6

Essential Aspects of Business Intelligence

The essential aspects of business intelligence are context analysis, business performance management, business process discovery, information system, organization intelligence and process mining. The method to analyze the environment of any business is known as context analysis. The topics discussed in this section are of great importance to broaden the existing knowledge on business intelligence.

Context Analysis

Context analysis is a method to analyze the environment in which a business operates. Environmental scanning mainly focuses on the macro environment of a business. But context analysis considers the entire environment of a business, its internal and external environment. This is an important aspect of business planning. One kind of context analysis, called SWOT analysis, allows the business to gain an insight into their strengths and weaknesses and also the opportunities and threats posed by the market within which they operate. The main goal of a context analysis, SWOT or otherwise, is to analyze the environment in order to develop a strategic plan of action for the business.

Context analysis also refers to a method of sociological analysis associated with Scheflen (1963) which believes that 'a given act, be it a glance at [another] person, a shift in posture, or a remark about the weather, has no intrinsic meaning. Such acts can only be understood when taken in relation to one another.' (Kendon, 1990: 16). This is not discussed here; only Context Analysis in the business sense is.

Define Market or Subject

The first step of the method is to define a particular market (or subject) one wishes to analyze and focus all analysis techniques on what was defined. A subject, for example, can be a newly proposed product idea.

Trend Analysis

The next step of the method is to conduct a trend analysis. Trend analysis is an analysis of macro environmental factors in the external environment of a business, also called PEST analysis. It consists of analyzing political, economical, social, technological and demographic trends. This can be done by first determining which factors, on each level, are relevant for the chosen subject and to score each item as to specify its importance. This allows the business to identify those factors that can influence them. They can't control these factors but they can try to cope with them by adapting themselves. The trends (factors) that are addressed in PEST analysis are Political, Economical, Social and Technological; but for context analysis Demographic trends are also of importance. Demographic trends are those factors that have to do with the population, like for example average age, religion, education etc. Demographic information is of importance if, for example during market

research, a business wants to determine a particular market segment to target. The other trends are described in environmental scanning and PEST analysis. Trend analysis only covers part of the external environment. Another important aspect of the external environment that a business should consider is its competition. This is the next step of the method, competitor analysis.

Competitor Analysis

As one can imagine, it is important for a business to know who its competition is, how they do their business and how powerful they are so that they can be on the defense and offense. In Competitor analysis a couple of techniques are introduced how to conduct such an analysis. Here I will introduce another technique which involves conducting four sub analyses, namely: determination of competition levels, competitive forces, competitor behavior and competitor strategy.

Competition Levels

Businesses compete on several levels and it is important for them to analyze these levels so that they can understand the demand. Competition is identified on four levels:

- Consumer needs: level of competition that refers to the needs and desires of consumers. A business should ask: What are the desires of the consumers?
- General competition: The kind of consumer demand. For example: do consumers prefer shaving with electric razor or a razor blade?
- Brand: This level refers to brand competition. Which brands are preferable to a consumer?
- Product: This level refers to the type of demand. Thus what types of products do consumers prefer?

Another important aspect of a competition analysis is to increase the consumer insight. For example: [Ducati] has, by interviewing a lot of their customers, concluded that their main competitor is not another bicycle, but sport-cars like [Porsche] or [GM]. This will of course influence the competition level within this business.

Competitive Forces

These are forces that determine the level of competition within a particular market. There are six forces that have to be taken into consideration, power of the competition, threat of new entrants, bargaining power of buyers and suppliers, threat of substitute products and the importance of complementary products. This analysis is described in Porter 5 forces analysis.

Competitor Behavior

Competitor behaviors are the defensive and offensive actions of the competition.

Competitor Strategy

These strategies refer to how an organization competes with other organizations. And these are: low price strategy and product differentiation strategy.

Opportunities and Threats

The next step, after the trend analysis and competitor analysis are conducted, is to determine threats and opportunities posed by the market. The trends analysis revealed a set of trends that can influence the business in either a positive or a negative manner. These can thus be classified as either opportunities or threats. Likewise, the competitor analysis revealed positive and negative competition issues that can be classified as opportunities or threats.

Organization Analysis

The last phase of the method is an analysis of the internal environment of the organization, thus the organization itself. The aim is to determine which skills, knowledge and technological fortes the business possesses. This entails conducting an internal analysis and a competence analysis.

Internal Analysis

The internal analysis, also called SWOT analysis, involves identifying the organizations strengths and weaknesses. The strengths refer to factors that can result in a market advantage and weaknesses to factors that give a disadvantage because the business is unable to comply with the market needs.

Competence Analysis

Competences are the combination of a business' knowledge, skills and technology that can give them the edge versus the competition. Conducting such an analysis involves identifying market related competences, integrity related competences and functional related competences.

SWOT-i Matrix

The previous sections described the major steps involved in context analysis. All these steps resulted in data that can be used for developing a strategy. These are summarized in a SWOT-i matrix. The trend and competitor analysis revealed the opportunities and threats posed by the market. The organization analysis revealed the competences of the organization and also its strengths and weaknesses. These strengths, weaknesses, opportunities and threats summarize the entire context analysis. A SWOT-i matrix, depicted in the table below, is used to depict these and to help visualize the strategies that are to be devised. SWOT- i stand for Strengths, Weaknesses, Opportunities, Threats and Issues. The Issues refer to strategic issues that will be used to devise a strategic plan.

	Opportunities (O_1, O_2, ..., O_n)	**Threats (T_1, T_2, ..., T_n)**
Strengths (S_1, S_2, ..., S_n)	$S_1O_1 \dots S_nO_1$ ... $S_1O_n \dots S_nO_n$	$S_1T_1 \dots S_nT_1$ ... $S_1T_n \dots S_nT_n$
Weaknesses (W_1, W_2, ..., W_n)	$W_1O_1 \dots W_nO_1$ ... $W_1O_n \dots W_nO_n$	$W_1T_1 \dots W_nT_1$ ... $W_1T_n \dots W_nT_n$

This matrix combines the strengths with the opportunities and threats, and the weaknesses with the opportunities and threats that were identified during the analysis. Thus the matrix reveals four clusters:

- Cluster strengths and opportunities: use strengths to take advantage of opportunities.
- Cluster strengths and threats: use strengths to overcome the threats
- Cluster weaknesses and opportunities: certain weaknesses hamper the organization from taking advantage of opportunities therefore they have to look for a way to turn those weaknesses around.
- Cluster weaknesses and threats: there is no way that the organization can overcome the threats without having to make major changes.

Strategic Plan

The ultimate goal of context analysis is to develop a strategic plan. The previous sections described all the steps that form the stepping stones to developing a strategic plan of action for the organization .The trend and competitor analysis gives insight to the opportunities and threats in the market and the internal analysis gives insight to the competences of the organization. And these were combined in the SWOT-i matrix. The SWOT-i matrix helps identify issues that need to be dealt with. These issues need to be resolved by formulating an objective and a plan to reach that objective, a strategy.

Example

Joe Arden is in the process of writing a business plan for his business idea, Arden Systems. Arden Systems will be a software business that focuses on the development of software for small businesses. Joe realizes that this is a tough market because there are many software companies that develop business software. Therefore, he conducts context analysis to gain insight into the environment of the business in order to develop a strategic plan of action to achieve competitive advantage within the market.

Define Market

First step is to define a market for analysis. Joe decides that he wants to focus on small businesses consisting of at most 20 employees.

Trend Analysis

Next step is to conduct trend analysis. The macro environmental factors that Joe should take into consideration are as follows:

- Political trend: Intellectual property rights
- Economical trend: Economic growth
- Social trend: Reduce operational costs; Ease for conducting business administration

- Technological trend: Software suites; Web applications
- Demographic trend: Increase in the graduates of IT related studies

Competitor Analysis

Following trend analysis is competitor analysis. Joe analyzes the competition on four levels to gain insight into how they operate and where advantages lie.

- Competition level:
    - Consumer need: Arden Systems will be competing on the fact that consumers want efficient and effective conducting of a business
    - Brand: There are software businesses that have been making business software for a while and thus have become very popular in the market. Competing based on brand will be difficult.
    - Product: They will be packaged software like the major competition.
- Competitive forces: Forces that can affect Arden Systems are in particular:
    - The bargaining power of buyers: the extent to which they can switch from one product to the other.
    - Threat of new entrants: it is very easy for someone to develop a new software product that can be better than Arden's.
    - Power of competition: the market leaders have most of the cash and customers; they have to power to mold the market.
- Competitor behavior: The focus of the competition is to take over the position of the market leader.
- Competitor strategy: Joe intends to compete based on product differentiation.

Opportunities and Threats

Now that Joe has analyzed the competition and the trends in the market he can define opportunities and threats.

- Opportunities:
    - Because the competitors focus on taking over the leadership position, Arden can focus on those segments of the market that the market leader ignores. This allows them to take over where the market leader shows weakness.
    - The fact that there are new IT graduates, Arden can employ or partner with someone that may have a brilliant idea.
- Threats:
    - IT graduates with fresh idea's can start their own software businesses and form a major competition for Arden Systems.

Organization Analysis

After Joe has identified the opportunities and threats of the market he can try to figure out what Arden System's strengths and weaknesses are by doing an organization analysis.

- Internal analysis:
  - Strength: Product differentiation
  - Weakness: Lacks innovative people within the organization
- Competence analysis:
  - Functional related competence: Arden Systems provides system functionalities that fit small businesses.
  - Market-related competence: Arden Systems has the opportunity to focus on a part of the market which is ignored.

SWOT-i Matrix

After the previous analyses, Joe can create a SWOT-i matrix to perform SWOT analysis.

	Opportunities	Threats
Strengths	Product differentiation, market leader ignores market segment	
Weaknesses		Lack of innovation, increase in IT graduates

Strategic Plan

After creating the SWOT-i matrix, Joe is now able to devise a strategic plan.

- Focus all software development efforts to that part of the market which is ignored by market leaders, small businesses.
- Employ recent innovative It graduates to stimulate the innovation within Arden Systems.

Business Performance Management

Business performance management is a set of performance management and analytic processes that enables the management of an organization's performance to achieve one or more pre-selected goals. Synonyms for "business performance management" include "corporate performance management (CPM)" and "enterprise performance management".

Business performance management is contained within approaches to business process management.

Business performance management has three main activities:

1. selection of goals,
2. consolidation of measurement information relevant to an organization's progress against these goals, and
3. interventions made by managers in light of this information with a view to improving future performance against these goals.

Although presented here sequentially, typically all three activities will run concurrently, with interventions by managers affecting the choice of goals, the measurement information monitored, and the activities being undertaken by the organization.

Because business performance management activities in large organizations often involve the collation and reporting of large volumes of data, many software vendors, particularly those offering business intelligence tools, market products intended to assist in this process. As a result of this marketing effort, business performance management is often incorrectly understood as an activity that necessarily relies on software systems to work, and many definitions of business performance management explicitly suggest software as being a definitive component of the approach.

This interest in business performance management from the software community is sales-driven - "The biggest growth area in operational BI analysis is in the area of business performance management."

Since 1992, business performance management has been strongly influenced by the rise of the balanced scorecard framework. It is common for managers to use the balanced scorecard framework to clarify the goals of an organization, to identify how to track them, and to structure the mechanisms by which interventions will be triggered. These steps are the same as those that are found in BPM, and as a result balanced scorecard is often used as the basis for business performance management activity with organizations.

In the past, owners have sought to drive strategy down and across their organizations, transform these strategies into actionable metrics and use analytics to expose the cause-and-effect relationships that, if understood, could give insight into decision-making.

History

Reference to non-business performance management occurs in Sun Tzu's *The Art of War*. Sun Tzu claims that to succeed in war, one should have full knowledge of one's own strengths and weaknesses as well as those of one's enemies. Lack of either set of knowledge might result in defeat. Parallels between the challenges in business and those of war include:

- collecting data - both internal and external
- discerning patterns and meaning in the data (analyzing)
- responding to the resultant information

Prior to the start of the Information Age in the late 20th century, businesses sometimes took the trouble to laboriously collect data from non-automated sources. As they lacked computing resources to properly analyze the data, they often made commercial decisions primarily on the basis of intuition.

As businesses started automating more and more systems, more and more data became available. However, collection often remained a challenge due to a lack of infrastructure for data exchange or due to incompatibilities between systems. Reports on the data gathered sometimes took months to generate. Such reports allowed informed long-term strategic decision-making. However, short-term tactical decision-making often continued to rely on intuition.

In 1989 Howard Dresner, a research analyst at Gartner, popularized "business intelligence" (BI) as an umbrella term to describe a set of concepts and methods to improve business decision-making by using fact-based support systems. Performance management builds on a foundation of BI, but marries it to the planning-and-control cycle of the enterprise - with enterprise planning, consolidation and modeling capabilities.

Increasing standards, automation, and technologies have led to vast amounts of data becoming available. Data warehouse technologies have allowed the building of repositories to store this data. Improved ETL and enterprise application integration tools have increased the timely collecting of data. OLAP reporting technologies have allowed faster generation of new reports which analyze the data. As of 2010, business intelligence has become the art of sieving through large amounts of data, extracting useful information and turning that information into actionable knowledge.

Definition and Scope

Business performance management consists of a set of management and analytic processes, supported by technology, that enable businesses to define strategic goals and then measure and manage performance against those goals. Core business performance management processes include financial planning, operational planning, business modeling, consolidation and reporting, analysis, and monitoring of key performance indicators linked to strategy.

