

# What to Teach Business Students in MIS Courses about Data and Information

*Zbigniew J Gackowski*  
*California State University Stanislaus, Turlock, California*

[zgacko@koko.csustan.edu](mailto:zgacko@koko.csustan.edu)

## Abstract

MIS textbooks on the market are overly technology laden, with oversimplified coverage of the fundamentals in general and in particular with regard to the role of data and information in business. This thought-provoking paper is born out of frustration with the disparity between what we already know on the subject and what actually is being offered to our students and clients. The first part of this paper examines of the most popular MIS textbooks and the empirical research on information quality, Besides four other referenced sources, two J. O'Brien's textbooks: one with six editions (O'Brien, 2004) and another with 11 editions (O'Brien, 2003) are used as the main frames of reference. The book "Quality Information and Knowledge" by Huang, Lee, and Wang (1999) serves as a recognized research framework. It is a product of the Information Quality Programs & Initiatives at MIT – the so-called MITIQ Program. The second part of the paper contains suggestions what to teach business students on data and information.

**Keywords:** Computer data, information, knowledge, MIS courses.

## Introduction

"The purpose of Informing Science is to encourage the sharing of knowledge and collaboration among the wide variety of fields that use information technology to inform clients" ([INSITE 2004](#)). From the viewpoint of instructors who teach Management Information Systems (MIS) courses to business students, those students are clients and future end users of such systems. The focus of this paper is a critical inquiry into what the most popular textbooks offer and a proposal of what one should teach business students about data and information and their role in business within the context of the wide proliferation of information technology. Obviously, the problem intersects many disciplines, such as business communications, data communications, information technology, education, management theory, and management science.

Having taught MIS courses to business students for several decades, this author has come to realize that the current textbooks are particularly deficient with regard to the role of end users. The texts are overly technology laden, with oversimplified coverage of the fundamentals in general

---

Material published as part of this journal, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at [Publisher@InformingScience.org](mailto:Publisher@InformingScience.org)

and in particular with regard to the role of data and information in business. A summary of research published by Huang et al., (1999 p. 4) supports this statement in the following words; "Many best-practice reports witness that information technology alone is not the driver for

knowledge management in companies today... Information and knowledge experienced by members of an organization should be the focus, not the system or technology per se. Technology and systems ... are facilitators”.

This paper is born out of frustration with the disparity between what we already know on the subject and what actually is being offered to our students and clients. Two of J. O’Brien’s textbooks: “[Management Information Systems](#)” (O’Brien, 2004), with six editions, and his “[Introduction to Computer Information Systems](#)” (O’Brien, 2003), with eleven editions, are used as the main frames of reference. Due to the exceptionally high number of editions, they are probably the most popular ones. The number of editions is a pretty simple but reliable indicator of popularity and the current trends. It clearly indicates that such a textbook has been on the market for a long time, was and still is actively used by instructors, and has been usually challenged by at least three reviewers before the publisher produced each edition. Additional references are made to Dock and Wetherbe (1988), Alter (2002), Post and Anderson (2003), and Malaga (2005). For the sake of brevity of this paper, it is practically impossible to include a broader survey of available MIS textbooks.

With regard to research in this field, “Quality Information and Knowledge” by K. Huang, Y. W. Lee and R. Y. Wang (1999) serves as a recognized framework. This book is a product of the Information Quality Programs & Initiatives at MIT – the so-called [MITIQ Program](#) (<http://mitiq.mit.edu>).

The main purpose of this paper is to show that the collective experiences and findings from first-class empirical studies about information quality should not only be included into current textbooks, but also complemented by a rational examination and analysis of the mutual interdependencies of the multidimensional aspects of information quality. Thus, we can arrive at a fuller understanding of the subject, a more complete frame of reference, a better understanding of the phenomenon of data and information quality within business environments, and at some useful guidelines on how to economize the examination of its corresponding attributes of quality.

This paper is divided into two main parts:

- First, an examination of the most popular MIS textbooks and the empirical research on information quality, and
- Second, a proposal of what one should teach business students in MIS courses on data and information.

Both parts represent a strongly contrasting view of the subject.

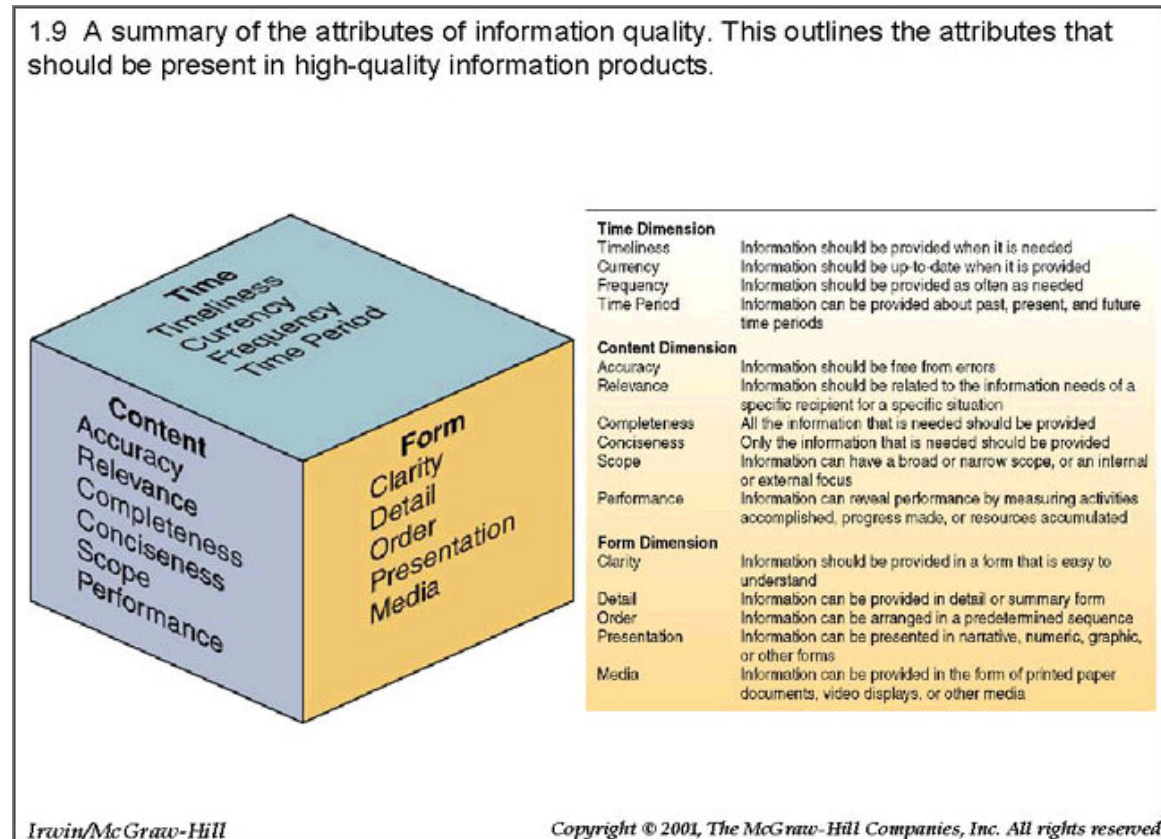
## **What MIS Textbooks and Empirical Research Offer on Data/Information Quality**

