiki/Data_warehouse)
[6] Source: https://www.ewsolutions.com/foundations-data-warehousing/

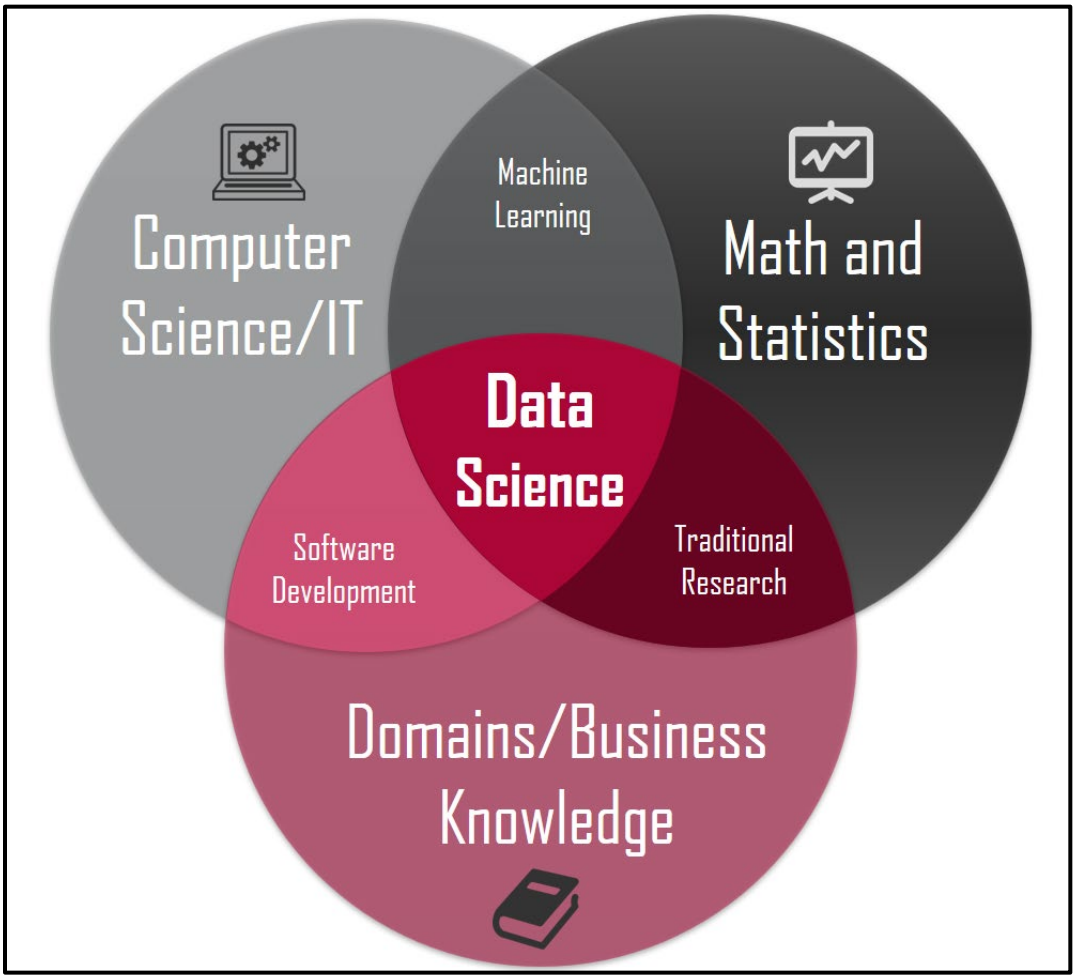
4. Time Variant
    4.1. Time variant doesn't mean that the data is never changed
    4.2. Business changes sometimes dictate changes to the data in the data warehouse (e.g. region reorganization, change in accounting procedure, etc.)
    4.3. Data in the data warehouse includes a time stamp to indicate its operational date and its inclusion date is recorded, along with any change date

Although each organization will have its own, specific, goals for a data warehouse, there are certain common goals which include:

- To enable users appropriate access to a homogenized and comprehensive view of the organization, supporting forecasting and decision-making processes at the enterprise
- To achieve information consistency. By bringing data from disparate data sources into a centralized database, the data warehouse provides a homogenized view of the organization's data. Users from across the organization making use of the data warehouse all view a single and consistent version of the enterprise's information that is available for analysis and reporting, at tactical and strategic decision-making levels.

In the end, all data warehouse goals should be aligned with the organization's business goals, so that business value of the data warehouse's development and maintenance can be justified and supported.