Business performance management involves consolidation of data from various sources, querying, and analysis of the data, and putting the results into practice.

Frameworks

Various frameworks for implementing business performance management exist. The discipline gives companies a top-down framework by which to align planning and execution, strategy and tactics, and business-unit and enterprise objectives. Reactions may include the Six Sigma strategy, balanced scorecard, activity-based costing (ABC), Objectives and Key Results (OKR), Total Quality Management, economic value-add, integrated strategic measurement and Theory of Constraints.

The balanced scorecard is the most widely adopted performance management framework.

Metrics and Key Performance Indicators

Some of the areas from which bank management may gain knowledge by using business performance management include:

- customer-related numbers:

    - new customers acquired
    - status of existing customers
    - attrition of customers (including breakup by reason for attrition)
- turnover generated by segments of the customers - possibly using demographic filters
- outstanding balances held by segments of customers and terms of payment - possibly using demographic filters
- collection of bad debts within customer relationships
- demographic analysis of individuals (potential customers) applying to become customers, and the levels of approval, rejections and pending numbers
- delinquency analysis of customers behind on payments
- profitability of customers by demographic segments and segmentation of customers by profitability
- campaign management
- real-time dashboard on key operational metrics
    - overall equipment effectiveness
- clickstream analysis on a website
- key product portfolio trackers
- marketing-channel analysis
- sales-data analysis by product segments
- callcenter metrics

Though the above list describes what a bank might monitor, it could refer to a telephone company or to a similar service-sector company.

Items of generic importance include:

1. consistent and correct KPI-related data providing insights into operational aspects of a company
2. timely availability of KPI-related data
3. KPIs designed to directly reflect the efficiency and effectiveness of a business
4. information presented in a format which aids decision-making for management and decision-makers
5. ability to discern patterns or trends from organized information

Business performance management integrates the company's processes with CRM or ERP. Companies should become better able to gauge customer satisfaction, control customer trends and influence shareholder value.

Application Software Types

People working in business intelligence have developed tools that ease the work of business performance management, especially when the business-intelligence task involves gathering and analyzing large amounts of unstructured data.

Tool categories commonly used for business performance management include:

- MOLAP — Multidimensional online analytical processing, sometimes simply called "analytics" (based on dimensional analysis and the so-called "hypercube" or "cube")
- scorecarding, dashboarding and data visualization
- data warehouses
- document warehouses
- text mining
- DM — data mining
- BPO — business performance optimization
- EPM — enterprise performance management
- EIS — executive information systems
- DSS — decision support systems
- MIS — management information systems
- SEMS — strategic enterprise management software
- EOI — Operational intelligence Enterprise Operational Intelligence Software

Design and Implementation

Questions asked when implementing a business performance management program include:

- Goal-alignment queries

  Determine the short- and medium-term purpose of the program. What strategic goal(s) of the organization will the program address? What organizational mission/vision does it relate to? A hypothesis needs to be crafted that details how this initiative will eventually improve results / performance (i.e. a strategy map).

- Baseline queries

  Assess current information-gathering competency. Does the organization have the capability to monitor important sources of information? What data is being collected and how is it being stored? What are the statistical parameters of this data, e.g., how much random variation does it contain? Is this being measured?

- Cost and risk queries

Estimate the financial consequences of a new BI initiative. Assess the cost of the present operations and the increase in costs associated with the BPM initiative. What is the risk that the initiative will fail? This risk assessment should be converted into a financial metric and included in the planning.

- Customer and stakeholder queries

  Determine who will benefit from the initiative and who will pay. Who has a stake in the current procedure? What kinds of customers / stakeholders will benefit directly from this initiative? Who will benefit indirectly? What quantitative / qualitative benefits follow? Is the specified initiative the best or only way to increase satisfaction for all kinds of customers? How will customer benefits be monitored? What about employees, shareholders, and distribution channel members?

- Metrics-related queries

  Information requirements need operationalization into clearly defined metrics. Decide which metrics to use for each piece of information being gathered. Are these the best metrics and why? How many metrics need to be tracked? If this is a large number (it usually is), what kind of system can track them? Are the metrics standardized, so they can be benchmarked against performance in other organizations? What are the industry standard metrics available?

- Measurement methodology-related queries

  Establish a methodology or a procedure to determine the best (or acceptable) way of measuring the required metrics. How frequently will data be collected? Are there any industry standards for this? Is this the best way to do the measurements? How do we know that?

- Results-related queries

  Monitor the BPM program to ensure that it meets objectives. The program itself may require adjusting. The program should be tested for accuracy, reliability, and validity. How can it be demonstrated that the BI initiative, and not something else, contributed to a change in results? How much of the change was probably random?

Business Process Discovery

Business process discovery (BPD) related to process mining is a set of techniques that automatically construct a representation of an organization's current business processes and its major process variations. These techniques use evidence found in the existing technology systems that run business processes within an organization.

Business Process Discovery Techniques

Business process discovery techniques embody the following properties:

- Emergent paradigm - Current methods are based on top-down structured manual interviews relying on second-hand representations of the business process/system behaviors. An automated discovery process relies on collecting data from the information system over a period of time. This data can then be analyzed to form a process model.
- Automated process discovery – By automating the analysis of the data, the subjectivity of current manual process analysis techniques is removed. The automated system has an ingrained methodology that — through repeated trials — has been shown to accurately discover processes and process variations without bias.
- Accurate information- Since the information is collected from the actual source it cannot be inaccurate, as opposed to gathering it from second party representation.
- Complete information - An automated process captures all the information that is occurring within the system and represents them by time, date, user, etc.... Since the information is collected from real-time interactions, it is not subject to lost or selective memory issues. This includes completeness regarding exceptions in the processes. Often, exceptions are treated as statistical "noise," which may exclude important inefficiencies in business processes.
- Standardized Process - The automated collection of information yields process data which can be grouped, quantified and classified. This supplies a basis for the development and monitoring of both current and new processes, to which benchmarks can be assigned. These benchmarks are the root of both new process design and the determination of problem root cause. Additionally, standardized process data can set the stage for efforts at continuous process improvement.

Application / Techniques

Business Process Discovery complements and builds upon the work in many other fields.

- Process discovery is one of the three main types of process mining. The other two types of process mining are conformance checking and model extension/enhancement. All of these techniques aim at extracting process related knowledge from event logs. In the case of process discovery, there is no prior process model; the model is discovered based on event logs. Conformance checking aims at finding differences between a given process model and event log. This way it is possible to quantify compliance and analyze discrepancies. Enhancement takes an a priori model and improves or extends it using information from the event log, e.g., show bottlenecks.
- Business process discovery is the next level of understanding in the emerging field of business analytics, which allows organizations to view, analyze and adjust the underlying structure and processes that go into day-to-day operations. This discovery includes information gathering of all of the components of a business process, including technology, people, department procedures and protocols.
- Business process discovery creates a process master which complements business process analysis (BPA). BPA tools and methodologies are well suited to top-down hierarchical process decomposition, and analysis of to-be processes. BPD provides a bottoms-up analysis

that marries to the top-down to provide a complete business process, organized hierarchically by BPA.

- Business Intelligence provides organizations with reporting and analytics on the data in their organizations. However, BI has no process model, awareness or analytics. BPD complements BI by providing an explicit process view to current operations, and providing analytics on that process model to help organizations identify and act upon business process inefficiencies, or anomalies.
- Web analytics are a limited example of BPD in that web analytics reconstruct the web-user's process as they interact with a Web-site. However, these analytics are limited to the process as is contained within the session, from the users perspective and with respect to just the web-based system and process.
- Business triage provides a framework for categorizing the processes identified by business process analysis (BPA) based on their relative importance to achieving a stated, measurable goal or outcome. Utilizing the same categories employed by military medical and disaster medical services, business processes are categorized as:
    - Essential/critical (red process) - Process essential for achieving outcomes/goals
    - Important/urgent (yellow process) - Process which speeds achieving outcomes/goals
    - Optional/supportive (green process) - Process not needed to achieve outcomes/goals

Resources are allocated based on the process category with resources first dedicated to red processes, then yellow processes and finally green processes. In the event that resources become limited, resources are first withheld from Green Processes, then Yellow Processes. Resources are only withheld from Red Processes if failure to achieve outcomes/goals is acceptable.

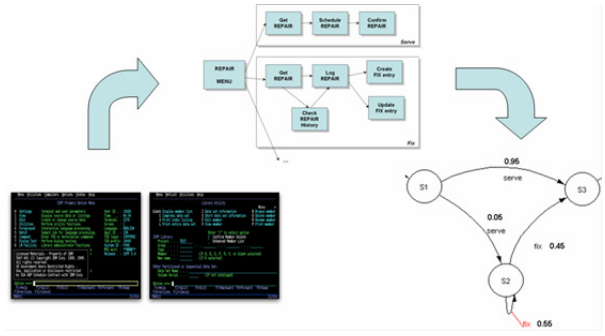
The Purpose / Example

A small example may illustrate the Business Process Discovery technology that is required today. Automated Business Process Discovery tools capture the required data, and transform it into a structured dataset for the actual diagnosis; A major challenge is the grouping of repetitive actions from the users into meaningful events. Next, these Business process discovery tools propose probabilistic process models. Probabilistic behavior is essential for the analysis and the diagnosis of the processes. The following shows an example where a probabilistic repair-process is recovered from user actions. The "as-is" process model shows exactly where the pain is in this business. Five percent faulty repairs is a bad sign, but worse, the repetitive fixes that are needed to complete those repairs are cumbersome.

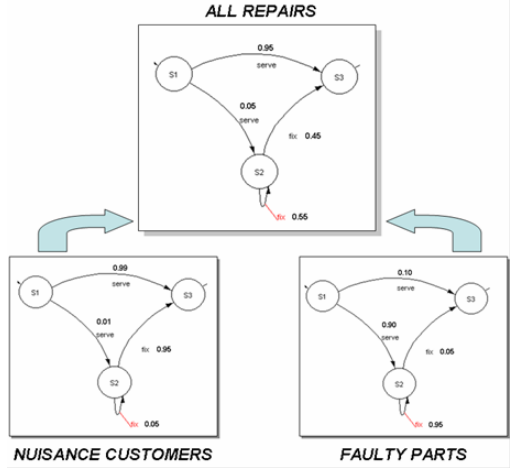
History

- Business intelligence (BI) emerged more than 20 years ago and is critical for reporting what is happening within an organization's systems. Yet current BI applications and data mining technologies are not always suited for evaluating the level of detail required to analyze unstructured data and the human dynamics of business processes.

- Six-Sigma and other quantitative approaches to business process improvement have been employed for over a decade with varying degrees of success. A major limitation to the success of these approaches is the availability of accurate data to form the basis of the analysis. With BPD, many six-sigma organizations are finding the ability to extend their analysis into major business processes effectively.
- Process mining According to researchers at Eindhoven University of Technology, (PM) emerged as a scientific discipline around 1990 when techniques like the Alpha algorithm made it possible to extract process models (typically represented as Petri nets) from event logs. However, the recognition of this wanna-be scientific discipline is extremely limited within few countries. As the hype of Process Mining carried by Eindhoven University of Technology growing, more and more criticisms have emerged pointing out that Process Mining is no more than a set of algorithms which solves a specific and simple business problem: business process discovery and auxiliary evaluation methods. Today, there are over 100 process mining algorithms that are able to discover process models that also include concurrency, e.g., genetic process discovery techniques, heuristic mining algorithms, region-based mining algorithms, and fuzzy mining algorithms.



A deeper analysis of the "as-is" process data may reveal which are the faulty parts that are responsible for the overall behavior in this example. It may lead to the discovery of subgroups of repairs that actually need management focus for improvement.



In this case, it would become obvious that the faulty parts are also responsible for the repetitive fixes. Similar applications have been documented, such as a Healthcare Insurance Provider case where in 4 months the ROI of Business Process Analysis was earned from precisely comprehending its claims handling process and discovering the faulty parts.

Information System

An information system (IS) is any organized system for the collection, organization, storage and communication of information. More specifically, it is the study of complementary networks that people and organizations use to collect, filter, process, create and distribute data.

"An information system (IS) is a group of components that interact to produce information"

A computer information system is a system composed of people and computers that processes or interprets information. The term is also sometimes used in more restricted senses to refer to only the software used to run a computerized database or to refer to only a computer system.

Information system is an academic study of systems with a specific reference to information and the complementary networks of hardware and software that people and organizations use to collect, filter, process, create and also distribute data. An emphasis is placed on an Information System having a definitive Boundary, Users, Processors, Stores, Inputs, Outputs and the aforementioned communication networks.

Any specific information system aims to support operations, management and decision-making. An information system is the information and communication technology (ICT) that an organization uses, and also the way in which people interact with this technology in support of business processes.

Some authors make a clear distinction between information systems, computer systems, and business processes. Information systems typically include an ICT component but are not purely concerned with ICT, focusing instead on the end use of information technology. Information systems are also different from business processes. Information systems help to control the performance of business processes.

Alter argues for advantages of viewing an information system as a special type of work system. A work system is a system in which humans or machines perform processes and activities using resources to produce specific products or services for customers. An information system is a work system whose activities are devoted to capturing, transmitting, storing, retrieving, manipulating and displaying information.

As such, information systems inter-relate with data systems on the one hand and activity systems on the other. An information system is a form of communication system in which data represent and are processed as a form of social memory. An information system can also be considered a semi-formal language which supports human decision making and action.

Information systems are the primary focus of study for organizational informatics.