### ***What do the most popular textbooks offer?***

O’Brien (2004 p. 15-16), the author of the two most popular textbooks on MIS judged by the number of editions, defines information as:

- “data placed in a meaningful and useful context”(glossary), and
- “the degree to which information has content, form, and time characteristics that give it value to specific end users.”

He states, “One way to answer the important question [about what information is] is to examine the characteristics or attributes of information quality.” He suggests 15 attributes of quality of information within a three-dimensional framework: **(1) time**, **(2) content**, and **(3) form**. One can find them with short explanations in Figure 1.



**Figure 1: O’Brien’s Summary of Attributes of Information Quality**

Similarly Malaga (2005, p. 8), while referring readers to well-known management consultants Davenport and Prusak (1997) suggests six qualities of information in the following sequence: **accuracy, timeliness, accessibility, engagement, application (relevant), and rarity.**

Dock and Wetherbe in “Computer Information Systems in Business” (1988, p. 36) suggest analyzing the following attributes: **accessibility, timeliness, relevance, accuracy, verifiability, completeness, and clarity.**

Alter (2002, p. 162-168) distinguishes four main factors to information usefulness: **information quality, accessibility, presentation, and security.** He further subdivides them into characteristics that are more specific and discusses them with examples. Certainly, among the referenced sources, Alter offers probably the broadest presentation of the subject.

### Comment

First, one can see, except for Alter, authors provide nearly no hints how business analysts, end users and students should examine and apply those attributes or characteristics of data/information. Several problems can be identified immediately:

- The **utility value of data/information** is not defined, and no hints are given how to measure it, except for Alter (2002, p. 162).
- In which sequence should the listed dimensions or factors be considered? O'Brien's enumeration sequence suggests time, content, and form. Alter's sequence suggests: quality, accessibility, presentation, and security. However, is it so?
- Within each of the dimensions or factors, what is the recommended sequence of considerations? Is timeliness indeed the most important one, does accuracy take precedence over relevance (O'Brien), or accuracy over completeness (Alter)?
- The referenced authors did not make clear which of the listed attributes:
  - are of **mandatory** nature, and must be met **unconditionally**,
  - cannot be met **fully**,
  - are **nice to have**, but are of secondary nature where one can tolerate some imperfections without losing much of their business utility.

Alter's (2002, p. 163) approach to this subject differs substantially from the other authors. The most drastic one is that information accessibility, presentation (form), and security, however discussed, are not considered components of information quality, but determinants of information usefulness on equal footing with quality. He does not mention relevance at all, but adds to the attributes considered by other authors the following ones: **precision, age, source, availability, admissibility, access restrictions, and encryption.**

Some authors, such as Post and Anderson (2003), are so enamored with the technological aspects of MIS that they do not refer to quality of data and information at all, which, nevertheless, are the *main* substance of all high-tech processing.

For students, who take general education courses on computer or management information systems (CIS or MIS), this subject provides an excellent opportunity to learn how, through critical examination, analytical thinking and the reasoning power of the human mind, one can arrive at useful conclusions. The purpose of this critical comment is not to discourage students from reading and using the quoted textbooks; at CSU Stanislaus O'Brien's books are required reading for them. The real purpose is to emphasize the importance of critical reading even of the best and most popular texts.

### ***What Does Empirical Research Offer?***

Huang et al. (1999 p. 13), unless otherwise specified, use the term "information" interchangeably with "data." After they reviewed the three approaches used in literature and in business practice to study information quality (IQ) (**intuitive, system, and empirical**), they decided to use the following (Huang et al., 1999, pp. 33-34)

- a system definition anchored in an ontological, logical foundation, and
- an empirical definition derived from the information consumer's perspective.

#### **A. The system and ontological approach**

The system definition of information quality:

- concentrates on the internal view intrinsic to data and information,

- is oriented toward system design and data production,
- is use independent,
- enables comparisons across applications, and
- may be used to guide the design of information systems by information quality objectives.

The fundamental role of an information system is to provide a representation of an application domain (real-world system) as perceived by the user. Representation deficiencies are defined in terms of the differences between the view of the real-world system as inferred from the information system and the view that is obtained by directly observing the real-world system. From various types of representation deficiencies, a set of information quality dimensions (**complete, unambiguous, meaningful, and correct**) is derived.

As shown in Table 1, based on the analysis of the mapping from states of the real-world system to states of the information system Huang et al. (1999, pp.39-40):

- identified **four potential representation deficiencies** derived from **four intrinsic** (i.e., system oriented) **information quality dimensions** (**complete, unambiguous, meaningful, and correct**, and
- associated them with two **sources of deficiencies** (**design and operation failure**) and related those sources to some **observed information problems**.