Business Intelligence – Information Analysis

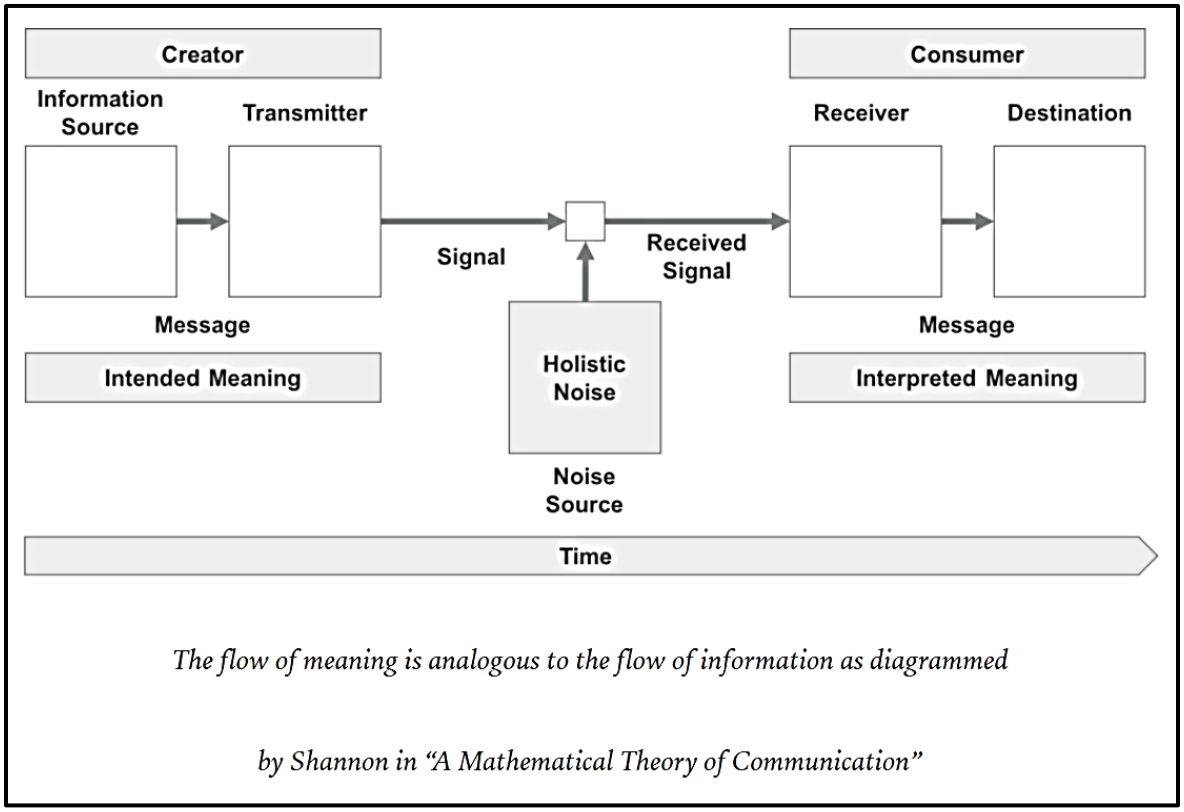


Information analysis within organizations takes on many names, one of which is business analytics. Business analytics utilizes an organization's data warehouse, and refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning[7]. Most recently however, the focus shifted from traditional business analytics towards data mining and analytics (commonly referred to nowadays as "data science"). Data science is an interdisciplinary field[8] that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured[9].

Data science is more discovery-driven. It provides insights into an organization's data by finding hidden patterns and relationships in large data sets and inferring rules from them to predict future behavior. The patterns and rules are used to guide decision making and forecast the effect of those decisions. The types of information obtainable from data mining include associations, sequences, classifications, clusters, and forecasts.

- Associations: occurrences linked to a single event
- Sequences: events that are linked over time
- Classification: recognizes patterns that describe the group to which an item belongs to by examining existing items that have been classified and by inferring a set of rules
- Clustering: works in a manner similar to classification when no groups have yet been defined
- Forecasting: uses a series of existing values to forecast what other (new) values will be

Business Intelligence – Delivery of Information and Knowledge



The flow of meaning is analogous to the flow of information as diagrammed

by Shannon in "A Mathematical Theory of Communication"

10

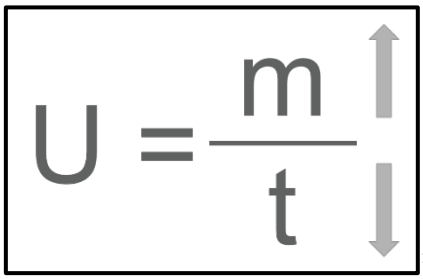
[7] Source: https://en.wikipedia.org/wiki/Business_analytics
[8] Source: https://towardsdatascience.com/introduction-to-statistics-e9d72d818745
[9] Source: https://en.wikipedia.org/wiki/Data_science
[10] Source: https://loupventures.com/defining-the-future-of-human-information-consumption/

Information (data that is readable by a human being) and knowledge (arrived at by analyzing information) is created to be consumed by some end user, this process of creation and consumption generally follows the below steps:

1. Information and knowledge is produced by the creator with an "intended meaning".
2. This "intended meaning" can lose some of its essence while in transmit to be consumed by the end user (this is what is shown above as the "holistic noise"). A decrease in holistic noise should mean a closer interpretation of the message by the consumer relative to the creator's intent. This distortion in the transfer of meaning from the one side to the other can be the result of the following factors:
   2.1. Lack of clarity in the message from the creator
   2.2. Inability to understand the message by the consumer
   2.3. Misinterpretation by the consumer
3. Information and knowledge reaches the end user, and the end user forms an "interpreted meaning" of the original "intended message"

The amount of utility (i.e., usefulness / meaning) the end user derives from the provided information / knowledge is paramount to the success of any organization. This is because the end user is the person that will act based on the provided information / knowledge in order to achieve the organization's outcomes. The amount of utility the end user derives from a provided information / knowledge can be measured using the below equation. Which translates into the following: Utility (U) of information is equal to the value of the meaning interpreted by the consumer of the information (m) divided by the time it takes to consume the information (t). This entails that the creator of information / knowledge needs to be aware of the following:

- The more meaning that the end user receives from the provided information / knowledge the better
- The less time that it takes the end user to consume the provided information / knowledge the better

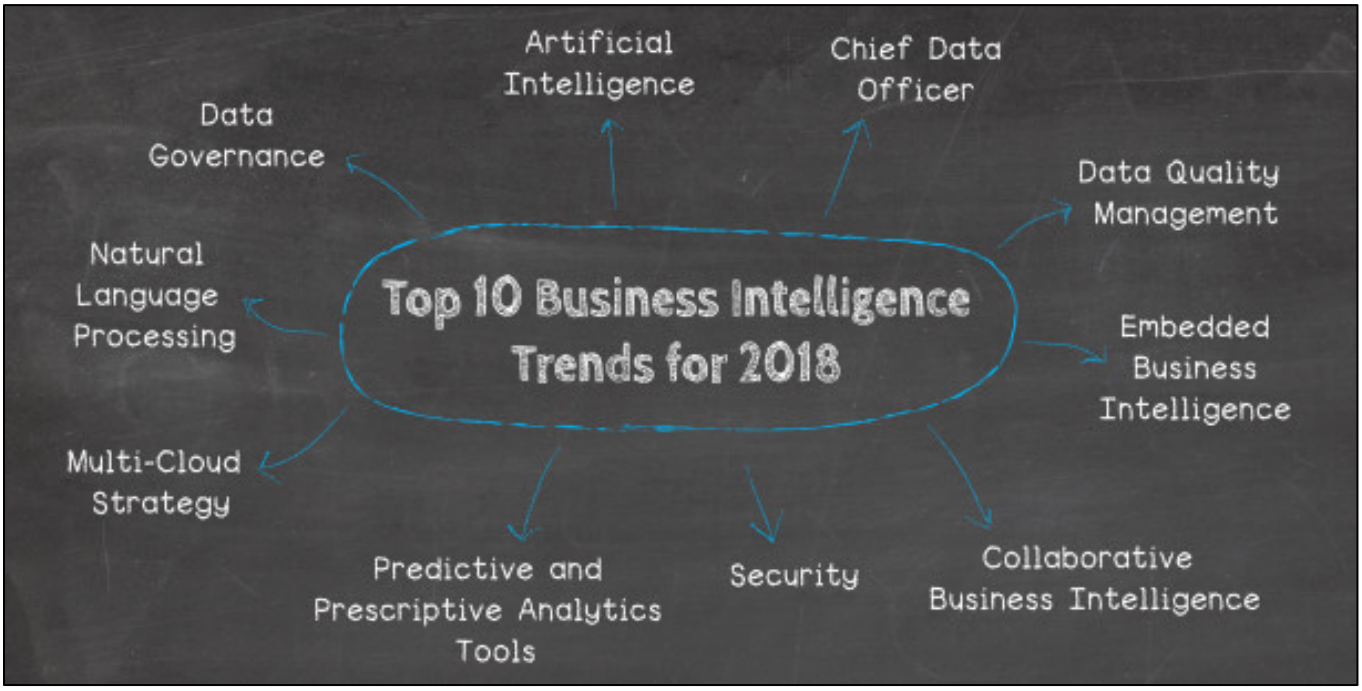


[11]

There are of course many formats in which information / knowledge can be delivered to an end user, but in all cases, the creator of information / knowledge should keep in mind the above-mentioned points in order to successfully deliver his message. Some of the formats in which a message can be delivered are the following:

- Tabular reports
- Parameterized reports (allow users to specify one or more parameters to customize the report)
- Reports that can be manipulated by end user (e.g., Excel pivot tables)
- Visual dashboards
- Images (e.g., infographic)
- Videos

[11] Source: https://loupventures.com/defining-the-future-of-human-information-consumption/



1) Artificial Intelligence

Artificial intelligence (AI) is the science aiming to make machines execute what is usually done by complex human intelligence. We are evolving from static, passive reports of things that have already happened to proactive analytics with live dashboards helping businesses to see what is happening at every second and give alerts when something is not how it should be.

2) Predictive and Prescriptive Analytics Tools

Business analytics of tomorrow is focused on the future and tries to answer the questions: what will happen? How can we make it happen? Accordingly, predictive and prescriptive analytics are by far the most discussed analytics trends among the BI professionals.

Predictive analytics is the practice of extracting information from existing data sets in order to forecast future probabilities. It's an extension of data mining which refers only to past data. Predictive analytics includes estimated future data and therefore always includes the possibility of errors from its definition. Applied to business, predictive analytics is used to analyze current data and historical facts in order to better understand customers, products and partners and to identify potential risks and opportunities for a company.

Prescriptive analytics goes a step further into the future. It examines data or content to determine what decisions should be made and what steps taken to achieve an intended goal. It is characterized by techniques such as graph analysis, simulation, complex event processing, neural networks, recommendation engines, heuristics, and machine learning. Prescriptive analytics tries to see what the effect of future decisions will be in order to adjust the decisions before they are actually made. This improves decision-making, as future outcomes are taken into

[12] **Source:** https://www.datapine.com/blog/business-intelligence-trends/

consideration in the prediction. Prescriptive analytics can help you optimize scheduling, production, inventory and supply chain design to deliver what your customers want in the most optimized way.