Overview

Silver et al. (1995) provided two views on IS that includes software, hardware, data, people, and procedures. Zheng provided another system view of information system which also adds processes and essential system elements like environment, boundary, purpose, and interactions. The Association for Computing Machinery defines "Information systems specialists [as] focus[ing] on inte-

grating information technology solutions and business processes to meet the information needs of businesses and other enterprises."

There are various types of information systems, for example: transaction processing systems, decision support systems, knowledge management systems, learning management systems, database management systems, and office information systems. Critical to most information systems are information technologies, which are typically designed to enable humans to perform tasks for which the human brain is not well suited, such as: handling large amounts of information, performing complex calculations, and controlling many simultaneous processes.

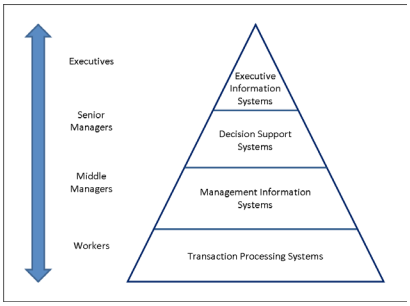
Information technologies are a very important and malleable resource available to executives. Many companies have created a position of chief information officer (CIO) that sits on the executive board with the chief executive officer (CEO), chief financial officer (CFO), chief operating officer (COO), and chief technical officer (CTO). The CTO may also serve as CIO, and vice versa. The chief information security officer (CISO) focuses on information security management.

The six components that must come together in order to produce an information system are:

1. Hardware: The term hardware refers to machinery. This category includes the computer itself, which is often referred to as the central processing unit (CPU), and all of its support equipments. Among the support equipments are input and output devices, storage devices and communications devices.
2. Software: The term software refers to computer programs and the manuals (if any) that support them. Computer programs are machine-readable instructions that direct the circuitry within the hardware parts of the system to function in ways that produce useful information from data. Programs are generally stored on some input / output medium, often a disk or tape.
3. Data: Data are facts that are used by programs to produce useful information. Like programs, data are generally stored in machine-readable form on disk or tape until the computer needs them.
4. Procedures: Procedures are the policies that govern the operation of a computer system. "Procedures are to people what software is to hardware" is a common analogy that is used to illustrate the role of procedures in a system.
5. People: Every system needs people if it is to be useful. Often the most over-looked element of the system are the people, probably the component that most influence the success or failure of information systems. This includes "not only the users, but those who operate and service the computers, those who maintain the data, and those who support the network of computers." <Kroenke, D. M. (2015). MIS Essentials. Pearson Education>
6. Feedback: it is another component of the IS, that defines that an IS may be provided with a feedback (Although this component isn't necessary to function).

Data is the bridge between hardware and people. This means that the data we collect is only data, until we involve people. At that point, data is now information.

Types of Information System



A four level

The "classic" view of Information systems found in the textbooks in the 1980s was of a pyramid of systems that reflected the hierarchy of the organization, usually transaction processing systems at the bottom of the pyramid, followed by management information systems, decision support systems, and ending with executive information systems at the top. Although the pyramid model remains useful, since it was first formulated a number of new technologies have been developed and new categories of information systems have emerged, some of which no longer fit easily into the original pyramid model.

Some examples of such systems are:

- data warehouses
- enterprise resource planning
- enterprise systems
- expert systems
- search engines
- geographic information system
- global information system
- office automation.

A computer(-based) information system is essentially an IS using computer technology to carry out some or all of its planned tasks. The basic components of computer-based information systems are:

- *Hardware*- these are the devices like the monitor, processor, printer and keyboard, all of which work together to accept, process, show data and information.
- *Software*- are the programs that allow the hardware to process the data.
- *Databases*- are the gathering of associated files or tables containing related data.
- *Networks*- are a connecting system that allows diverse computers to distribute resources.
- *Procedures*- are the commands for combining the components above to process information and produce the preferred output.

The first four components (hardware, software, database, and network) make up what is known as the information technology platform. Information technology workers could then use these components to create information systems that watch over safety measures, risk and the management of data. These actions are known as information technology services.

Certain information systems support parts of organizations, others support entire organizations, and still others, support groups of organizations. Recall that each department or functional area within an organization has its own collection of application programs, or information systems. These functional area information systems (FAIS) are supporting pillars for more general IS namely, business intelligence systems and dashboards. As the name suggest, each FAIS support a particular function within the organization, e.g.: accounting IS, finance IS, production/operation management (POM) IS, marketing IS, and human resources IS. In finance and accounting, managers use IT systems to forecast revenues and business activity, to determine the best sources and uses of funds, and to perform audits to ensure that the organization is fundamentally sound and that all financial reports and documents are accurate. Other types of organizational information systems are FAIS, Transaction processing systems, enterprise resource planning, office automation system, management information system, decision support system, expert system, executive dashboard, supply chain management system, and electronic commerce system. Dashboards are a special form of IS that support all managers of the organization. They provide rapid access to timely information and direct access to structured information in the form of reports. Expert systems attempt to duplicate the work of human experts by applying reasoning capabilities, knowledge, and expertise within a specific domain.

Information System Development

Information technology departments in larger organizations tend to strongly influence the development, use, and application of information technology in the organizations. A series of methodologies and processes can be used to develop and use an information system. Many developers now use an engineering approach such as the system development life cycle (SDLC), which is a systematic procedure of developing an information system through stages that occur in sequence. Recent research aims at enabling and measuring the ongoing, collective development of such systems within an organization by the entirety of human actors themselves. An information system can be developed in house (within the organization) or outsourced. This can be accomplished by outsourcing certain components or the entire system. A specific case is the geographical distribution of the development team (offshoring, global information system).

A computer-based information system, following a definition of Langefors, is a technologically implemented medium for:

- recording, storing, and disseminating linguistic expressions,
- as well as for drawing conclusions from such expressions.

Geographic information systems, land information systems, and disaster information systems are examples of emerging information systems, but they can be broadly considered as spatial information systems. System development is done in stages which include:

- Problem recognition and specification

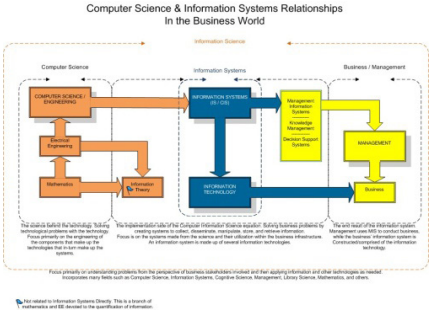
- Information gathering
- Requirements specification for the new system
- System design
- System construction
- System implementation
- Review and maintenance.

As an Academic Discipline

The field of study called *information systems* encompasses a variety of topics including systems analysis and design, computer networking, information security, database management and decision support systems. *Information management* deals with the practical and theoretical problems of collecting and analyzing information in a business function area including business productivity tools, applications programming and implementation, electronic commerce, digital media production, data mining, and decision support. *Communications and networking* deals with the telecommunication technologies. Information systems bridges business and computer science using the theoretical foundations of information and computation to study various business models and related algorithmic processes on building the IT systems within a computer science discipline. Computer information system(s) (CIS) is a field studying computers and algorithmic processes, including their principles, their software and hardware designs, their applications, and their impact on society, whereas IS emphasizes functionality over design.

Several IS scholars have debated the nature and foundations of Information Systems which has its roots in other reference disciplines such as Computer Science, Engineering, Mathematics, Management Science, Cybernetics, and others. Information systems also can be defined as a collection of hardware, software, data, people and procedures that work together to produce quality information.

Differentiating IS from Related Disciplines



Information Systems relationship to Information Technology, Computer Science, Information Science, and Business.

Similar to computer science, other disciplines can be seen as both related and foundation disciplines of IS. The domain of study of IS involves the study of theories and practices related to

the social and technological phenomena, which determine the development, use, and effects of information systems in organization and society. But, while there may be considerable overlap of the disciplines at the boundaries, the disciplines are still differentiated by the focus, purpose, and orientation of their activities.

In a broad scope, the term *Information Systems* is a scientific field of study that addresses the range of strategic, managerial, and operational activities involved in the gathering, processing, storing, distributing, and use of information and its associated technologies in society and organizations. The term information systems is also used to describe an organizational function that applies IS knowledge in industry, government agencies, and not-for-profit organizations. *Information Systems* often refers to the interaction between algorithmic processes and technology. This interaction can occur within or across organizational boundaries. An information system is the technology an organization uses and also the way in which the organizations interact with the technology and the way in which the technology works with the organization's business processes. Information systems are distinct from information technology (IT) in that an information system has an information technology component that interacts with the processes' components.

One problem with that approach is that it prevents the IS field from being interested in non-organizational use of ICT, such as in social networking, computer gaming, mobile personal usage, etc. A different way of differentiating the IS field from its neighbours is to ask, "Which aspects of reality are most meaningful in the IS field and other fields?" This approach, based on philosophy, helps to define not just the focus, purpose and orientation, but also the dignity, destiny and responsibility of the field among other fields. *International Journal of Information Management*, 30, 13-20.

Career Pathways

Information Systems have a number of different areas of work:

- IS strategy
- IS management
- IS development
- IS iteration
- IS organization

There is a wide variety of career paths in the information systems discipline. "Workers with specialized technical knowledge and strong communications skills will have the best prospects. Workers with management skills and an understanding of business practices and principles will have excellent opportunities, as companies are increasingly looking to technology to drive their revenue."

Information technology is important to the operation of contemporary businesses, it offers many employment opportunities. The information systems field includes the people in organizations who design and build information systems, the people who use those systems, and the people responsible for managing those systems. The demand for traditional IT staff such as programmers, business analysts, systems analysts, and designer is significant. Many well-paid jobs exist in areas of Information technology. At the top of the list is the chief information officer (CIO).

The CIO is the executive who is in charge of the IS function. In most organizations, the CIO works with the chief executive officer (CEO), the chief financial officer (CFO), and other senior executives. Therefore, he or she actively participates in the organization's strategic planning process.

Research

Information systems research is generally interdisciplinary concerned with the study of the effects of information systems on the behaviour of individuals, groups, and organizations. Hevner et al. (2004) categorized research in IS into two scientific paradigms including *behavioural science* which is to develop and verify theories that explain or predict human or organizational behavior and *design science* which extends the boundaries of human and organizational capabilities by creating new and innovative artifacts.

Salvatore March and Gerald Smith proposed a framework for researching different aspects of Information Technology including outputs of the research (research outputs) and activities to carry out this research (research activities). They identified research outputs as follows:

1. *Constructs* which are concepts that form the vocabulary of a domain. They constitute a conceptualization used to describe problems within the domain and to specify their solutions.
2. A *model* which is a set of propositions or statements expressing relationships among constructs.
3. A *method* which is a set of steps (an algorithm or guideline) used to perform a task. Methods are based on a set of underlying constructs and a representation (model) of the solution space.
4. An *instantiation* is the realization of an artifact in its environment.

Also research activities including:

1. *Build* an artifact to perform a specific task.
2. *Evaluate* the artifact to determine if any progress has been achieved.
3. Given an artifact whose performance has been evaluated, it is important to determine why and how the artifact worked or did not work within its environment. Therefore, *theorize* and *justify* theories about IT artifacts.

Although Information Systems as a discipline has been evolving for over 30 years now, the core focus or identity of IS research is still subject to debate among scholars. There are two main views around this debate: a narrow view focusing on the IT artifact as the core subject matter of IS research, and a broad view that focuses on the interplay between social and technical aspects of IT that is embedded into a dynamic evolving context. A third view calls on IS scholars to pay balanced attention to both the IT artifact and its context.

Since the study of information systems is an applied field, industry practitioners expect information systems research to generate findings that are immediately applicable in practice. This is not always the case however, as information systems researchers often explore behavioral issues in much more depth than practitioners would expect them to do. This may render information systems research results difficult to understand, and has led to criticism.

In the last ten years the business trend is represented by the considerable increasing of Information Systems Function (ISF) role, especially with regard the enterprise strategies and operations supporting. It became a key-factor to increase productivity and to support new value creation. To study an information system itself, rather than its effects, information systems models are used, such as EATPUT.

The international body of Information Systems researchers, the Association for Information Systems (AIS), and its Senior Scholars Forum Subcommittee on Journals (23 April 2007), proposed a 'basket' of journals that the AIS deems as 'excellent', and nominated: *Management Information Systems Quarterly* (MISQ), *Information Systems Research* (ISR), *Journal of the Association for Information Systems* (JAIS), *Journal of Management Information Systems* (JMIS), *European Journal of Information Systems* (EJIS), and *Information Systems Journal* (ISJ).

A number of annual information systems conferences are run in various parts of the world, the majority of which are peer reviewed. The AIS directly runs the International Conference on Information Systems (ICIS) and the Americas Conference on Information Systems (AMCIS), while AIS affiliated conferences include the Pacific Asia Conference on Information Systems (PACIS), European Conference on Information Systems (ECIS), the Mediterranean Conference on Information Systems (MCIS), the International Conference on Information Resources Management (Conf-IRM) and the Wuhan International Conference on E-Business (WHICEB). AIS chapter conferences include Australasian Conference on Information Systems (ACIS), Information Systems Research Conference in Scandinavia (IRIS), Information Systems International Conference (ISICO), Conference of the Italian Chapter of AIS (itAIS), Annual Mid-Western AIS Conference (MWAIS) and Annual Conference of the Southern AIS (SAIS). EDSIG, which is the special interest group on education of the AITP, organizes the Conference on Information Systems and Computing Education and the Conference on Information Systems Applied Research which are both held annually in November.