**Table 1 Intrinsic Information Quality Dimensions and Observed Problems**

<b>Dimensions</b>	<b>Nature of Associated Deficiency</b>	<b>Source of Deficiency</b>	<b>Observed Information Problems</b>
Complete	Improper representation: missing information system states	Design failure	Loss of information about the application domain
Unambiguous	Improper representation: multiple real-world states mapped to the same information system state	Design failure	Insufficient information: the data can be interpreted in more than one way.
Meaningful	Meaningless information system (IS) state and garbling (Mapping to a meaningless IS state)	Design failure and Operation failure	It is not possible to interpret the data in a meaningful way.
Correct	Garbling (mapping to a wrong information system state)	Operation failure	The data derived from the IS do not conform to those used to create these data.

(Sources: Huang et al., 1999, p.41; Wand & Wang, 1996)

### **Comment**

First, one must be aware of fundamental terminological differences between O'Brien and Huang. O'Brien's attributes of information quality are Huang's dimensions (later grouped into information quality categories) and vice versa, but O'Brien's dimensions (time, content, form), are not

used by Huang. This and other disparities, listed later, constitute a formidable challenge for instructors to develop a consistent presentation of the subject, and for business students to comprehend it.

The first weakness pertains to **completeness** as defined above. Despite professing the ontological approach, Wand and Wang (1996) did not address the most acute problem of completeness in real life situations. Every business manager, field commander, and scientist is aware that completeness of information in the real world is simply unattainable, and it is not simply a design failure. It is the result of the limitations of human cognition in science, and the limitations of intelligence in business and military operations. A current anecdotal definition of a “good manager” reads: – *a manager capable of making good decisions with incomplete information*. In business organizations in cutthroat competition and in warfare, the critical blow most frequently comes from a danger, direction, or factors not recognized in time. Business students, the future end users of data/information, should be made aware of this reality.

Other dimensions such as **unambiguous, meaningful, correct** are defined precisely within the strict context for mapping. In business practice, however, **meaningful** and **correct** mapping of the real world to the information system state is only part of the problem. Even with perfectly meaningful and correct mapping, as defined by the authors, another type of mapping follows immediately – the mapping of the information system state to the decision maker’s mindset. At that time, completely different distortions frequently occur. Usually they are of a psychological, social, and political nature. Those distortions cannot be entirely avoided, but by careful design of proper organizational procedures, and proper checks and balances, they can be minimized to some degree. In short, they should not be ignored when teaching the subject to business students. They should be prepared to design and implement the adequate preventive measures or at least require other professionals to do it.

### **B. The empirical approach from information customer perspective**

The empirical definition of information quality is based on the information consumer’s perspective, and on the Total Quality Management (TQM) literature. In this view, information quality should not be defined by providers or custodians of information, but instead, by information consumers. Information quality is defined as information that is fit for use by information consumers. Information is treated as a product. While most information consumers do not purchase information, they choose to use or not use information (Huang, et al., 1999, pp. 42-43). Garvin (1987) goes even further and says “high quality means pleasing the consumers, not just protecting them from annoyances.”

Huang et al. (1999, p. 44), using qualitative analysis, examined 42 information quality projects from three leading-edge data-rich organizations, which are leaders with regard to attention to information quality. Each project served as a mini case, and it was analyzed with respect to the identified dimensions of quality.

These authors refer to a “case study” as an empirical inquiry that investigates a contemporary phenomenon within its real-life context. They emphasize that the study was done within a “larger information system’s context” to cover the organizational processes, procedures, and roles employed in collecting, processing, distributing, and using data. They developed a framework (see Table 2), which contains four **information quality categories (intrinsic, contextual, representational, accessibility)**, and two or more associated **information quality dimensions**.

**Table 2 Information Quality Categories and Dimensions**

<b>Information Quality Categories</b>	<b>Information Quality Dimensions</b>
Intrinsic IQ	Accuracy, objectivity, believability, reputation
Contextual IQ	Relevancy, value-added, timeliness, completeness, amount of information
Representational IQ	Interpretability, ease of understanding, concise representation, consistent representation
Accessibility IQ	Access, security

(Source: Wang & Strong, 1996)

**Intrinsic information quality** denotes that information has quality in its own right. Accuracy is merely one of the four dimensions underlying this category.

**Contextual information quality** highlights the requirement that information quality must be considered within the context of the task; it must be relevant, timely, complete, and appropriate in terms of amount to add value.

**Representational and accessibility information quality** emphasize the importance of the delivery system. It must be accessible but secure. It must present information in a way that is interpretable, easy to understand, concise, and consistently represented.

Huang et al. (1999, p. 56) claim that they:

- defined the concepts of information quality objectively and subjectively,
- provided the essential vocabulary for identifying IQ problems, and
- formed the foundations for measuring, analyzing, improving information quality in a continuous cycle.

### **Comment**

There is no doubt that their first-class empirical study constitutes enormous practical progress. It should find its proper place in MIS textbooks and replace the current rather eclectic enumeration of attributes of information quality with no guidelines on how to use them. From the business end user's viewpoint, the first inconsistency of the above referenced research is that **completeness** is listed twice:

- First, it is defined as a mapping or design deficiency and listed as an independent intrinsic dimension of information quality in Table 1.
- Second, it is not defined explicitly, but listed as a contextual dimension of information quality in Table 2.

The glossary of the text does not contain a definition of **completeness**. One can find a definition of *incompleteness*, but it pertains only to incompleteness of mapping, which is not of contextual nature. Later in the text, **contextual incompleteness** is explained as missing data, whatever that may mean, due to operational problems within the boundaries of the mini case. Even within the contextual category, the purely empirical approach neglects entirely the most difficult strategic

problem of information completeness in real business life - the deficiency of business intelligence.

Another weakness is that most attributes or dimensions of information quality are defined insufficiently or not at all. They are usually explained only by example within the limited context of particular mini-cases. This time, to further increase confusion, **information dimension** is defined as a **set** of information attributes that represent a single aspect or construct of information quality.