3) Natural Language Processing

This trend is highly related to the previous business intelligence trends mentioned. Natural Language Processing, NLP or computational linguistics, is a branch of artificial intelligence related to the understanding of human language(s) by a computer program. It is based on linguistics and deep learning, a type of AI that works with pattern recognition to improve the program's understanding by analyzing massive amount of data to find correlations that are relevant. Deep learning is a more intuitive and flexible approach that learns how to identify a speaker's intents, a bit like when a child learns how to speak with his/her environment.

The technology underlying this human-computer relationship and understanding is changing our society dramatically. We can already see the applications with virtual assistants like Siri, Cortana or Alexa, or the incredible development of customer service chatbots that can help and answer clients more accurately every day.

That will also transform the way we do business intelligence in a future nearer than we probably imagine. Not only will the interface be changed, but the way we interact with it. Imagine you are no data scientist, nor an IT professional, but you need to work on your humongous amounts of data gathered on many different databases that are centralized in one place. And you can do that by simply asking the software your business questions, orally. Like you would ask your colleague. Like you would ask yourself before digging in, going through, drilling down the mountains of data, finding the valuable information, organizing it, and then visualizing it thanks to modern dashboards. With a simple oral interaction with a BI chatbot, all this time and efforts would be spared, might also be less biased and more accurate than with a human.

Applying natural language processing to business intelligence would let you spend more time on more critical tasks where the human cannot (yet?) be removed, like actually asking the right data analysis questions, or elaborating the company's business intelligence strategy. With NLP, you won't just ask the question in natural language, you will also receive the response in natural language.

However, several challenges remain for NLP and advances made will help business analyze and learn from a greater range of data sources, at a faster pace, and thus increase productivity and give them a competitive advantage.

4) Data Quality Management (DQM)

DQM consist of acquiring the data, implementing advanced data processes, distributing the data effectively and managing oversight data.

5) The Multi-Cloud Strategy

There are more and more organizations moving their data and all of their applications to the cloud. Gartner states that by 2019, the cloud will be the common strategy for 70% of the companies – while it was less than 10% in 2016. When evaluating hosting environment, you take risk, speed, costs and complexity into account, which makes it even harder to pick one solution fitting all your needs.

6) Data Governance

According to the DGI (Data Governance Institute), data governance is "the exercise of decision-making and authority for data-related matters." In other words, it is the control over any data entry that has to be made accordingly to particular standards. Data access, and security issues don't all deal with data breaches. In 2018, organizations will increase focus on data governance and data quality. As data is only useful when accessible, organizations will increasingly look to strike a balance between data access and security. They also must learn to remain agile and adapt as the business changes.

New data preparation tools and methods will help fuel this trend and decrease the cultural gap between business and technology. Organizations are learning that data governance can help nurture a culture of analytics and

meet business needs. Also, people are more likely to dig into their data when they have centralized, clean, and fast data sources.

7) Security

8) Growing Importance of the CDO

We can safely affirm that today, data and analytics are becoming core to every business. Every company has had a Chief Information Officer who supervised all the information management assets and security issues. But today, the data and analytics volume and role are getting so big that a new position emerged: the CDO, or Chief Data Officer, assigned to data management only. Gartner even says that 90% of large companies will have a CDO role by 2019.

A Chief Data Officer's role is to create a system developed in a way that data can be leveraged across any business units within a company. A CDO needs to empower all the users with trusted, clean, and ready-to-use data. They have to ensure that value can be extracted and are hence outcome-focused.

The role of a CDO is essential for the good management of the information assets a company has, as well as to improve efficiency in the data analyses to get a competitive advantage in their BI strategy.

9) Embedded Business Intelligence

Embedded BI refers to the integration of a BI tool or selected features, into another business application to fill the gaps in the application's analytics or reporting functionality. Thanks to it, you can turn raw data into interactive dashboards, enhancing the user experience with real-time analytics and innovative data visualizations, enabling people to make data-driven decisions faster and on their own.

These capabilities may be located outside of the application, but they have to be easily accessible from inside the app so that the user doesn't have to alternate between systems and become accustomed to another user interface and structure. This way, embedded dashboards add features that are usually specific to BI software, enriching the application and making things simple for the user who also won't need to install or adapt to a new tool. The time between collection of data and analysis of it is also shortened.

10) Collaborative Business Intelligence

Today, managers and workers need to interact differently as they face an always-more competitive environment. More and more we see a new kind of business intelligence rising: the collaborative BI. It is a combination of collaboration tools, including social media and other 2.0 technologies, with online BI tools. This is developed in a context of enhanced collaboration addressing the new challenges the fast-track business provides, where more analyses is done and reports edited. When talking about collaborative BI, the term "self-service BI" quickly pops up in the sense that those self-service BI tools do not require an IT team to access, interpret and understand all the data.

These BI tools make the sharing easier in generating automated reports that can be scheduled at specific times and to specific people. For instance; they enable you to set up intelligent alerts, share public or embedded dashboards with a flexible level of interactivity. All these possibilities are accessible on all devices which enhances the decision-making and problem-solving processes.