The Impact on Economic Models

- Microeconomic theory model
- Transaction cost theory
- Agency theory

Organizational Intelligence

Organizational Intelligence (OI) is the capability of an organization to comprehend and conclude knowledge relevant to its business purpose. In other words, it is the intellectual capacity of the entire organizations. With relevant organizational intelligence comes great potential value for companies and therefore organizations find study where their strengths and weaknesses lie in responding to change and complexity. Organizational Intelligence embraces both knowledge management and organizational learning, as it is the application of knowledge management concepts to a business environment, additionally including learning mechanisms, comprehension models and business value network models, such as the balanced scorecard concept. Organizational Intelligence consists of the ability to make sense of complex situations and act effectively, to interpret and act upon relevant events and signals in the environment. It also includes the ability to develop,

share and use knowledge relevant to its business purpose as well as the ability to reflect and learn from experience

While organizations in the past have been viewed as compilations of tasks, products, employees, profit centers and processes, today they are seen as intelligent systems that are designed to manage knowledge. Scholars have shown that organizations engage in learning processes using tacit forms of intuitive knowledge, hard data stored in computer networks and information gleaned from the environment, all of which are used to make sensible decisions. Because this complex process involves large numbers of people interacting with diverse information systems, organizational intelligence is more than the aggregate intelligence of organizational members; it is the intelligence of the organization itself as a larger system.

Organizational Intelligence vs Operational Intelligence

Organizational Intelligence and operational intelligence are usually seen as subsets of business analytics, since both are types of know-how that have the goal of improving business performance across the enterprise. Operational Intelligence is often linked to or compared with real-time business intelligence (BI) since both deliver visibility and insight into business operations. Operational Intelligence differs from BI in being primarily activity-centric, whereas BI is primarily data-centric and relies on a database (or Hadoop cluster) as well as after-the-fact and report-based approaches to identifying patterns in data. By definition, Operational Intelligence works in real-time and transforms unstructured data streams—from log file, sensor, network and service data—into real-time, actionable intelligence.

While Operational Intelligence is activity-focused and BI is data-focused, Organizational Intelligence differs from these other approaches in being workforce- or organization-focused. Organizational Intelligence helps companies understand the relationships that drive their business—by identifying communities as well as employee workflow and collaborative communications patterns across geographies, divisions, and internal and external organizations.

Information Process

There are many aspects that organizations must consider in the three steps that they take to gain information. Without these considerations, organizations may experience strategic challenges.

Acquiring Information

First of all, organizations must acquire applicable information to make beneficial predictions. An organization must ask what they already know and need to know. They must also know the timeframe in which the information is needed and where and to find it. To make the best judgements, they must also evaluate the value of the information. Seemingly valuable information that costs more to find than gain from can hurt the company. If judged valuable, the organization must find the most efficient means of acquiring it.

Processing Information

After acquiring the right information, an organization must know how to properly process it. They

need to know how they can make new information more retrievable and how they can make sure that the information gets disseminated to the right people. The organization must figure out how to secure it and how long and if long, how they need to preserve it.

Utilization of Information

The last step includes the utilization of the information. An organization should ask themselves if they are looking at the right information and if so, if they are placing them in the right context. They must consider the possible environmental changes alter the informational value and determine all the relevant connections and patterns. Not forgetting to know if they are including the right people in the decision making process and if there are any technology that can improve the decision making.

Organizational Ignorance

There are briefly four dimensions of problems that many organizations face when dealing with information. This is also referred to as organizational ignorance.

Uncertainty

An organization may be uncertain when it does not possess enough or the right information. To exemplify, a company may be uncertain in a competitive landscape because it does not have enough information to see how the competitors will act. This does not imply that the context of the situation is complex or unclear. Uncertainty can even exist when the range of possibilities is small and simple. There are different degrees of uncertainty. First of all an organization can be completely determined (complete certainty), have some probabilities (risk), probabilities estimated with lesser confidence (subjective uncertainty), unknown probabilities (traditional uncertainty) or undefined (complete uncertainty). However even with the lack of clarity, uncertainty assumes that the context of the problem is clear and well-understood.

Complexity

An organization may be processing more information than they can manage. Complexity doesn't always correlate with vagueness or unpredictability. Rather, it occurs when there are too much or when the scope is too large to process. Organizations with complexity problems have interrelated variables, solutions and methods. Managing these problems is dependent of the individuals and the organizations. For instance, uninformed and novices must deal with each elements and relationships one by one but experts can perceive the situation better and find familiar patterns more easily. Organizations facing complexity must have the capacity to locate, map, collect, share, exploit on what the organizations need to know.

Ambiguity

An organization may not have a conceptual framework for interpreting the information. If uncertainty represents not having answers, and complexity represents difficulty in finding them, ambiguity represents not being able to formulate the right questions. Ambiguity cannot be resolved by increasing the amount of information. An organization must be able to interpret and explain the

information in collective agreement. Hypotheses should be continuously made and discussed and key communication activities such as face-to-face conversations must be made. Resolving ambiguity in the earlier stages than competitors gives organizations much advantage because it helps organizations to make more appropriate and strategic decisions and have better awareness.

Equivocality

An organization may be having competing frameworks for interpreting a job. Equivocality refers to multiple interpretations of the field. Each interpretation is unambiguous but differ from each other and they may be mutually exclusive or in conflict. Equivocality result not only because everyone's experiences and values are unique but also from unreliable or conflicting preferences and goals, different interests or vague roles and responsibilities.

Information Organization and Culture

A culture of the organization describes how the organization will work in order to succeed. It can simply be described as the organization's atmosphere or values. Organizational culture is important because it can be used as a successful leadership tool to shape and improve the organization. Once the culture is settled, it can be used by the leader to deliver his/her vision to the organization. Moreover, if the leader deeply understands the organizational culture, he/she can also use it to predict a future outcome in certain situations.

Control

An organization with control culture is company oriented and reality oriented. They will succeed by controlling and keeping restrictions. The organization will value timeliness of information, security and hierarchical standardization. They make plans and maintain a process. This organization has stability, predictability and authority. For example, an organization with control culture can be monarchy.

Competence

An organization with competence culture is company oriented and possibility oriented. They will succeed by being the best with exclusivity of the information. The organization values efficiency, accuracy and achievement. They look for creativity and expertise from the people in the organization. For example, an organization with competence culture can be...

Cultivation

An organization with cultivation culture is people oriented and possibility oriented. They will succeed by growing people, who fulfill the shared vision. The organization values self-actualization and brilliance. They also prioritizes the idea from people. For example, an organization with cultivation culture can be technological utopianism.

Collaboration

An organization with collaboration culture is people oriented and reality oriented. They will succeed by working together. The organization values affiliation and teamwork. They also prioritizes

people in the organization. This organization has accessibility and inclusiveness of information. For example, an organization with collaboration culture can be anarchy.

Organizational Intelligence and Innovation

An organization's leadership effectiveness is closely related to the organization's intelligence and innovation. There are six leadership factors that determines organization' atmosphere : flexibility (how freely people can communicate with each other and innovate), responsibility (sense of loyalty to the organization), the standards set by people in the organization, appropriate feedback and rewards, the clear vision shared by people and the amount of commitment to the goal. Combination of these factors result in six different leadership styles: Coercive/Commanding, Authoritative/Visionary, Affiliative, Democratic, Coaching and Pacesetting.

Furthermore, organizational intelligence is a collection of individual intelligence. The leadership style of the organization and its atmosphere are related to the organization's innovation. Innovation happens when there are new information getting shared and processed efficiently in the organization.

Theories

Round Table

In King Arthur's Round Table, Harvard professor David Perkins uses the metaphor of the Round Table to discuss how collaborative conversations create smarter organizations. The Round Table is one of the most familiar stories of Arthurian legend since it's meant to signal the shift in power from a king who normally sat at the head of a long table and made long pronouncements while everyone else listened. By reducing hierarchy and making collaboration easier, Arthur discovered an important source of power—organizational intelligence—that allowed him to unite medieval England.

Lawnmower Paradox

The lawnmower paradox, another metaphor from Perkins' book, describes the fact that, while pooling physical effort is easy, pooling mental effort is hard. "It's a lot easier for 10 people to collaborate on mowing a large lawn than for 10 people to collaborate on designing a lawnmower." An organization's intelligence is reflected by the types of conversations—face-to-face and electronic, from the mailroom to the boardroom—which members have with one another. "At the top, top level, organizational intelligence depends on ways of interacting with one another that show good knowledge processing and positive symbolic conduct."

Harold Wilensky argued that organizational intelligence benefited from healthy argument and constructive rivalry.

Data Visualization

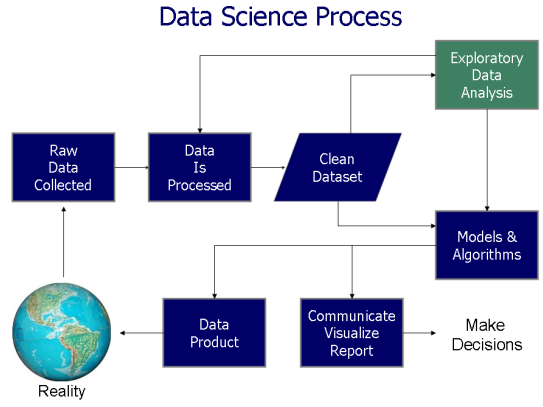
Data visualization or data visualisation is viewed by many disciplines as a modern equivalent

of visual communication. It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information".

A primary goal of data visualization is to communicate information clearly and efficiently via statistical graphics, plots and information graphics. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships in the data for one or more variables.

Data visualization is both an art and a science . It is viewed as a branch of descriptive statistics by some, but also as a grounded theory development tool by others. The rate at which data is generated has increased. Data created by internet activity and an expanding number of sensors in the environment, such as satellites, are referred to as "Big Data". Processing, analyzing and communicating this data present a variety of ethical and analytical challenges for data visualization. The field of data science and practitioners called data scientists have emerged to help address this challenge.

Overview



Data visualization is one of the steps in analyzing data and presenting it to users.

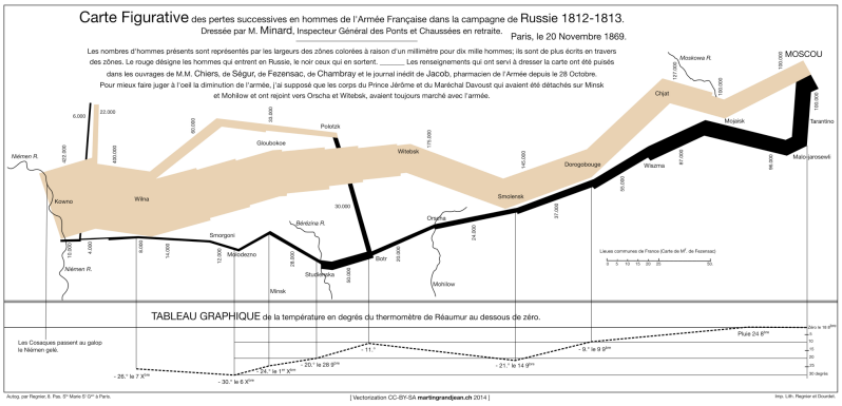
Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines or bars) contained in graphics. The goal is to communicate information clearly and efficiently to users. It is one of the steps in data analysis or data science. According to Friedman (2008) the "main goal of data visualization is to communicate information clearly and effectively through graphical means. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information".

Indeed, Fernanda Viegas and Martin M. Wattenberg have suggested that an ideal visualization should not only communicate clearly, but stimulate viewer engagement and attention.

Not limited to the communication of an information, a well-crafted data visualization is also a way to a better understanding of the data (in a data-driven research perspective), as it helps uncover trends, realize insights, explore sources, and tell stories.

Data visualization is closely related to information graphics, information visualization, scientific visualization, exploratory data analysis and statistical graphics. In the new millennium, data visualization has become an active area of research, teaching and development. According to Post et al. (2002), it has united scientific and information visualization.

Characteristics of Effective Graphical Displays



Charles Joseph Minard's 1869 diagram of Napoleon's March - an early example of an information graphic.

Professor Edward Tufte explained that users of information displays are executing particular *analytical tasks* such as making comparisons or determining causality. The *design principle* of the information graphic should support the analytical task, showing the comparison or causality.

In his 1983 book *The Visual Display of Quantitative Information*, Edward Tufte defines 'graphical displays' and principles for effective graphical display in the following passage: "Excellence in statistical graphics consists of complex ideas communicated with clarity, precision and efficiency. Graphical displays should:

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production or something else
- avoid distorting what the data has to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure

- serve a reasonably clear purpose: description, exploration, tabulation or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

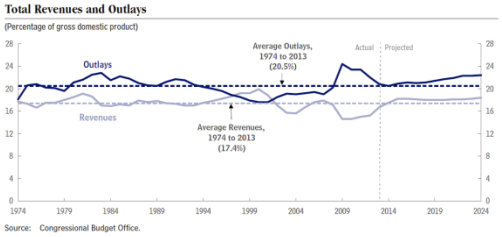
Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statistical computations."

For example, the Minard diagram shows the losses suffered by Napoleon's army in the 1812–1813 period. Six variables are plotted: the size of the army, its location on a two-dimensional surface (x and y), time, direction of movement, and temperature. The line width illustrates a comparison (size of the army at points in time) while the temperature axis suggests a cause of the change in army size. This multivariate display on a two dimensional surface tells a story that can be grasped immediately while identifying the source data to build credibility. Tufte wrote in 1983 that: "It may well be the best statistical graphic ever drawn."