This is not a criticism of the conduct of this empirical study, but an indication of its inevitable limitations. The authors are aware of some of them, when they emphasize, “the disadvantage of empirical approach is that the correctness or completeness of the results cannot be proven based on fundamental principles” (Huang et al., 1999, p. 34).

After adding up all the suggestions listed in the referenced sources, we face about 25 attributes of information quality, or at least its usefulness. In the referenced research, attributes are referred to as dimensions. The main purpose of this paper is to show that by an analytical examination of the accumulated findings, whether empirical or rational, one can arrive at a fuller understanding of the subject, discover several mutual interdependencies among those aspects and develop an economic framework for their practical examination.

Part two of this paper constitutes the author’s proposal of “What to teach business students in MIS or CIS courses about data and information.” It is presented for discussion and critique with the purpose of arriving at a better consensus on what educators should offer business students on this subject.

## **What to Teach on Data and Information – Author’s Proposal**

One way the human mind deals with a complex reality is by building a symbolic model of that reality. Such a model should adequately represent or faithfully reflect that reality. In addition, with the information technology available, the *purpose of computer data, information and knowledge in business* is to represent in a symbolic manner the **business reality** management must deal with.

**Business reality** encompasses the **business system** and its **environment**. Within that reality, one can distinguish two major types of entities:

- **business objects** (customer, employee, inventory item, etc.), and
- **business events** (sales order placed by a customer, promotion or demotion of an employee, shipping of an inventory item, etc.).

Each entity under consideration must be symbolically represented with regard to its:

- **identity** (customer ID, inventory item ID, transaction ID, etc.), and the
- states of their relevant **attributes** (age, address, telephone number, date, etc).

Computers can only recognize, store, process, and present certain types of physical states, which are usually of the binary, that is two-state, nature (surge in voltage, direction of magnetic field, etc.) that the computer engineers designed them for – the so-called **computer bits**. Those bits and combinations thereof, by *computer engineers’ design*, can represent other symbols such as letters, digits, special characters, pixels, etc. that are more friendly to human senses and the human mind than computer bits. Next, only by *computer end-users’ design and determination*,



those symbols and combinations thereof, called *data values*, may represent some aspects of business reality.

## **Computer Data<sup>1</sup>**

In computers, **data values** have to be stored on portions of computer storage media called **data fields**. Hence, within the context of computerized processing, one has to view *data fields* as identifiable **storage containers** capable of accommodating a class of data values that symbolically should represent some aspects of business reality. With regard to data fields, we have to distinguish data field **types** and data field **occurrences**.

A *data field type* is identified for:

- human use by an end-user-assigned **data type name** that usually one can find in report or table column headers, and for
- computer use by an entry in **database dictionary** usually followed by a complete or default definition of the **data field type**.

A *data field occurrence* is any one of many data fields of the same type designated to accommodate data values of the same type, such as customer telephone number, or customer name. They are identified for:

- *human use* in the source code of a computer program usually by a **subscripted name of the data type** or in spreadsheets by a **column letter and row number identifier of the cell**, and for
- *computer use* - by an **absolute** or a **relative storage address**.

Thus, **computer stored and processed data** are only collections of symbolic representations of something ultimately and exclusively *defined and determined by computer end-users*. This definition makes it completely clear that the determination of computer data values is the *exclusive responsibility of end-users*.

In this task, end-users can be assisted by professionals such as systems analysts, programmer analysts, database administrators etc., but no more than tax preparers can assist taxpayers. Preparers cannot be liable for the veracity of the content of tax returns they helped to prepare. They are liable only for the correctness of the procedures they used to determine the tax liability according to the tax law then in effect.

All this has to be clearly said and deeply ingrained in the end-users' minds. In-class discussions invariably demonstrate that most business students are not sufficiently aware of *business end users' responsibility for data*. Too frequently, computers are blamed for deficiencies that mostly are of human, and more specifically, of end-user origin. Due to ever-increasing reliability of computer hardware, computers rarely fail, and when they do, they are designed to notify us. *Data values stored in computers are factual representations of business reality only when end-users*

---

<sup>1</sup> In traditional business communications, **data** (derived from datum that in Latin means "given") are given or admitted facts or premises. This definition, however, is inadequate, even largely misleading with regard to data values stored in computers or on other media. It deceptively suggests that the data values stored and processed represent given or admitted facts and premises. This assumption may be true only under rigorously defined *organizational conditions*. For veracity of computer-stored data values cannot be verified or validated by *computational procedures*; it can be assured only by **computer end-users**.

*make them such.*

## **Information**

In most cases, in everyday business communications, the terms **data** and **information** are used interchangeably. Indeed, frequently there is no difference, with no consequences whatsoever. Students of business information systems, however, should be aware of and should be sensitive to situations where the difference is *perceivable* and the *business consequences are substantial*. To clearly delineate that difference in the business context one may use *decision situations* as an adequate frame of reference, for that is what managers face all the time.

Even a partial and imperfect description of a decision situation constitutes what decision-makers already know about it - it constitutes what is *given* to them. A collection of **given** or **admitted facts** and **premises** are *data* in the conventional meaning of the term. Recipients of data values and decision-makers consider themselves as informed only when those values convey something unknown to them. Imagine an extreme case of a system that delivers only data values that represent what decision-makers already know! It is a pure waste of resources, a useless service with no other business consequences than waste. No change or progress occurs. In such situations, decision-makers do not perceive themselves as being informed. In real life, such a situation never exists in a pure form, but there is plenty of evidence that most printed reports end up in wastebaskets without being read at all.

Of course, the data already available also require storing, processing, retrieving, and presenting in various forms with the purpose of increasing their usability, but usability of data is not a synonym for information. All this constitutes the old familiar **data processing**, not information processing in the strict meaning of the term. Yes, one may derive from the available data some additional data values not present in your database or memory; however, this requires that decision makers possess additional knowledge about the possible relationships among them, about the patterns of the represented fragment of business reality. The additional data values are derived due to additional knowledge of rules applicable to the case. Once they are known, and a programmer programs a computer to compute them accordingly, one will always obtain the same results as expected. It is not going to change the decision situation.