External Forces Driving the Need for an Enhanced BI

Banking has been a prolific industry for innovation concerning information systems and technologies[13]. BI, more specifically, is showing increased significance within the banking sector. This is partly because of the following external factors:

1) The global financial crises of 2007-2009

Managing risks is key to the survival and flourishing of any financial institution. The financial industry learned many lessons from the recent crises and have come up with the following two standards: Basel III and IFRS 9.

Basel III is a set of international banking regulations developed by the Bank for International Settlements in order to promote stability in the international financial system. The purpose of Basel III is to reduce the ability of banks to damage the economy by taking on excess risk.[14]

IFRS 9 is an International Financial Reporting Standard (IFRS) promulgated by the International Accounting Standards Board (IASB). It addresses the accounting for financial instruments. It contains three main topics: classification and measurement of financial instruments, impairment of financial assets and hedge accounting. It will replace the earlier IFRS for financial instruments, IAS 39, when it becomes effective in 2018. However, early adoption is allowed[15].

Many banks initially approached IFRS convergence as an accounting challenge first, then as a risk challenge. They did not consider the significant role played by IT systems, processes and infrastructure. In general, the impact on IT from IFRS conversions creates a need for additional data, revised calculation tables and a new governance framework. To facilitate these challenges, IT will need to modify, remap, reconfigure or even start anew with different systems. Here are the four major IT challenges facing banks as they prepare for IFRS 9[16].

2) Data governance issues (GDPR)

The General Data Protection Regulation ("GDPR") is a regulation in EU law on data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA). It also addresses the export of personal data outside the EU and EEA areas. The GDPR aims primarily to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU[17].

The financial sector is one of the more highly regulated industries, but many banks have nonetheless been thrown off by the complexity of the EU General Data Protection Regulation (GDPR). The Regulation, which takes effect on 25 May 2018, overhauls the way organizations handle personal data. It includes countless requirements, we will outline three essential steps that banks can take[18]:

1. Documenting a lawful basis for processing
2. Hire a data protection officer (DPO)
3. Prepare for the right to data portability

[13] Source: Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation
[14] Source: https://www.investopedia.com/articles/economics/10/understanding-basel-3-regulations.asp
[15] Source: https://en.wikipedia.org/wiki/IFRS_9
[16] Source: https://www.sas.com/en_my/insights/articles/risk-fraud/better-it-and-data-practices-can-prepare-banks-for-ifrs-9-convergence.html
[17] Source: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
[18] Source: https://www.itgovernance.eu/blog/en/how-banks-should-prepare-for-the-gdpr

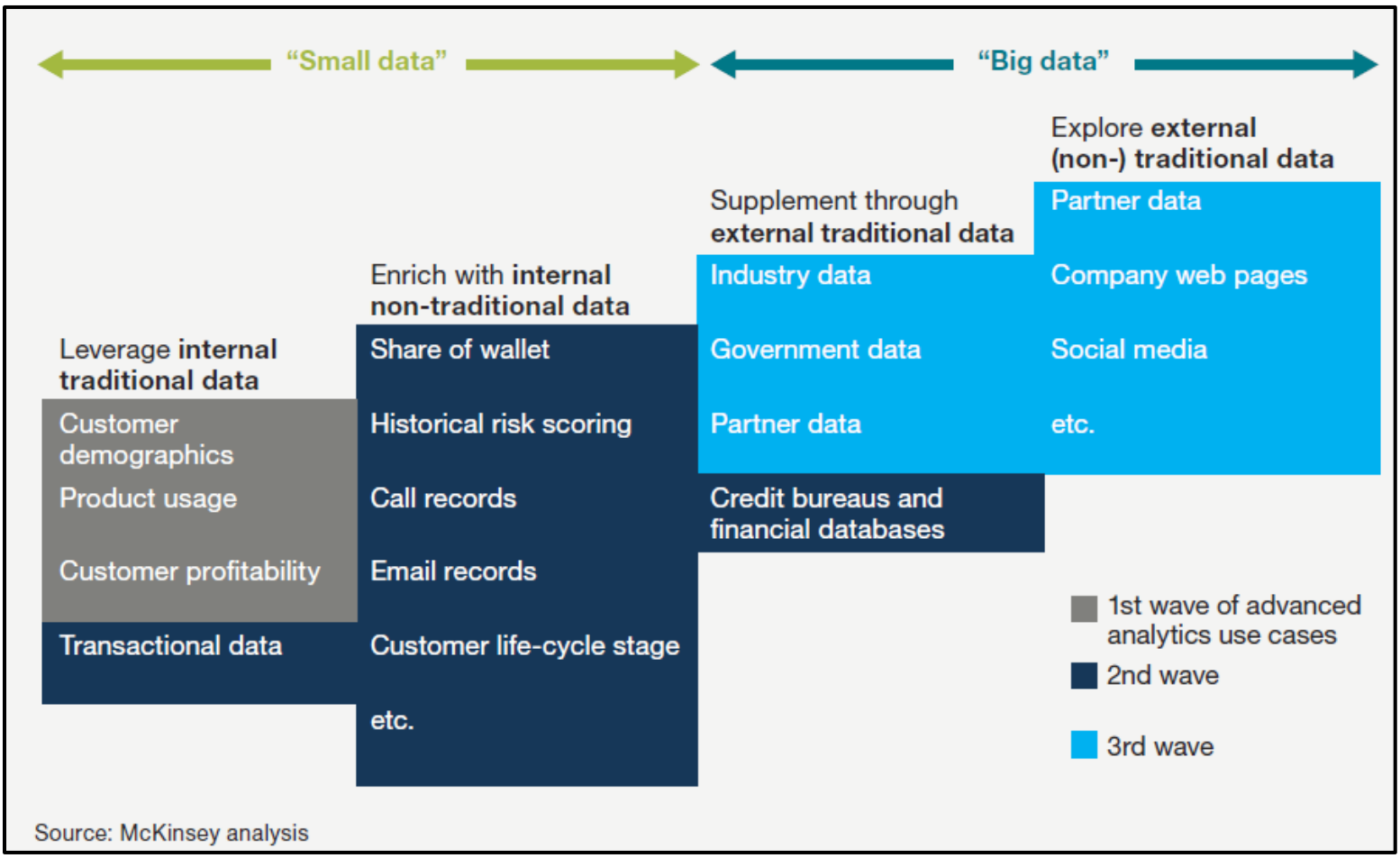
3) Foreign Account Tax Compliance Act (FATCA)