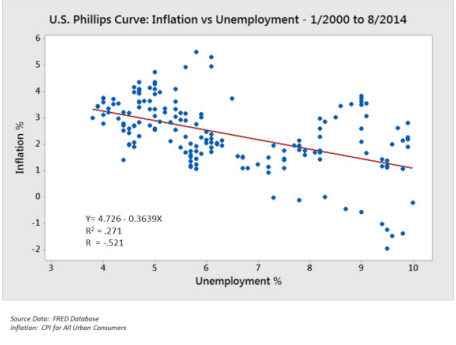
Not applying these principles may result in misleading graphs, which distort the message or support an erroneous conclusion. According to Tufte, chartjunk refers to extraneous interior decoration of the graphic that does not enhance the message, or gratuitous three dimensional or perspective effects. Needlessly separating the explanatory key from the image itself, requiring the eye to travel back and forth from the image to the key, is a form of "administrative debris." The ratio of "data to ink" should be maximized, erasing non-data ink where feasible.

The Congressional Budget Office summarized several best practices for graphical displays in a June 2014 presentation. These included: a) Knowing your audience; b) Designing graphics that can stand alone outside the context of the report; and c) Designing graphics that communicate the key messages in the report.

Quantitative Messages



A time series illustrated with a line chart demonstrating trends in U.S. federal spending and revenue over time.



A scatterplot illustrating negative correlation between two variables (inflation and unemployment) measured at points in time.

Author Stephen Few described eight types of quantitative messages that users may attempt to understand or communicate from a set of data and the associated graphs used to help communicate the message:

1. Time-series: A single variable is captured over a period of time, such as the unemployment rate over a 10-year period. A line chart may be used to demonstrate the trend.
2. Ranking: Categorical subdivisions are ranked in ascending or descending order, such as a ranking of sales performance (the *measure*) by sales persons (the *category*, with each sales person a *categorical subdivision*) during a single period. A bar chart may be used to show the comparison across the sales persons.
3. Part-to-whole: Categorical subdivisions are measured as a ratio to the whole (i.e., a percentage out of 100%). A pie chart or bar chart can show the comparison of ratios, such as the market share represented by competitors in a market.
4. Deviation: Categorical subdivisions are compared against a reference, such as a comparison of actual vs. budget expenses for several departments of a business for a given time period. A bar chart can show comparison of the actual versus the reference amount.
5. Frequency distribution: Shows the number of observations of a particular variable for given interval, such as the number of years in which the stock market return is between intervals such as 0-10%, 11-20%, etc. A histogram, a type of bar chart, may be used for this analysis. A boxplot helps visualize key statistics about the distribution, such as median, quartiles, outliers, etc.
6. Correlation: Comparison between observations represented by two variables (X,Y) to determine if they tend to move in the same or opposite directions. For example, plotting unemployment (X) and inflation (Y) for a sample of months. A scatter plot is typically used for this message.
7. Nominal comparison: Comparing categorical subdivisions in no particular order, such as the sales volume by product code. A bar chart may be used for this comparison.
8. Geographic or geospatial: Comparison of a variable across a map or layout, such as the unemployment rate by state or the number of persons on the various floors of a building. A cartogram is a typical graphic used.

Analysts reviewing a set of data may consider whether some or all of the messages and graphic types above are applicable to their task and audience. The process of trial and error to identify meaningful relationships and messages in the data is part of exploratory data analysis.

Visual Perception and Data Visualization

A human can distinguish differences in line length, shape orientation, and color (hue) readily without significant processing effort; these are referred to as "pre-attentive attributes." For example, it may require significant time and effort ("attentive processing") to identify the number of times the digit "5" appears in a series of numbers; but if that digit is different in size, orientation, or color, instances of the digit can be noted quickly through pre-attentive processing.

Effective graphics take advantage of pre-attentive processing and attributes and the relative strength of these attributes. For example, since humans can more easily process differences in line length than surface area, it may be more effective to use a bar chart (which takes advantage of line length to show comparison) rather than pie charts (which use surface area to show comparison).

Human Perception/Cognition and Data Visualization

There is a human side to data visualization. With the "studying [of] human perception and cognition ..." we are better able to understand the target of the data which we display. Cognition refers to processes in human beings like perception, attention, learning, memory, thought, concept formation, reading, and problem solving. The basis of data visualization evolved because as a picture is worth a thousand words, data displayed graphically allows for an easier comprehension of the information. Proper visualization provides a different approach to show potential connections, relationships, etc. which are not as obvious in non-visualized quantitative data. Visualization becomes a means of data exploration. Human brain neurons involve multiple functions but 2/3 of the brain's neurons are dedicated to vision. With a well-developed sense of sight, analysis of data can be made on data, whether that data is quantitative or qualitative. Effective visualization follows from understanding the processes of human perception and being able to apply this to intuitive visualizations is important. Understanding how humans see and organize the world is critical to effectively communicating data to the reader. This leads to more intuitive designs.

History of Data Visualization

There is a history of data visualization: beginning in the 2nd century C.E. with data arrangement into columns and rows and evolving to the initial quantitative representations in the 17th century. According to the Interaction Design Foundation, French philosopher and mathematician René Descartes laid the ground work for Scotsman William Playfair. Descartes developed a two-dimensional coordinate system for displaying values, which in the late 18th century Playfair saw potential for graphical communication of quantitative data. In the second half of the 20th century, Jacques Bertin used quantitative graphs to represent information "intuitively, clearly, accurately, and efficiently". John Tukey and more notably Edward Tufte pushed the bounds of data visualization. Tukey with his new statistical approach: exploratory data analysis and Tufte with his book "The Visual Display of Quantitative Information", the path was paved for refining data visualization techniques for more than statisticians. With the progression of technology came the progression of data visualization; starting with hand drawn visualizations and evolving into more technical applications – including interactive designs leading to software visualization. Programs like SAS, SOFA, R, Minitab, and more allow for data visualization in the field of statistics. Other data visualization applications, more focused and unique to individuals, programming languages such as D3, Python and JavaScript help to make the visualization of quantitative data a possibility.

Terminology

Data visualization involves specific terminology, some of which is derived from statistics. For example, author Stephen Few defines two types of data, which are used in combination to support a meaningful analysis or visualization:

- Categorical: Text labels describing the nature of the data, such as "Name" or "Age". This

term also covers qualitative (non-numerical) data.

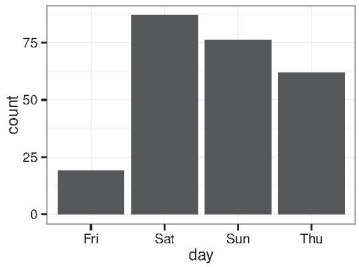
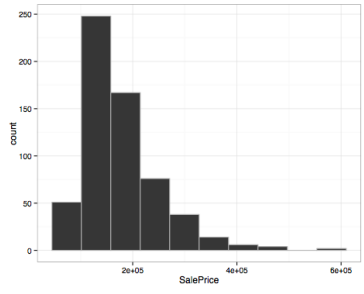
- Quantitative: Numerical measures, such as "25" to represent the age in years.

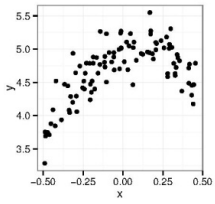
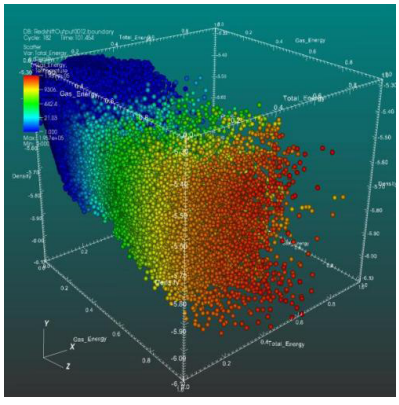
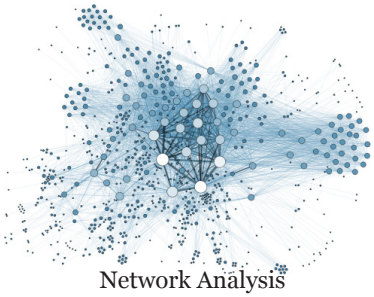
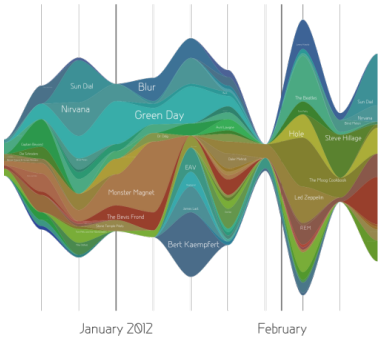
Two primary types of information displays are tables and graphs.

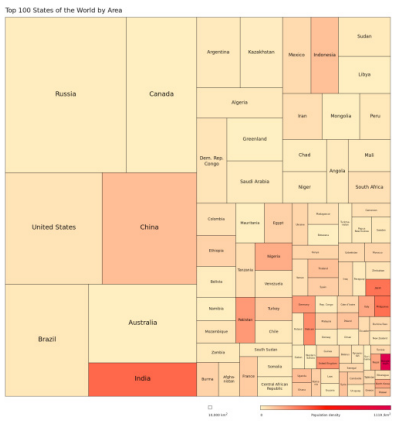
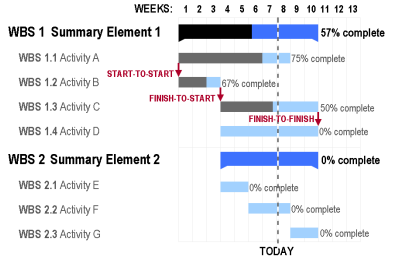
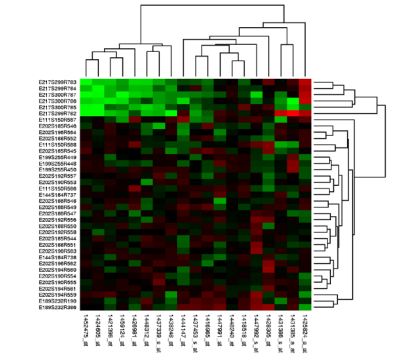
- A *table* contains quantitative data organized into rows and columns with categorical labels. It is primarily used to look up specific values. In the example above, the table might have categorical column labels representing the name (a *qualitative variable*) and age (a *quantitative variable*), with each row of data representing one person (the sampled *experimental unit* or *category subdivision*).
- A *graph* is primarily used to show relationships among data and portrays values encoded as *visual objects* (e.g., lines, bars, or points). Numerical values are displayed within an area delineated by one or more *axes*. These axes provide *scales* (quantitative and categorical) used to label and assign values to the visual objects. Many graphs are also referred to as *charts*.

KPI Library has developed the "Periodic Table of Visualization Methods," an interactive chart displaying various data visualization methods. It includes six types of data visualization methods: data, information, concept, strategy, metaphor and compound.

Examples of Diagrams used for Data Visualization

	Name	Visual Dimensions	Example Usages
 Bar chart of tips by day of week	Bar chart	• length/count • category • (color)	• Comparison of values, such as sales performance for several persons or businesses in a single time period. For a single variable measured over time (trend) a line chart is preferable.
 Histogram of housing prices	Histogram	• bin limits • count/length • (color)	• Determining frequency of annual stock market percentage returns within particular ranges (bins) such as 0-10%, 11-20%, etc. The height of the bar represents the number of observations (years) with a return % in the range represented by the bin.

 Basic scatterplot of two variables	Scatter plot	• x position • y position • (symbol/ glyph) • (color) • (size)	• Determining the relationship (e.g., correlation) between unemployment (x) and inflation (y) for multiple time periods.
 Scatter Plot	Scatter plot (3D)	• position x • position y • position z • color	
 Network Analysis	Network	• nodes size • nodes color • ties thickness • ties color • spatialization	• Finding clusters in the network (e.g. grouping Facebook friends into different clusters). • Determining the most influential nodes in the network (e.g. A company wants to target a small group of people on Twitter for a marketing campaign).
 Streamgraph	Streamgraph	• width • color • time (flow)	

 Treemap	Treemap	• size • color	• disk space by location / file type
 Gantt Chart	Gantt chart	• color • time (flow)	• schedule / progress, e.g. in project planning
 Heat map	Heat map	• row • column • cluster • color	• Analyzing risk, with green, yellow and red representing low, medium, and high risk, respectively.

Other Perspectives

There are different approaches on the scope of data visualization. One common focus is on information presentation, such as Friedman (2008) presented it. In this way Friendly (2008) presumes two main parts of data visualization: statistical graphics, and thematic cartography. In this line the "Data Visualization: Modern Approaches" (2007) article gives an overview of seven subjects of data visualization:

- Articles & resources
- Displaying connections
- Displaying data

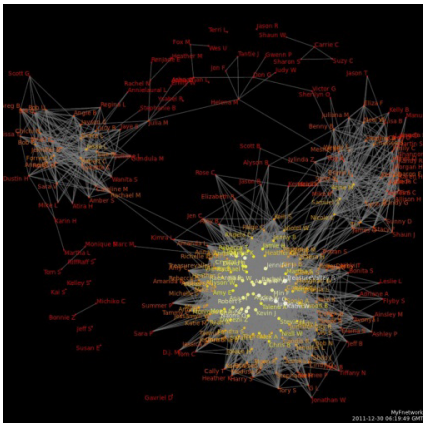
- Displaying news
- Displaying websites
- Mind maps
- Tools and services

All these subjects are closely related to graphic design and information representation.