There is a wide spread infamous myth that information is **data processed** to a useful and meaningful form. It haunts most textbooks on Management Information Systems; is misleading, oversimplified, and should be dispelled. No amount of processing of available data, even if performed with supercomputers, will result in something that would change the decision situation under consideration and possible outcomes of decisions made based on those data. Yes, it may present the same data in a more agreeable, user-friendly, easy to use format that save end users' time and effort to interpret; it may increase their *usability*, but not their *informational content*. Usability is *not* synonymous to informativeness. Processing of known data cannot generate information that may change decision situations. It only facilitates usage of data and their cost effectiveness, as it will be explained later. Let us not leave students with that false and superficial impression.

The exception is data mining, which represents research with the purpose of finding new, not yet known patterns and new relations among the available data. This, however, is **knowledge acquisition**, not conventional data or information processing.

Decision-makers do experience a real difference only when the incoming data values are *new* to them, *change* the value of available data, and/or the perceived likelihood or probability of something pertaining to the decision situation under consideration; in other words, when the received data values *change* the decision-maker's mind about the situation. Only then does actual **inform-**

**ing** take place. Hence, from the *end-users' or managers' viewpoint*, data values are **informative**, or in a more rigorous manner, one can say they carry, convey, or contain **information** only under the following circumstances, when they represent:

1. something *new* or *unknown* and significantly relevant to the situation, and/or
2. a significant *change* of the known relevant data, and/or
3. a significant *change* of the level of the perceived uncertainty about or the probability of any aspect of business reality.

All of those aspects have to be carefully considered. One can see that **data** (the given, the known) describe the *current* state of a decision situation, and only those *incoming* data values that significantly *change* what is known are the real **information** from the viewpoint of **information theory** and decision makers. Thus, only *some data values* acquire the property of being considered **information**. It is determined by the situation, similar to the temperature of an object in the physical world or economic value of something to an individual or society. Even something given, known, hence data to one individual, when communicated to others, may be information to them, for they do not yet know it. There is an analogy with the news. What is news to one person may be history to another.

All the above examples referred to human recipients. Thus, one can easily get a false impression that information is a phenomenon in human communications only. The opposite is true. Information is an *objective physical phenomenon, which makes a difference independently of the human mind*. In business, we measure the impact of something by the changes in results. Therefore, information that communicates a significant change manifests its impact exactly the same way whether its recipient is a robot, any numerically controlled device, any computer-run decision model, or any rational human. As long the incoming **signals** (Defined as any perceivable, observable change of a physical state used in communication devices and computers to represent data values.) represent the current states of a device, they do not change or affect anything. Changes in results or outcomes may occur only when the incoming signals change the state of the device or the parameters of the model that describe the decision situation or the actions of a human. Only then they are relevant to the model and are considered informative to any recipient, whether a human, an animal, or a robot.

## **Knowledge**

Knowledge is more than data and information. It entails also rules for using them, rules of reasoning. Even when one possess relevant data and can acquire any necessary additional information, if one does not know how to use them, how to think about them, one is not yet knowledgeable. Within the context of previous definitions, **knowledge K** can be defined as the logical sum of **data D** known by decision makers, the **additional information I** they may acquire, and the **rules of reasoning R** known to them

$$\mathbf{K} = \mathbf{D} + \mathbf{I} + \mathbf{R}.$$

Those three components of knowledge may become the content of a **knowledge base** of an **expert system** that can assist **expert decision makers** in problem solving within a specified domain of expertise.

The following simple example of data, information, and a single rule of reasoning may elucidate the concept of **knowledge**. An out-of California Company knows its range of products or services, the structure of its distribution system, etc. This represents the available **data D**. The company knows also that according to the current law any trade by mail over state borders is not sub-

ject to state taxes. The State of California, in its never-ending search for more tax money, has just imposed the California State sales tax on everything sold by mail or Internet to customers in California. This constitutes the incoming **information I**, which changes the decision situation, and in consequence, the operating results of the company. This changes the rules of computing and reporting business activities over Internet in the State of California. The latter constitutes the added **rule of reasoning R**, which now has to be applied to all future financial projections prepared for the company.

### ***Quantity and Usefulness of Data/Information***

How much information any data value contains or conveys is measured by its surprise effect and by its probability or frequency, similar to news. The less probable and the more unusual an event is, the more newsworthy it is. On the other hand, when data values are transmitted over a distance, then from the viewpoint of the receiving station all data values transmitted are new, something unknown. This is the subject of mathematical **theory of communication**<sup>2</sup> developed by Shannon (1949). That theory deals with the **quantity of information** and problems arising in transmission of messages, which usually consist of one or more data values. Twenty years later, Mazur (1970) developed a more generalized **qualitative communication theory**, which eliminates the assumptions of probabilities and leads to the same conclusions. (In science, when one must resort to probabilities, it indicates that one does not yet possess full knowledge of the internal mechanics of the phenomenon.)

All the considerations on usefulness that follow pertain equally to data and/or information. For the usefulness of an incoming piece of information manifests exactly the same way as for a piece of equally useful data that for any reason has been lost and cannot be used anymore for one's benefit. In business, only *useful data/information* is worth considering. **Usefulness**, however, is contextual, depending heavily on the situation. How may usefulness be perceived in different situations?

- For *general education* purposes any message, (which usually consists of one or more data values) that broadens students' perception of the world, society, and community is useful.
- For *designers of decision support systems* only **data/information** that change the outcome of a decision situation under consideration are useful.
- For *business people* only **data/information** that change the results of their actions or their business operations for the better are useful.

Hence, in a business environment the **usefulness of data/information** is of foremost interest to end users and managers. It should be viewed the same way by *students of business information systems*, by *business systems analysts*, and while discussing **information quality**, as well.