The Foreign Account Tax Compliance Act (FATCA) is designed to combat tax evasion by U.S. persons holding investments in offshore accounts. There are two aspects to FATCA[19]:

1. Foreign financial assets: certain U.S. taxpayers holding specified financial assets outside the United States must report those assets to the IRS.
2. Reporting by foreign financial institutions: FATCA requires foreign financial institutions to report directly to the IRS certain information about financial accounts held by U.S. taxpayers, or by foreign entities in which U.S. taxpayers hold a substantial ownership interest.

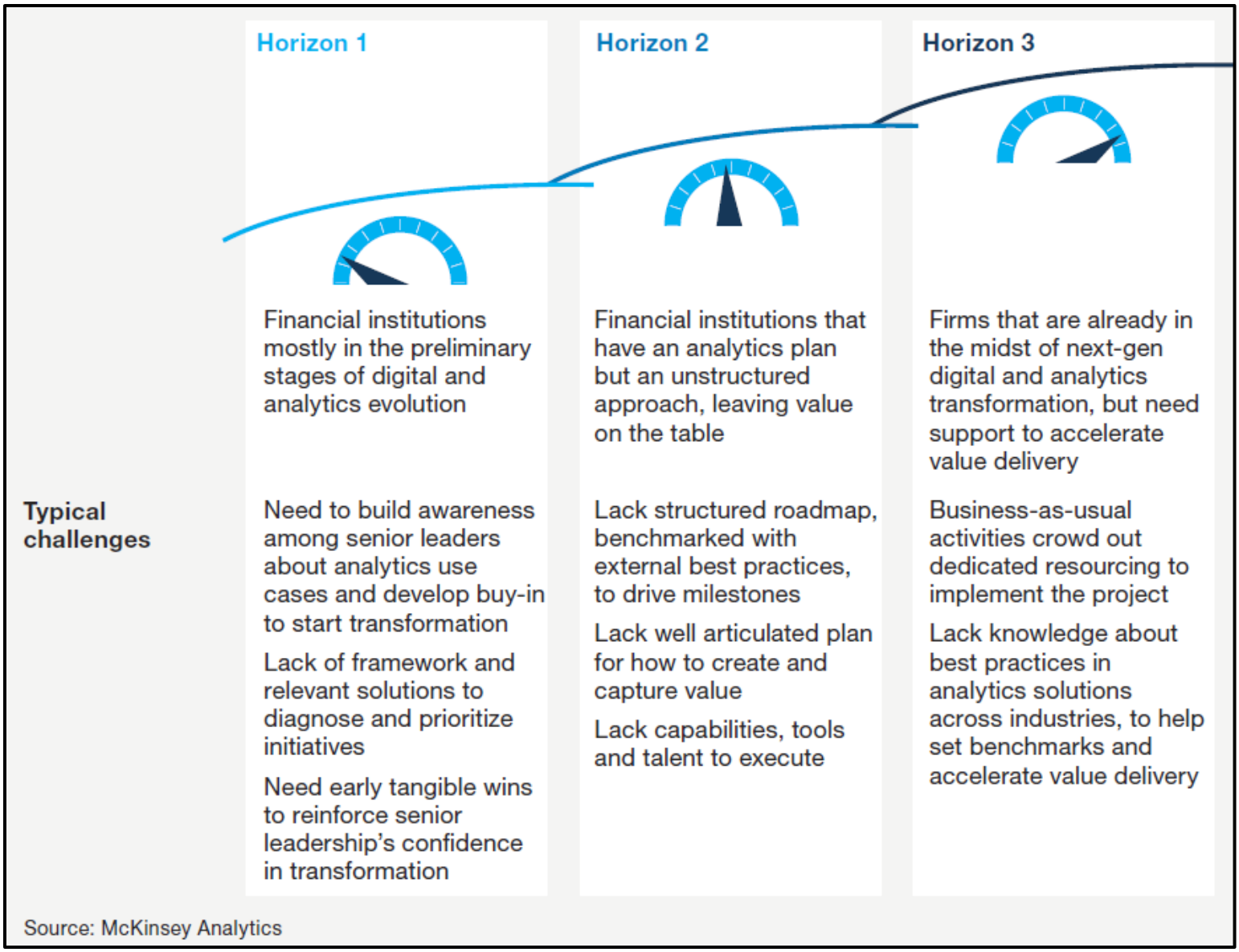
BI Within Banks, A Closer Look[20]

In a financial institution, data are collected from various sources and are consolidated in successive waves of increasing complexity, from "small data" such as customer transactions to "big data" such as customer social media interactions with the bank.