On the other hand, from a computer science perspective, Frits H. Post (2002) categorized the field into a number of sub-fields:

- Information visualization
- Interaction techniques and architectures
- Modelling techniques
- Multiresolution methods
- Visualization algorithms and techniques
- Volume visualization

Data Presentation Architecture



A data visualization from social media

Data presentation architecture (DPA) is a skill-set that seeks to identify, locate, manipulate, format and present data in such a way as to optimally communicate meaning and proper knowledge.

Historically, the term *data presentation architecture* is attributed to Kelly Lautt: "Data Presentation Architecture (DPA) is a rarely applied skill set critical for the success and value of Business Intelligence. Data presentation architecture weds the science of numbers, data and statistics in discovering valuable information from data and making it usable, relevant and actionable with the arts of data visualization, communications, organizational psychology and change management in order to provide business intelligence solutions with the data scope, delivery timing, format and visualizations that will most effectively support and drive operational, tactical and strategic behaviour toward understood business (or organizational) goals. DPA is neither an IT nor a busi-

ness skill set but exists as a separate field of expertise. Often confused with data visualization, data presentation architecture is a much broader skill set that includes determining what data on what schedule and in what exact format is to be presented, not just the best way to present data that has already been chosen (which is data visualization). Data visualization skills are one element of DPA."

Objectives

DPA has two main objectives:

- To use data to provide knowledge in the most efficient manner possible (minimize noise, complexity, and unnecessary data or detail given each audience's needs and roles)
- To use data to provide knowledge in the most effective manner possible (provide relevant, timely and complete data to each audience member in a clear and understandable manner that conveys important meaning, is actionable and can affect understanding, behavior and decisions)

Scope

With the above objectives in mind, the actual work of data presentation architecture consists of:

- Creating effective delivery mechanisms for each audience member depending on their role, tasks, locations and access to technology
- Defining important meaning (relevant knowledge) that is needed by each audience member in each context
- Determining the required periodicity of data updates (the currency of the data)
- Determining the right timing for data presentation (when and how often the user needs to see the data)
- Finding the right data (subject area, historical reach, breadth, level of detail, etc.)
- Utilizing appropriate analysis, grouping, visualization, and other presentation formats

Related Fields

DPA work shares commonalities with several other fields, including:

- Business analysis in determining business goals, collecting requirements, mapping processes.
- Business process improvement in that its goal is to improve and streamline actions and decisions in furtherance of business goals
- Data visualization in that it uses well-established theories of visualization to add or highlight meaning or importance in data presentation.
- Graphic or user design: As the term DPA is used, it falls just short of design in that it does

not consider such detail as colour palates, styling, branding and other aesthetic concerns, unless these design elements are specifically required or beneficial for communication of meaning, impact, severity or other information of business value. For example:

  - choosing locations for various data presentation elements on a presentation page (such as in a company portal, in a report or on a web page) in order to convey hierarchy, priority, importance or a rational progression for the user is part of the DPA skill-set.
  - choosing to provide a specific colour in graphical elements that represent data of specific meaning or concern is part of the DPA skill-set

- Information architecture, but information architecture's focus is on unstructured data and therefore excludes both analysis (in the statistical/data sense) and direct transformation of the actual content (data, for DPA) into new entities and combinations.
- Solution architecture in determining the optimal detailed solution, including the scope of data to include, given the business goals
- Statistical analysis or data analysis in that it creates information and knowledge out of data

Data Profiling

Data profiling is the process of examining the data available in an existing information data source (e.g. a database or a file) and collecting statistics or small but informative summaries about that data. The purpose of these statistics may be to:

1. Find out whether existing data can easily be used for other purposes
2. Improve the ability to search the data by tagging it with keywords, descriptions, or assigning it to a category
3. Give metrics on data quality including whether the data conforms to particular standards or patterns
4. Assess the risk involved in integrating data for new applications, including the challenges of joins
5. Discover metadata of the source database, including value patterns and distributions, key candidates, foreign-key candidates, and functional dependencies
6. Assess whether known metadata accurately describes the actual values in the source database
7. Understanding data challenges early in any data intensive project, so that late project surprises are avoided. Finding data problems late in the project can lead to delays and cost overruns.
8. Have an enterprise view of all data, for uses such as master data management where key data is needed, or data governance for improving data quality.

Introduction

Data profiling refers to analyzing candidate data sources for a data warehouse in order to clarify the structure, content, relationships and derivation rules of the data. Profiling helps not only to understand anomalies and assess data quality, but also to discover, register, and assess enterprise metadata. Thus, the purpose of data profiling is both to validate metadata when it is available and to discover metadata when it is not. The result of the analysis is used both strategically, to determine suitability of the candidate source systems and give the basis for an early go/no-go decision, and tactically, to identify problems for later solution design, and to level sponsors' expectations.

How to do Data Profiling

Data profiling utilizes different kinds of descriptive statistics such as minimum, maximum, mean, mode, percentile, standard deviation, frequency, and variation as well as other aggregates such as count and sum. Additional metadata information obtained during data profiling could be the data type, length, discrete values, uniqueness, occurrence of null values, typical string patterns, and abstract type recognition. The metadata can then be used to discover problems such as illegal values, misspelling, missing values, varying value representation, and duplicates.

Different analyses are performed for different structural levels. E.g. single columns could be profiled individually to get an understanding of frequency distribution of different values, type, and use of each column. Embedded value dependencies can be exposed in a cross-columns analysis. Finally, overlapping value sets possibly representing foreign key relationships between entities can be explored in an inter-table analysis.

Normally, purpose-built tools are used for data profiling to ease the process. The computation complexity increases when going from single column, to single table, to cross-table structural profiling. Therefore, performance is an evaluation criterion for profiling tools.

When to Conduct Data Profiling

According to Kimball, data profiling is performed several times and with varying intensity throughout the data warehouse developing process. A light profiling assessment should be undertaken as soon as candidate source systems have been identified immediately after the acquisition of the DW/BI business requirements. The purpose is to clarify at an early stage if the right data is available at the appropriate detail level and that anomalies can be handled subsequently. If this is not the case the project may be terminated.

More detailed profiling is done prior to the dimensional modelling process in order to see what is required to convert data into the dimensional model. Detailed profiling extends into the ETL system design process in order to determine what data to extract and which filters to apply.

Additionally, data may be conducted in the data warehouse development process after data has been loaded into staging, the data marts, etc. Conducting data at these stages helps ensure that data cleaning and transformations have been done correctly according to requirements.

Benefits

The benefits of data profiling are to improve data quality, shorten the implementation cycle of major projects, and improve understanding of data for users. Discovering business knowledge embedded in data itself is one of the significant benefits derived from data profiling. Data profiling is one of the most effective technologies for improving data accuracy in corporate databases.

Although data profiling is effective and useful for each sector of our daily life, it can be challenging not to slip into "analysis paralysis".

Data Cleansing

Data cleansing, data cleaning, or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. Data cleansing may be performed interactively with data wrangling tools, or as batch processing through scripting.

After cleansing, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores. Data cleansing differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at the time of entry, rather than on batches of data.

The actual process of data cleansing may involve removing typographical errors or validating and correcting values against a known list of entities. The validation may be strict (such as rejecting any address that does not have a valid postal code) or fuzzy (such as correcting records that partially match existing, known records). Some data cleansing solutions will clean data by cross checking with a validated data set. A common data cleansing practice is data enhancement, where data is made more complete by adding related information. For example, appending addresses with any phone numbers related to that address. Data cleansing may also involve activities like, harmonization of data, and standardization of data. For example, harmonization of short codes (st, rd, etc.) to actual words (street, road, etcetera). Standardization of data is a means of changing a reference data set to a new standard, ex, use of standard codes.

Motivation

Administratively, incorrect or inconsistent data can lead to false conclusions and misdirected investments on both public and private scales. For instance, the government may want to analyze population census figures to decide which regions require further spending and investment on infrastructure and services. In this case, it will be important to have access to reliable data to avoid erroneous fiscal decisions.

In the business world, incorrect data can be costly. Many companies use customer information databases that record data like contact information, addresses, and preferences. For instance, if

the addresses are inconsistent, the company will suffer the cost of resending mail or even losing customers.

The profession of forensic accounting and fraud investigating uses data cleansing in preparing its data and is typically done before data is sent to a data warehouse for further investigation.

There are packages available so you can cleanse/wash address data while you enter it into your system. This is normally done via an API and will prompt staff as they type the address.

Data Quality

High-quality data needs to pass a set of quality criteria. Those include:

- Validity: The degree to which the measures conform to defined business rules or constraints. When modern database technology is used to design data-capture systems, validity is fairly easy to ensure: invalid data arises mainly in legacy contexts (where constraints were not implemented in software) or where inappropriate data-capture technology was used (e.g., spreadsheets, where it is very hard to limit what a user chooses to enter into a cell). Data constraints fall into the following categories:
    - *Data-Type Constraints* – e.g., values in a particular column must be of a particular datatype, e.g., Boolean, numeric (integer or real), date, etc.
    - *Range Constraints:* typically, numbers or dates should fall within a certain range. That is, they have minimum and/or maximum permissible values.
    - *Mandatory Constraints:* Certain columns cannot be empty.
    - *Unique Constraints:* A field, or a combination of fields, must be unique across a dataset. For example, no two persons can have the same social security number.
    - *Set-Membership constraints*: The values for a column come from a set of discrete values or codes. For example, a person's gender may be Female, Male or Unknown (not recorded).
    - *Foreign-key constraints*: This is the more general case of set membership. The set of values in a column is defined in a column of another table that contains unique values. For example, in a US taxpayer database, the "state" column is required to belong to one of the US's defined states or territories: the set of permissible states/territories is recorded in a separate States table. The term foreign key is borrowed from relational database terminology.
    - Regular expression patterns: Occasionally, text fields will have to be validated this way. For example, phone numbers may be required to have the pattern (999) 999-9999.
    - Cross-field validation: Certain conditions that utilize multiple fields must hold. For example, in laboratory medicine, the sum of the components of the differential white blood cell count must be equal to 100 (since they are all percentages). In a hospital database, a patient's date of discharge from hospital cannot be earlier than the date of admission.

- Decleansing is detecting errors and syntactically removing them for better programming.
- Accuracy: The degree of conformity of a measure to a standard or a true value. Accuracy is very hard to achieve through data-cleansing in the general case, because it requires accessing an external source of data that contains the true value: such "gold standard" data is often unavailable. Accuracy has been achieved in some cleansing contexts, notably customer contact data, by using external databases that match up zip codes to geographical locations (city and state), and also help verify that street addresses within these zip codes actually exist.
- Completeness: The degree to which all required measures are known. Incompleteness is almost impossible to fix with data cleansing methodology: one cannot infer facts that were not captured when the data in question was initially recorded. (In some contexts, e.g., interview data, it may be possible to fix incompleteness by going back to the original source of data, i,e., re-interviewing the subject, but even this does not guarantee success because of problems of recall - e.g., in an interview to gather data on food consumption, no one is likely to remember exactly what one ate six months ago. In the case of systems that insist certain columns should not be empty, one may work around the problem by designating a value that indicates "unknown" or "missing", but supplying of default values does not imply that the data has been made complete.
- Consistency: The degree to which a set of measures are equivalent in across systems. Inconsistency occurs when two data items in the data set contradict each other: e.g., a customer is recorded in two different systems as having two different current addresses, and only one of them can be correct. Fixing inconsistency is not always possible: it requires a variety of strategies - e.g., deciding which data were recorded more recently, which data source is likely to be most reliable (the latter knowledge may be specific to a given organization), or simply trying to find the truth by testing both data items (e.g., calling up the customer).
- Uniformity: The degree to which a set data measures are specified using the same units of measure in all systems. In datasets pooled from different locales, weight may be recorded either in pounds or kilos, and must be converted to a single measure using an arithmetic transformation.

The term Integrity encompasses accuracy, consistency and some aspects of validation but is rarely used by itself in data-cleansing contexts because it is insufficiently specific. (For example, "referential integrity" is a term used to refer to the enforcement of foreign-key constraints above.)

The Process of Data Cleansing

- Data auditing: The data is audited with the use of statistical and database methods to detect anomalies and contradictions: this eventually gives an indication of the characteristics of the anomalies and their locations. Several commercial software packages will let you specify constraints of various kinds (using a grammar that conforms to that of a standard programming language, e.g., JavaScript or Visual Basic) and then generate code that checks the data for violation of these constraints. This process is referred to below in the bullets "workflow specification" and "workflow execution." For users who lack access to high-end cleansing software, Microcomputer database packages such as Microsoft Access

or File Maker Pro will also let you perform such checks, on a constraint-by-constraint basis, interactively with little or no programming required in many cases.

- Workflow specification: The detection and removal of anomalies is performed by a sequence of operations on the data known as the workflow. It is specified after the process of auditing the data and is crucial in achieving the end product of high-quality data. In order to achieve a proper workflow, the causes of the anomalies and errors in the data have to be closely considered.
- Workflow execution: In this stage, the workflow is executed after its specification is complete and its correctness is verified. The implementation of the workflow should be efficient, even on large sets of data, which inevitably poses a trade-off because the execution of a data-cleansing operation can be computationally expensive.
- Post-processing and controlling: After executing the cleansing workflow, the results are inspected to verify correctness. Data that could not be corrected during execution of the workflow is manually corrected, if possible. The result is a new cycle in the data-cleansing process where the data is audited again to allow the specification of an additional workflow to further cleanse the data by automatic processing.