There is, however, another trap. Quizzes show that many, if not most, students of MIS enter the course with an impression that the more information the better. The intrinsic relationship between the quantity of information and its usefulness is much subtler. In general, there is nearly none. Without going into mathematical aspects of how the quantity and the usefulness of data information are defined, at the undergraduate level it should suffice to explain that both formulas

---

<sup>2</sup> The theory of the probability of transmission of messages with specified accuracy when the bits of information constituting the messages are subject, with certain probabilities, to transmission failure, distortion, and accidental additions.

are completely different and show no common elements. In simpler terms, on one hand, one bit of data or information may be of extremely high usefulness, at least to some end users, and another megabyte of data or information may be of limited usefulness. On the other hand, there has to be at least some amount of data or information that can carry or can be associated with some kind of usefulness. ***Data/Information Requirements and End-Users' Responsibility***

Practically all types of business information systems whether for **operations support** or **management support** require:

- a constant access to organizations' data usually organized in **databases/data warehouses**, and a continuous
- handling of incoming information relevant to the **business reality**.

The first point is a subject of **Data Administration, Database and Data-Warehouse Management** handled by computer information specialists. Of course, all of these should be designed and operated with input from all types of end-users. Usually it occurs on a one-time basis, when principles, policies, and standards are being established. This subject is also referred to as **Data Resource Management**.

The second point is a continuous effort of acquiring, and collecting information about all the perceived changes taking place within and without the organization (in its environment) – business events. Here, an *active ongoing* participation of all types of end users is indispensable, for this is a domain, where computer information specialists play only a secondary role. They can, however, aid, support, and facilitate the efforts of end- users by suggesting proper use of cost-effective input devices, information gathering techniques, data entry verification, and data validation techniques, etc.

Data and information collection and acquisition is a *fulltime* and *exclusive* responsibility of end users. It can be outsourced only with utmost precaution. It means lowering one's guard and making organizations vulnerable to practically invisible influences by outsiders.

Hence, data and information requirements of business organizations have to be developed primarily by end-users. Again, computer information specialists can be of assistance. Students of business information systems should be aware that practically all known textbooks on this subject do not clearly delineate the boundaries, the necessary division of labor and responsibilities between computer end users and computer information system professionals. Whenever end users, mostly due to complacency or convenience, do not strictly discharge their responsibility the quality of information services will suffer. In such a case, new business information systems may be developed without paying sufficient attention to their main business purpose. They will be developed at the convenience of computer information specialists, who naturally pay primary attention to technical aspects and technical expediency. Methods and techniques of identifying and defining data and information requirements are not discussed here.

### ***Attributes of Data/Information Quality***

Usually, MIS textbooks cover the subject of data/information quality superficially, if at all, despite the importance of this subject to end-users. Even worse, they cover rarely the fundamentals of how to define usefulness of data and/or information in the business environment and how to articulate and examine all the attributes of data/information quality that determine it or contribute to it, and finally how to measure the outcome. In any business environment, one requires more than “fit to use” (Huang, et al., 1999, p. 42) or “pleasing the customer” (Garvin, 1987). It also

should be cost effective to use. No known textbook explains that in a business environment the *single most important and cumulative measure of quality* is the **expected cost effectiveness of data/information**.

**Utility value of data/information** (mentioned by O'Brien, but not defined in any referenced source, except for Alter (2002, p. 162) who considers it more elegant than practical)

We assume that decision makers, while making decisions and acting accordingly, use some data **D** known to them. Any incoming piece of information **I** significantly relevant to the decision situation changes what they know from the previous state **D** to state **D + I**. In reverse, *the same pertains to any lost data item*. In such a situation the *utility value of data/information V(I)* (Kofler, 1968) may be defined as the difference between the **value of results V<sub>R</sub>** of actions or business operations while acting with **V<sub>R</sub>(D + I)** and without it **V<sub>R</sub>(D)**, that is

$$V(I) = V_R(D + I) - V_R(D) \quad (1)$$

Utility value of data/information is an important attribute for any end-user of business information systems. It may be calculated only under the assumption that *the results of business operations can be measured and evaluated*, but not necessarily in monetary units. The same formula covers the utility value of an incoming significantly relevant information item as of a lost piece of previously available significantly relevant data.

**Procurement cost of data/information** (not in referenced sources, except for Alter (2002))

Nothing, however, comes from nothing. The decision-maker must consider the *procurement cost of data/information C(I)*, such as the cost of counting, measuring, researching, retrieving, or acquiring it.

**Net business utility value of data/information** (not in referenced sources, except for Alter)

If the utility value of data/information **V(I)** and their procurement cost **C(I)** can be expressed in the same monetary units, then one can calculate the *net utility value of data/information V<sub>N</sub>(I)* by subtracting its procurement cost **C(I)** from its utility value **V(I)**, that is

$$V_N(I) = V(I) - C(I) \quad (2)$$

The formula calculates the *absolute* net benefit resulting from its use.

**Cost effectiveness of data/information** (not in referenced sources, except for Alter (2002))

To obtain a *relative* measure of **benefits attributed to data/information I** one should compute the *cost effectiveness of the data/information C<sub>E</sub>(I)* by dividing its **utility value V(I)** by its **procurement cost C(I)**, that is

$$C_E(I) = V(I) / C(I) \quad (3)$$

Any ratio less than 1 means a losing proposition. When the ratio equals one, it is a breakeven proposition, which in a profit-oriented business is considered a loser, too. (When it does not generate any surplus, and when in addition one takes into account the risk associated with any new venture, with certainty it also becomes a loser). Even any value of the ratio between 1 and 1 + the current discount rate (usually the percentage rate that banks charge for lending you money) is unappealing from the business viewpoint. One can safely obtain a similar result while acting without the particular data/information and leaving with a bank the money necessary to procure it.