[19] Source: http://www.taxamerican.com/fatca.html
[20] Source: McKinsey on Payments, Special Edition on Advanced Analytics in Banking

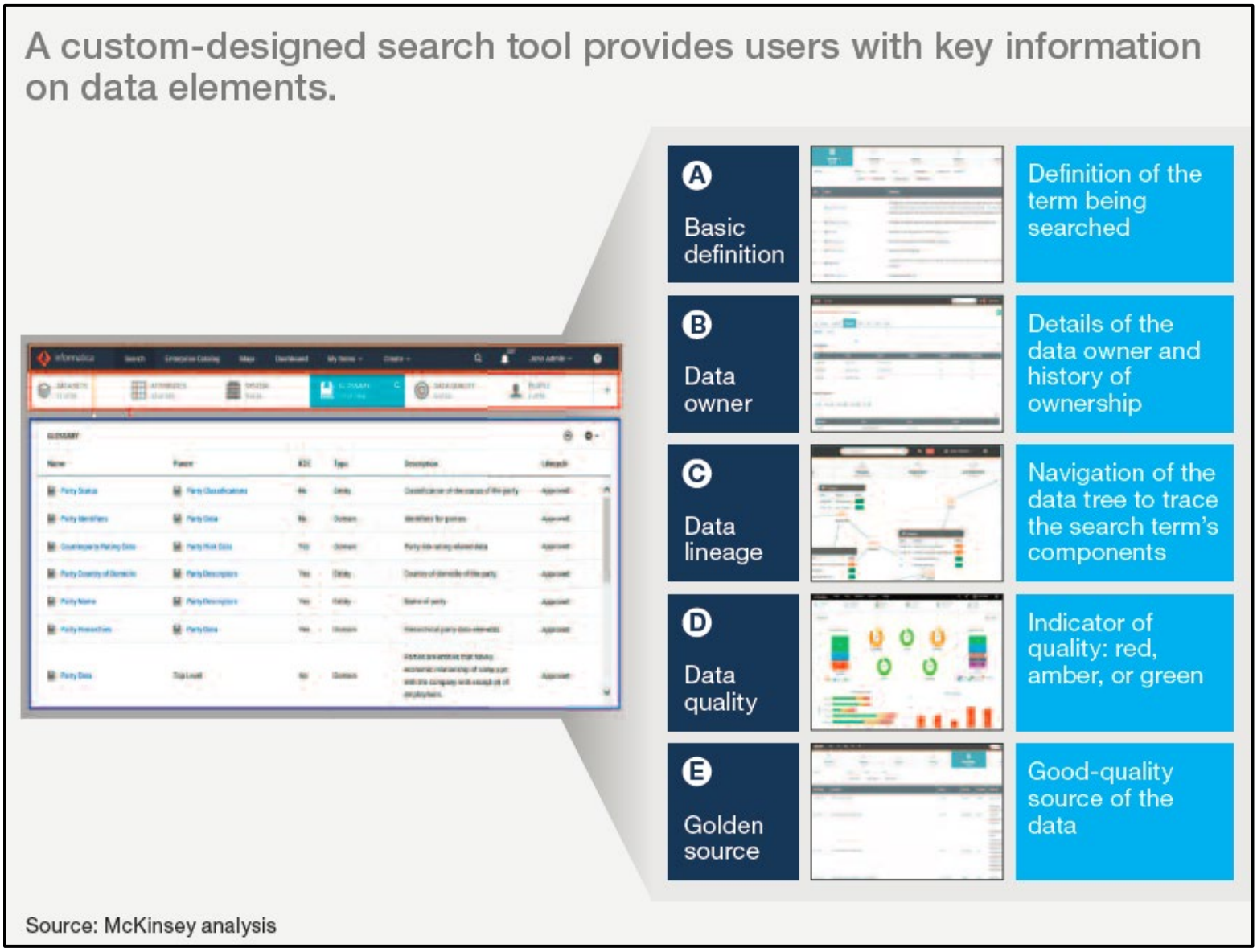
When looking at financial institutions in general, most financial institutions are situated on one of the three following horizons.