Good quality source data has to do with "Data Quality Culture" and must be initiated at the top of the organization. It is not just a matter of implementing strong validation checks on input screens, because almost no matter how strong these checks are, they can often still be circumvented by the users. There is a nine-step guide for organizations that wish to improve data quality:

- Declare a high level commitment to a data quality culture
- Drive process reengineering at the executive level
- Spend money to improve the data entry environment
- Spend money to improve application integration
- Spend money to change how processes work
- Promote end-to-end team awareness
- Promote interdepartmental cooperation
- Publicly celebrate data quality excellence
- Continuously measure and improve data quality

Decleanse

Parsing: for the detection of syntax errors. A parser decides whether a string of data is acceptable within the allowed data specification. This is similar to the way a parser works with grammars and languages.

- Data transformation: Data transformation allows the mapping of the data from its given format into the format expected by the appropriate application. This includes value conversions or translation functions, as well as normalizing numeric values to conform to minimum and maximum values.

- Duplicate elimination: Duplicate detection requires an algorithm for determining whether data contains duplicate representations of the same entity. Usually, data is sorted by a key that would bring duplicate entries closer together for faster identification.
- Statistical methods: By analyzing the data using the values of mean, standard deviation, range, or clustering algorithms, it is possible for an expert to find values that are unexpected and thus erroneous. Although the correction of such data is difficult since the true value is not known, it can be resolved by setting the values to an average or other statistical value. Statistical methods can also be used to handle missing values which can be replaced by one or more plausible values, which are usually obtained by extensive data augmentation algorithms.

Data Cleansing System

The essential job of this system is to find a suitable balance between fixing dirty data and maintaining the data as close as possible to the original data from the source production system. This is a challenge for the Extract, transform, load architect.

The system should offer an architecture that can cleanse data, record quality events and measure/control quality of data in the data warehouse.

A good start is to perform a thorough data profiling analysis that will help define to the required complexity of the data cleansing system and also give an idea of the current data quality in the source system(s).

Quality Screens

Part of the data cleansing system is a set of diagnostic filters known as quality screens. They each implement a test in the data flow that, if it fails records an error in the Error Event Schema. Quality screens are divided into three categories:

- Column screens. Testing the individual column, e.g. for unexpected values like NULL values; non-numeric values that should be numeric; out of range values; etc.
- Structure screens. These are used to test for the integrity of different relationships between columns (typically foreign/primary keys) in the same or different tables. They are also used for testing that a group of columns is valid according to some structural definition it should adhere.
- Business rule screens. The most complex of the three tests. They test to see if data, maybe across multiple tables, follow specific business rules. An example could be, that if a customer is marked as a certain type of customer, the business rules that define this kind of customer should be adhered.

When a quality screen records an error, it can either stop the dataflow process, send the faulty data somewhere else than the target system or tag the data. The latter option is considered the best solution because the first option requires, that someone has to manually deal with the issue each time it occurs and the second implies that data are missing from the target system (integrity) and

it is often unclear, what should happen to these data.

Criticism of Existing Tools and Processes

The main reasons cited are:

- Project costs: costs typically in the hundreds of thousands of dollars
- Time: lack of enough time to deal with large-scale data-cleansing software
- Security: concerns over sharing information, giving an application access across systems, and effects on legacy systems

Error Event Schema

This schema is the place, where all error events thrown by quality screens, are recorded. It consists of an Error Event Fact table with foreign keys to three dimension tables that represent date (when), batch job (where) and screen (who produced error). It also holds information about exactly when the error occurred and the severity of the error. In addition there is an Error Event Detail Fact table with a foreign key to the main table that contains detailed information about in which table, record and field the error occurred and the error condition.

Challenges and Problems

- Error correction and loss of information: The most challenging problem within data cleansing remains the correction of values to remove duplicates and invalid entries. In many cases, the available information on such anomalies is limited and insufficient to determine the necessary transformations or corrections, leaving the deletion of such entries as a primary solution. The deletion of data, though, leads to loss of information; this loss can be particularly costly if there is a large amount of deleted data.
- Maintenance of cleansed data: Data cleansing is an expensive and time-consuming process. So after having performed data cleansing and achieving a data collection free of errors, one would want to avoid the re-cleansing of data in its entirety after some values in data collection change. The process should only be repeated on values that have changed; this means that a cleansing lineage would need to be kept, which would require efficient data collection and management techniques.
- Data cleansing in virtually integrated environments: In virtually integrated sources like IBM's DiscoveryLink, the cleansing of data has to be performed every time the data is accessed, which considerably increases the response time and lowers efficiency.
- Data-cleansing framework: In many cases, it will not be possible to derive a complete data-cleansing graph to guide the process in advance. This makes data cleansing an iterative process involving significant exploration and interaction, which may require a framework in the form of a collection of methods for error detection and elimination in addition to data auditing. This can be integrated with other data-processing stages like integration and maintenance.

Process Mining

Process mining is a process management technique that allows for the analysis of business processes based on event logs. During process mining, specialized data-mining algorithms are applied to event log datasets in order to identify trends, patterns and details contained in event logs recorded by an information system. Process mining aims to improve process efficiency and understanding of processes. Process mining is also known as *Automated Business Process Discovery* (ABPD).

Overview

Process mining techniques are often used when no formal description of the process can be obtained by other approaches, or when the quality of existing documentation is questionable. For example, application of process mining methodology to the audit trails of a workflow management system, the transaction logs of an enterprise resource planning system, or the electronic patient records in a hospital can result in models describing processes, organizations, and products. Event log analysis can also be used to compare event logs with *prior* model(s) to understand whether the observations conform to a prescriptive or descriptive model.

Contemporary management trends such as BAM (Business Activity Monitoring), BOM (Business Operations Management), and BPI (business process intelligence) illustrate the interest in supporting diagnosis functionality in the context of Business Process Management technology (e.g., Workflow Management Systems and other *process-aware* information systems).

Application

Process mining follows the options established in business process engineering, then goes beyond those options by providing feedback for business process modeling:

- process analysis filters, orders and compresses logfiles for further insight into the connex of process operations.
- process design may be supported by feedback from process monitoring (action or event recording or logging)
- process enactment uses results from process mining based on logging for triggering further process operations

Classification

There are three classes of process mining techniques. This classification is based on whether there is a prior model and, if so, how the prior model is used during process mining.

- *Discovery*: Previous (*a priori*) models do not exist. Based on an event log, a new model is constructed or discovered based on low-level events. For example, using the alpha algorithm (a didactically driven approach). Many established techniques exist for automatically constructing process models (for example, Petri net, pi-calculus expression) based on an event log. Recently, process mining research has started targeting the other perspectives

(e.g., data, resources, time, etc.). One example is the technique described in (Aalst, Reijers, & Song, 2005), which can be used to construct a social network.

- *Conformance checking*: Used when there is an *a priori* model. The existing model is compared with the process event log; discrepancies between the log and the model are analyzed. For example, there may be a process model indicating that purchase orders of more than 1 million Euro require two checks. Another example is the checking of the so-called "four-eyes" principle. Conformance checking may be used to detect deviations to enrich the model. An example is the extension of a process model with performance data, i.e., some *a priori* process model is used to project the potential bottlenecks. Another example is the *decision miner* described in (Rozinat & Aalst, 2006b) which takes an *a priori* process model and analyzes every choice in the process model. For each choice the event log is consulted to see which information is typically available the moment the choice is made. Then classical data mining techniques are used to see which data elements influence the choice. As a result, a decision tree is generated for each choice in the process.
- *Extension*: Used when there is an *a priori* model. The model is extended with a new aspect or perspective, so that the goal is *not* to check conformance, but rather to improve the existing model. An example is the extension of a process model with performance data, i.e., some prior process model dynamically annotated with performance data.

Software for Process Mining

A software framework for the evaluation of process mining algorithms has been developed at the Eindhoven University of Technology by Wil van der Aalst and others, and is available as an open source toolkit.

- Process Mining
- ProM Framework
- ProM Import Framework

Process Mining functionality is also offered by the following commercial vendors:

- Interstage Automated Process Discovery, a Process Mining service offered by Fujitsu, Ltd. as part of the Interstage Integration Middleware Suite.
- Disco is a complete Process Mining software by Fluxicon.
- ARIS Process Performance Manager, a Process Mining and Process Intelligence Tool offered by Software AG as part of the Process Intelligence Solution.
- QPR ProcessAnalyzer, Process Mining software for Automated Business Process Discovery (ABPD).
- Perceptive Process Mining, the Process Mining solution by Perceptive Software (formerly Futura Reflect / Pallas Athena Reflect).
- Celonis Process Mining, the Process Mining solution offered by Celonis
- SNP Business Process Analysis, the SAP-focused Process Mining solution by SNP Schnei-

der-Neureither & Partner AG

- minit is a Process Mining software offered by Gradient ECM
- myInvenio cloud and on-premises solution by Cognitive Technology Ltd.
- LANA is a process mining tool featuring discovery and conformance checking.
- ProcessGold Enterprise Platform, an integration of Process Mining & Business Intelligence.

Competitive Intelligence

Competitive intelligence is the action of defining, gathering, analyzing, and distributing intelligence about products, customers, competitors, and any aspect of the environment needed to support executives and managers making strategic decisions for an organization.

Competitive intelligence essentially means understanding and learning what is happening in the world outside the business so one can be as competitive as possible. It means learning as much as possible, as soon as possible, about one's industry in general, one's competitors, or even one's county's particular zoning rules. In short, it empowers anticipating and facing challenges head on.

Key points of this definition:

1. Competitive intelligence is an ethical and legal business practice, as opposed to industrial espionage, which is illegal.
2. The focus is on the external business environment
3. There is a process involved in gathering information, converting it into intelligence and then using it in decision making. Some CI professionals erroneously emphasise that if the intelligence gathered is not usable or actionable, it is not intelligence.

A more focused definition of CI regards it as the organizational function responsible for the early identification of risks and opportunities in the market before they become *obvious*. Experts also call this process the early signal analysis. This definition focuses attention on the difference between dissemination of widely available factual information (such as market statistics, financial reports, newspaper clippings) performed by functions such as libraries and information centers, and competitive intelligence which is a *perspective* on developments and events aimed at yielding a competitive edge.

The term CI is often viewed as synonymous with competitor analysis, but competitive intelligence is more than analyzing competitors: it is about making the organization more competitive relative to its entire environment and stakeholders: customers, competitors, distributors, technologies, and macroeconomic data.

Historic Development

The literature associated with the field of competitive intelligence is best exemplified by the detailed bibliographies that were published in the Society of Competitive Intelligence Professionals' refereed academic journal called *The Journal of Competitive Intelligence and Management*. Al-

though elements of organizational intelligence collection have been a part of business for many years, the history of competitive intelligence arguably began in the U.S. in the 1970s, although the literature on the field pre-dates this time by at least several decades. In 1980, Michael Porter published the study *Competitive-Strategy: Techniques for Analyzing Industries and Competitors* which is widely viewed as the foundation of modern competitive intelligence. This has since been extended most notably by the pair of Craig Fleisher and Babette Bensoussan, who through several popular books on competitive analysis have added 48 commonly applied competitive intelligence analysis techniques to the practitioner's tool box. In 1985, Leonard Fuld published his best selling book dedicated to competitor intelligence. However, the institutionalization of CI as a formal activity among American corporations can be traced to 1988, when Ben and Tamar Gilad published the first organizational model of a formal corporate CI function, which was then adopted widely by US companies. The first professional certification program (CIP) was created in 1996 with the establishment of The Fuld-Gilad-Herring Academy of Competitive Intelligence in Cambridge, Massachusetts.

In 1986 the *Society of Competitive Intelligence Professionals* (SCIP) was founded in the United States and grew in the late 1990s to around 6,000 members worldwide, mainly in the United States and Canada, but with large numbers especially in the UK and Australia. Due to financial difficulties in 2009, the organization merged with Frost & Sullivan under the Frost & Sullivan Institute. SCIP has since been renamed "Strategic & Competitive Intelligence Professionals" to emphasise the strategic nature of the subject, and also to refocus the organisation's general approach, while keeping the existing SCIP brandname and logo. A number of efforts have been made to discuss the field's advances in post-secondary (university) education, covered by several authors including Blenkhorn & Fleisher, Fleisher, Fuld, Prescott, and McGonagle. Although the general view would be that competitive intelligence concepts can be readily found and taught in many business schools around the globe, there are still relatively few dedicated academic programs, majors, or degrees in the field, a concern to academics in the field who would like to see it further researched. These issues were widely discussed by over a dozen knowledgeable individuals in a special edition of the Competitive Intelligence Magazine that was dedicated to this topic. In France, a Specialized Master in Economic Intelligence and Knowledge Management was created in 1995 within the CERAM Business School, now SKEMA Business School, in Paris, with the objective of delivering a full and professional training in Economic Intelligence. A Centre for Global Intelligence and Influence was created in September 2011 in the same School.

On the other hand, practitioners and companies regard professional accreditation as especially important in this field. In 2011, SCIP recognized the Fuld-Gilad-Herring Academy of Competitive Intelligence's CIP certification process as its global, dual-level (CIP-I and CIP-II) certification program.