When the value of the additional results attributed to **data/information I** *cannot* be expressed in monetary units as the corresponding costs, we can still calculate the *cost effectiveness of data or*

*information*  $C_E(I)$  by dividing the **quantity of attributed results**  $Q(I)$ , instead of monetary value  $V(I)$ , by its **procurement cost**  $C(I)$ ,

$$C_E(I) = Q(I) / C(I) \quad (4)$$

Thus, a comparison of cost effectiveness of different choices of information support is still possible. The higher the ratio of the additional results to the procurement cost of the data and/or information itself or the information service offered the better the option under consideration. The applicability of this ratio is limited for usually one cannot compare it with the cost effectiveness of other projects and other business activities.

**Expected cost effectiveness of data/information** (not mentioned in referenced sources, with some exception for Alter (2002, p. 456) who implicitly refers to and discusses various risks)

However, any new business venture, whether information related or not, is risk-prone. Thus any proposal submitted to end users and their management should include the calculation of the **expected cost effectiveness of information**  $EC_E(I)$  by dividing the **expected utility value of information**  $EV(I) = V(I) * (1 - \text{risk factor or failure rate})$  by the **expected procurement cost of information**  $EC(I) = C(I) * (1 + \text{relative average cost overruns})$ , that is

$$EC_E(I) = EV(I) / EC(I) \quad (5)$$

This ratio has to be greater than the same ratio for any project competing for the same business resources, greater than the current ratio for the entire business unit, and greater than  $1 +$  the current discount rate. Only then does an offered proposal make business sense.

In conclusion, MIS textbooks should explain that the **expected cost effectiveness of data or information** is the *most important cumulative* indicator of the **quality of data/information** or any **information service offered**. None of the referenced sources, intended mainly for business students, considers this subject.

### ***The Determining and Contributing Factors to Cost Effectiveness of Data/Information***

Most textbooks and the empirical research list under different names several attributes of information that should be considered. The major question is, however, how to go about this task, and how to consider them in real life situations.

1. Which of them are mandatory (must be met in any case)? If not, eliminate that piece of information from further consideration.
2. Which of the mandatory ones have to be considered first? If not met, this also renders unnecessary any further analysis.
3. Which of them are not fully attainable? Hence, the decision makers have to trade off.
4. Which of them are optional or nice to have? Hence, one can tolerate some imperfections without losing much of their cost effectiveness.

Taking into account the sources referenced in Part 1 of this paper, one faces about 25 attributes of data/information quality or their equivalents: dimensions, factors or characteristics. The purpose of this part of the paper is to put the attributes in a logical order. First, they are divided into two major categories: the primary or mandatory ones and the remaining ones, which are of rather secondary or auxiliary nature. Second, they are presented in a sequence that economizes their examination. That sequence is based on a logical interdependence among them. Thus, when after examination of any mandatory one, one arrives at the conclusion that the corresponding require-

ment cannot be met; the examination of all the remaining attributes should be discontinued saving time and effort. It turns out that there exists a logical interdependence among at least the primary attributes of data/information quality.

Alter (2002, p. 163) lists availability as a characteristic of accessibility. Therefore, it is necessary to make an introductory remark that availability, not in Alter's sense "*whether it can be reached with minimum effort*", but whether an information item can be available at all, is here an implied assumption. Only an available substance can be subject of any practical examination.

### (1) **Interpretable, recognizable, legible** (Representational category – by Wang)

For any message or statement, which usually consists of one or more data values, to be useful at all, it has to be **interpretable, recognizable or legible**. In practice, it means whether the received data value fits a state with some attributed or associated meaning in the human mind of the receiving individual, or a state that triggers automatically a designed sequence of state transitions in the receiving numerically controlled device. This term does not cover the issue of how easy or difficult the interpretation is. This attribute is frequently omitted probably as obvious and not worth mentioning. Interpretability is the very first primary and mandatory requirement that must be met *unconditionally*. When for any reasons the decision maker or the receiving device is unable to interpret the data/information item, it is lost and it should be excluded from any further consideration.

Interpretability is contextual, e.g., a more educated recipient or a different receiving device may be able to interpret it. Within the representational category, Wang lists also ease of understanding, conciseness, and consistency. O'Brien within the form dimension lists clarity, detail, order, type of presentation, and media used. Alter lists format and level of summarization. All of them are nice to have, and they may actually add value. They are, however, of secondary nature and in this paper, they will be listed under the collective name **ease and effectiveness of use**, except for conciseness that is an aspect subordinate to relevance.

### (2) **Significantly relevant** (Content dimension – by O'Brien, contextual dimension - by Wang, engagement – by Davenport and Prusak, skipped - by Alter)

It is of primary importance that the content of messages must be *significantly* (above an assumed or acceptable threshold level) **relevant** to the decision situation under consideration. In practice, it means that it somehow affects the decision situation, and subsequently significantly changes the results of the decisions made and the corresponding actions taken. If not, all its remaining attributes are irrelevant, too. Hence, relevance is the *second primary mandatory* attribute that must be met unconditionally, with no exceptions. However, relevance can be quantified or at least ranked. One may ask how sensitive the model of a decision situation under consideration is to the usage of any specific data value, whether it has any significant impact on the results worth consideration. Hence, one can say in a more rigorous manner that a data/ information item in a specific situation may be qualitatively relevant but quantitatively irrelevant, when its impact is considered negligible. If so, one should also stop examining it.

O'Brien lists separately **currency**, or **pertinence to the proper time**. After a short reflection, however, one can see they are only aspects of relevance. When information is not current from the view of the decision situation, then it is irrelevant. With regard to **conciseness**, one may say: the less concise a message, the more redundant and irrelevant components it contains.