And when analyzing the typical challenges faced in improving data quality at the enterprise level, we see the following reasons ranked by importance:

1. Lack of front office controls (e.g., poor quality of data entry at system of origin with no / limited validation)
2. Inefficient data architecture (e.g., multiple data warehouses with no common data model, legacy systems, complex lineage)
3. Lack of business buy-in for value of data transformation
4. Data doesn't get enough board and senior management attention (e.g., seen as an IT issue, not considered a business asset)

Moreover, one of the key issues that banks face regularly is related to data management. The following illustration shows a sample solution to such an issue:



BI Within Banks, Towards Successful Implementation

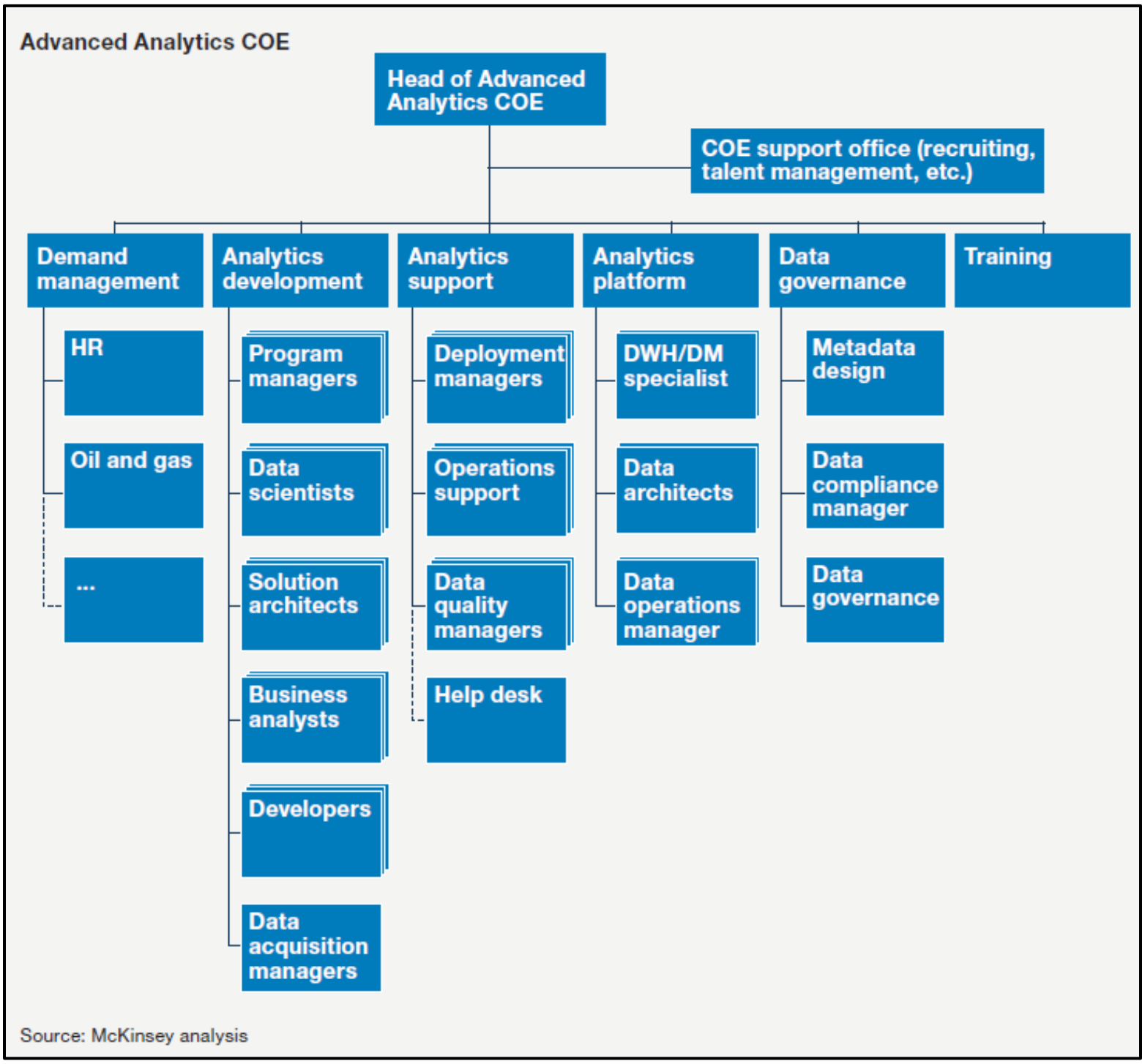
It is difficult to implement a state-of-the-art, enterprise level BI within a financial institution and have the whole organization understand the importance of data. A study in 2017 was conducted and reached that the following are the factors that determine BI implementation in organizations[21]:

- User satisfaction
- Cost reduction
- Time optimization

[21] Source: Factors to Determine Business Intelligence Implementation in Organizations, By Michael S. Moreno Saavedra, and Christian Bach

The evidence from the study points towards the idea that to adopt the business intelligence approach in an enterprise, it has to have a great significance on the business processes staring from improving personnel satisfaction in their work, cutting down the budget cost of the firm, and completing quicker responses to customers, executives, and suppliers.

In the end, the following chart shows an organizational blueprint of the advanced analytics Center Of Excellence (COE).



References

1. Management Information Systems – managing the digital firm – 15th edition, chapter 6
2. Management Information Systems – managing the digital firm – 15th edition, chapter 12
3. TDW, business intelligence executive briefing
4. Top 10 Analytics And Business Intelligence Trends for 2019 (https://www.datapine.com/blog/business-intelligence-trends/)
5. McKinsey on Payments Special Edition on Advanced Analytics in Banking
6. Factors to Determine Business Intelligence Implementation in Organizations, By Michael S. Moreno Saavedra, and Christian Bach