Global developments have also been uneven in competitive intelligence. Several academic journals, particularly the *Journal of Competitive Intelligence and Management* in its third volume, provided coverage of the field's global development. For example, in 1997 the *École de guerre économique* (fr) *(*School of economic warfare*) was founded in Paris, France. It is the first European institution which teaches the tactics of economic warfare within a globalizing world. In Germany, competitive intelligence was unattended until the early 1990s. The term "competitive intelligence" first appeared in German literature in 1997. In 1995 a German SCIP chapter was*

founded, which is now second in terms of membership in Europe. In summer 2004 the Institute for Competitive Intelligence was founded, which provides a postgraduate certification program for Competitive Intelligence Professionals. Japan is currently the only country that officially maintains an economic intelligence agency (JETRO). It was founded by the Ministry of International Trade and Industry (MITI) in 1958.

Accepting the importance of competitive intelligence, major multinational corporations, such as ExxonMobil, Procter & Gamble, and Johnson and Johnson, have created formal CI units. Importantly, organizations execute competitive intelligence activities not only as a safeguard to protect against market threats and changes, but also as a method for finding new opportunities and trends.

Organizations use competitive intelligence to compare themselves to other organizations ("competitive benchmarking"), to identify risks and opportunities in their markets, and to pressure-test their plans against market response (war gaming), which enable them to make informed decisions. Most firms today realize the importance of knowing what their competitors are doing and how the industry is changing, and the information gathered allows organizations to understand their strengths and weaknesses.

One of the major activities involved in corporate competitive intelligence is use of ratio analysis, using key performance indicators (KPI). Organizations compare annual reports of their competitors on certain KPI and ratios, which are intrinsic to their industry. This helps them track their performance, vis-a-vis their competitors.

The actual importance of these categories of information to an organization depends on the contestability of its markets, the organizational culture, the personality and biases of its top decision makers, and the reporting structure of competitive intelligence within the company.

Strategic Intelligence (SI) focuses on the longer term, looking at issues affecting a company's competitiveness over the course of a couple of years. The actual time horizon for SI ultimately depends on the industry and how quickly it's changing. The general questions that SI answers are, 'Where should we as a company be in X years?' and 'What are the strategic risks and opportunities facing us?' This type of intelligence work involves among others the identification of weak signals and application of methodology and process called Strategic Early Warning (SEW), first introduced by Gilad, followed by Steven Shaker and Victor Richardson, Alessandro Comai and Joaquin Tena, and others. According to Gilad, 20% of the work of competitive intelligence practitioners should be dedicated to strategic early identification of weak signals within a SEW framework.

Tactical Intelligence: the focus is on providing information designed to improve shorter-term decisions, most often related with the intent of growing market share or revenues. Generally, it is the type of information that you would need to support the sales process in an organization. It investigates various aspects of a product/product line marketing:

- Product – what are people selling?
- Price – what price are they charging?
- Promotion – what activities are they conducting for promoting this product?
- Place – where are they selling this product?

- Other – sales force structure, clinical trial design, technical issues, etc.

With the right amount of information, organizations can avoid unpleasant surprises by anticipating competitors' moves and decreasing response time. Examples of competitive intelligence research is evident in daily newspapers, such as the *Wall Street Journal*, *Business Week*, and *Fortune*. Major airlines change hundreds of fares daily in response to competitors' tactics. They use information to plan their own marketing, pricing, and production strategies.

Resources, such as the Internet, have made gathering information on competitors easy. With a click of a button, analysts can discover future trends and market requirements. However competitive intelligence is much more than this, as the ultimate aim is to lead to competitive advantage. As the Internet is mostly public domain material, information gathered is less likely to result in insights that will be unique to the company. In fact there is a risk that information gathered from the Internet will be misinformation and mislead users, so competitive intelligence researchers are often wary of using such information.

As a result, although the Internet is viewed as a key source, most CI professionals should spend their time and budget gathering intelligence using primary research—networking with industry experts, from trade shows and conferences, from their own customers and suppliers, and so on. Where the Internet is used, it is to gather sources for primary research as well as information on what the company says about itself and its online presence (in the form of links to other companies, its strategy regarding search engines and online advertising, mentions in discussion forums and on blogs, etc.). Also, important are online subscription databases and news aggregation sources which have simplified the secondary source collection process. Social media sources are also becoming important—providing potential interviewee names, as well as opinions and attitudes, and sometimes breaking news (e.g., via Twitter).

Organizations must be careful not to spend too much time and effort on old competitors without realizing the existence of any new competitors. Knowing more about your competitors will allow your business to grow and succeed. The practice of competitive intelligence is growing every year, and most companies and business students now realize the importance of knowing their competitors.

According to Arjan Singh and Andrew Beurschgens in their 2006 article in the *Competitive Intelligence Review*, there are four stages of development of a competitive intelligence capability with a firm. It starts with "stick fetching", where a CI department is very reactive, up to "world class", where it is completely integrated in the decision-making process.

Recent Trends

The technical advances in massive parallel processing offered by the Hadoop "big data" architecture has allowed the creation of multiple platforms for named-entity recognition such as the Apache Projects OpenNLP and Apache Stanbol. The former includes pre-trained statistical parsers that can discern elements key to establishing trends and evaluating competitive position and responding appropriately. Public information mining from *SEC.gov*, Federal Contract Awards, social media (Twitter, Reddit, Facebook, and others), vendors, and competitor websites now permit real-time counterintelligence as a strategy for horizontal and vertical market expansion and

product positioning. This occurs in an automated fashion on massive marketplaces such as Amazon.com and their classification and prediction of product associations and purchase probability.

Similar Fields

Competitive intelligence has been influenced by national strategic intelligence. Although national intelligence was researched 50 years ago, competitive intelligence was introduced during the 1990s. Competitive-intelligence professionals can learn from national-intelligence experts, especially in the analysis of complex situations. Competitive intelligence may be confused with (or seen to overlap) environmental scanning, business intelligence and market research. Craig Fleisher questions the appropriateness of the term, comparing it to business intelligence, competitor intelligence, knowledge management, market intelligence, marketing research and strategic intelligence.

Fleisher suggests that business intelligence has two forms. Its narrow (contemporary) form is more focused on information technology and internal focus than CI, while its broader (historical) definition is more inclusive than CI. Knowledge management (KM), when improperly achieved, is seen as an information-technology driven organizational practice relying on data mining, corporate intranets and mapping organizational assets to make it accessible to organization members for decision-making. CI shares some aspects of KM; they are human-intelligence- and experience-based for a more-sophisticated qualitative analysis. KM is essential for effective change. A key effective factor is a powerful, dedicated IT system executing the full intelligence cycle.

Market intelligence (MI) is industry-targeted intelligence developed in real-time aspects of competitive events taking place among the four Ps of the marketing mix (pricing, place, promotion and product) in the product (or service) marketplace to better understand the market's attractiveness. A time-based competitive tactic, MI is used by marketing and sales managers to respond to consumers more quickly in the marketplace. Fleisher suggests it is not distributed as widely as some forms of CI, which are also distributed to non-marketing decision-makers. Market intelligence has a shorter time horizon than other intelligence areas, and is measured in days, weeks, or (in slower-moving industries) months.

Market research is a tactical, method-driven field consisting of neutral, primary research of customer data (beliefs and perceptions) gathered in surveys or focus groups, and is analyzed with statistical-research techniques. CI draws on a wider variety (primary and secondary) of sources from a wider range of stakeholders (suppliers, competitors, distributors, substitutes and media) to answer existing questions, raise new ones and guide action.

Ben Gilad and Jan Herring lay down a set of prerequisites defining CI, distinguishing it from other information-rich disciplines such as market research or business development. They show that a common body of knowledge and a unique set of tools (key intelligence topics, business war games and blindspots analysis) distinguish CI; while other sensory activities in a commercial firm focus on *one* segment of the market (customers, suppliers or acquisition targets), CI synthesizes data from all high-impact players (HIP).

Gilad later focused his delineation of CI on the difference between information and intelligence. According to him, the common denominator among organizational sensory functions (whether they are called market research, business intelligence or market intelligence) is that they deliver

information rather than intelligence. Intelligence, says Gilad, is a perspective on facts rather than the facts themselves. Unique among corporate functions, competitive intelligence has a perspective of risks and opportunities for a firm's performance; as such, it (not information activities) is part of an organization's risk-management activity.

Ethics

Ethics has been a long-held issue of discussion among CI practitioners. Essentially, the questions revolve around what is and is not allowable in terms of CI practitioners' activity. A number of excellent scholarly treatments have been generated on this topic, most prominently addressed through Society of Competitive Intelligence Professionals publications. The book *Competitive Intelligence Ethics: Navigating the Gray Zone* provides nearly twenty separate views about ethics in CI, as well as another 10 codes used by various individuals or organizations. Combining that with the over two dozen scholarly articles or studies found within the various CI bibliographic entries, it is clear that no shortage of study has gone into better classifying, understanding and addressing CI ethics.

Competitive information may be obtained from public or subscription sources, from networking with competitor staff or customers, disassembly of competitor products or from field research interviews. Competitive intelligence research is distinguishable from industrial espionage, as CI practitioners generally abide by local legal guidelines and ethical business norms.

Outsourcing

Outsourcing has become a big business for competitive intelligence professionals. Companies like Aperio Intelligence, Cast Intelligence, Fuld & Co, Black Cube, Securitas and Emissary Investigative Services offer corporate intelligence services.

References

- Ward, J. & Peppard, J. (2002). Situation Analysis. In Strategic Planning for information systems. (pp. 82 - 83). England: John Wiley & Sons. ISBN 978-0-470-84147-1
- D'Atri A., De Marco M., Casalino N. (2008). "Interdisciplinary Aspects of Information Systems Studies", Physica-Verlag, Springer, Germany, pp. 1–416, doi:10.1007/978-3-7908-2010-2 ISBN 978-3-7908-2009-6
- Verna Allee, The Knowledge Evolution: Expanding Organizational Intelligence, Butterworth-Heinemann (1997) ISBN 0-7506-9842-X
- David Perkins, King Arthur's Round Table: How Collaborative Conversations Create Smart Organizations, Wiley (2002) ISBN 0-4712-3772-8
- Tufte, Edward (1983). The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press. ISBN 0-9613921-4-2.
- [David Loshin (2003), "Business Intelligence: The Savvy Manager's Guide, Getting Onboard with Emerging IT", Morgan Kaufmann Publishers, ISBN 9781558609167].
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., Becker, B. The Data Warehouse Lifecycle Toolkit, Wiley Publishing, Inc., 2008. ISBN 978-0-470-14977-5
- McGonagle, John J. and Carolyn M. Vella (2003). The Manager's Guide to Competitive Intelligence. Westport CT: Greenwood Publishing Group. p. 184. ISBN 1567205712.

7

Operational Intelligence: Technological Components

Operational intelligence has a number of aspects that have been elucidated in this chapter. Some of these features are complex event processing, business process management, metadata and root cause analysis. The components discussed in this text are of great importance to broaden the existing knowledge on operational intelligence.

Operational Intelligence

Operational intelligence (OI) is a category of real-time dynamic, business analytics that delivers visibility and insight into data, streaming events and business operations. Operational Intelligence solutions run queries against streaming data feeds and event data to deliver real-time analytic results as operational instructions. Operational Intelligence provides organizations the ability to make decisions and immediately act on these analytic insights, through manual or automated actions.

Purpose

The purpose of OI is to monitor business activities and identify and detect situations relating to inefficiencies, opportunities, and threats and provide operational solutions. Some definitions define operational intelligence an event-centric approach to delivering information that empowers people to make better decisions.

In addition, these metrics act as the starting point for further analysis (drilling down into details, performing root cause analysis — tying anomalies to specific transactions and of the business activity).

Sophisticated OI systems also provide the ability to associate metadata with metrics, process steps, channels, etc. With this, it becomes easy to get related information, e.g., “retrieve the contact information of the person that manages the application that executed the step in the business transaction that took 60% more time than the norm,” or “view the acceptance/rejection trend for the customer who was denied approval in this transaction,” or “Launch the application that this process step interacted with.”

Features

Different operational intelligence solutions may use many different technologies and be implemented in different ways. This section lists the common features of an operational intelligence solution:

- Real-time monitoring
- Real-time situation detection
- Real-time dashboards for different user roles
- Correlation of events
- Industry-specific dashboards
- Multidimensional analysis
    - Root cause analysis
    - Time Series and trend analysis
- Big Data Analytics: Operational Intelligence is well suited to address the inherent challenges of Big Data. Operational Intelligence continuously monitors and analyzes the variety of high velocity, high volume Big Data sources. Often performed in memory, OI platforms and solutions then present the incremental calculations and changes, in real-time, to the end-user.

Technology Components

Operational intelligence solutions share many features, and therefore many also share technology components. This is a list of some of the commonly found technology components, and the features they enable:

- Business activity monitoring (BAM) - Dashboard customization and personalization
- Complex event processing (CEP) - Advanced, continuous analysis of real-time information and historical data
- Business process management (BPM) - To perform model-driven execution of policies and processes defined as Business Process Model and Notation (BPMN) models
- Metadata framework to model and link events to resources
- Multi-channel publishing and notification
- Dimensional database
- Root cause analysis
- Multi-protocol event collection

Operational intelligence is a relatively new market segment (compared to the more mature business intelligence and business process management segments). In addition to companies that produce dedicated and focussed products in this area, there are numerous companies in adjacent areas that provide solutions with some OI components.

Operational intelligence places complete information at one's fingertips, enabling one to make smarter decisions in time to maximize impact. By correlating a wide variety of events and data from both streaming feeds and historical data silos, operational intelligence helps orga-