Separately, as not belonging to the factor of information quality as defined by Alter (2002, p. 163), he lists additional characteristics of usefulness of information, such as **admissibility, access restrictions and encryption**. Rightly, admissibility of any piece of information should be taken into account. Again, after a sort reflection one can clearly see that it constitutes only an-



other aspect of relevance. For instance, if age by law cannot be used as a valid factor in hiring decisions, or a judge ruled it as inadmissible, subsequently such an information item becomes irrelevant to the respective decision situation. Similarly, if lack of security, whether by encryption or password protection may render an affected information item useless, it also becomes irrelevant. Hence, the security issue may become a precondition of relevance. In some situations, relevance of information must be assured by restricting access to it only to authorized users. Even worse, unrestricted access to some information may make the business organization vulnerable to various adverse effects, and then its relevance becomes even more prominent. The conclusion is that Alter's **admissibility** and **security** of information, which are very valid concerns, are simply aspects *subordinate* to **relevance** of information.

**(3) Credible (true, accurate – meant as free from error)** (Intrinsic category – by Wang, accuracy – by Alter and O'Brien, source - by Alter)

The adjective **true** means consistent with reality. While probing a message or a single data value whether it is true, one should examine it in several ways:

- How **credible, believable or reputable** the source is,
- How **objective or unbiased** the collection and acquisition process was, and finally,
- How **faithful** the mapping from real world states to respective data values was: **complete** (Missing data value for existing real world states.), **unambiguous** (Multiple states of the real world not mapped to the same state of data value.), **meaningful** (Real world state mapped to a meaningful data value.), and **correct** (Real world state not mapped to a wrong data value.) as accurately described by Wand and Wang (1996).

**Credibility** of information is the third *primary and mandatory attribute*, after relevance.

The first problem with credibility is that it carries even a greater weight than it initially appears. Lack of credibility may also have a very destructive and negative impact. Misinformation, whether spread irresponsibly or planted on purpose, (for instance, for personal gain or worse to cause *harm and damage*.) is not simply negligible, Credibility is *rarely fully attainable*, particularly, when obtained from external sources over which one has little control. Even internal sources cannot be trusted either, unless carefully designed and implemented organizational procedures protect decision-makers from receiving **misinformation**. Verification of sources and using information from *preferred* sources is insufficient, contrary to Alter's suggestion (2002, p. 163).

The second problem with credibility is lack of **objectivity** or **bias**, which is a systematic inaccuracy due to approaches and methods used in selecting of primary sources, measuring points, observation points, and finally collecting, processing and presenting data. The resulting distortion may be either unintended due to ignorance or introduced intentionally. In both cases, the results of such distortions may be significant, and in the latter case, deceptive and damaging.

The third problem with credibility is how **accurate** the representation of reality is. One encounters **random errors** in all results of measuring and observation. They also may be due to deficiencies of mapping. The typical, very gross measure of **inaccuracy** in this sense is the error rate. One calculates it by dividing the number of errors by the number of items. A much more useful measure is the expected cost of dealing with their consequences. One may calculate it by multiplying the number of items by the probability of each type of error by the average cost of dealing with those errors. This measure of **inaccuracy** provides the end users with a good idea how important each type of error is. One may reduce many of those errors by using check digits, error self-detection codes, error self-correcting codes, etc. A good example is how barcode readers

may considerably reduce most mapping errors, except for completeness of mapping. End users of information systems, even business systems analysts, need not to have the expertise how to deal with such a situation, but they should be taught to recognize the need for such measures.

Another problem with accuracy is the use of adequate precision to represent data. For numerical data, it is measured by the number of significant digits. The precision of pictures and images one measures by the number of dots per inch. This measure is commonly used to describe the precision of printers, computer screens, scanners, etc. Hence, precision is a subordinate, not an independent, aspect of accuracy. (Alter, 2002, p. 163 - see his list of characteristics of information quality). It is an aspect of accuracy as accuracy is an aspect of information credibility. Insufficient precision of data/information presentation may compromise its accuracy to such a degree that it becomes useless. There is a trap associated with accuracy of information in both of its meanings: as free from error, and as precise in its presentation. Generally, it is overrated (Wang, 1996). Students of business administration and computer information systems are rarely warned that *unchecked* efforts to increase the level of accuracy and precision of anything can become counterproductive. The ultimate determination of the indispensable and economically justified level of accuracy and precision of data/information strongly depends on its utility value. Hence, for a decision situation under consideration, it should be postponed until the examination of timeliness and completeness of data or information has been concluded. Only then, the relative importance of each relevant data value can be properly ranked and later its economic level of accuracy and precision determined. (See item (6), Economic level of accuracy and precision).

Continuous monitoring of procedures pertaining to information acquisition, verification, validation, auditing, double-checking, etc., must be implemented. If credibility of information is in doubt, it makes all other considerations immaterial. **Credibility** of information is *a real and serious problem*; it is not a given, *it must be assured*.

**(4) Timely available** (Contextual category– by Wang, time dimension - by O'Brien, related to age of information – by Alter)

In ever-changing business reality time is of the essence. **Timeliness** is here defined as delivered sufficiently in advance to enable taking effective action. It is the fourth *primary and mandatory* attribute. Even with all remaining attributes as perfect as possible, when timeliness cannot be assured, the impact of information delivered late, may be null. If data/information are not available when needed, it does not make sense to ask about any other remaining attributes of quality. According to Huang's et al. (1999) research, information consumers perceive lack of timeliness as an **accessibility** problem.

There is, however, another trap. Meeting mandatory requirements usually does not add value; it makes the data value only acceptable, but timeliness is scalable. One may receive the necessary data/information on time, but more or less in advance. Any additional time available may be used for making decisions with a lesser hast, and for taking actions that are better prepared. It may yield better results and add value.

Alter (2002, p. 163) mentions age of information, which on its own merit, without knowing the volatility of the represented state is of no value. It should be rather considered as another aspect of its accuracy, hence, credibility.

**(5) Complete** (Contextual category– by Wang, content dimension - by O'Brien, characteristic of quality – by Alter)

Within the context of decision situations **completeness** measures the degree to which the relevant data values are available. In its quantitative dimension, it can be measured in percentage points (1 - 100%). It is the ratio of the sum of all results that can be attained and attributed to the

respective pieces of relevant data/information gathered and the sum of all results that could be attained, if all identified relevant pieces of data/information were available for decision making.

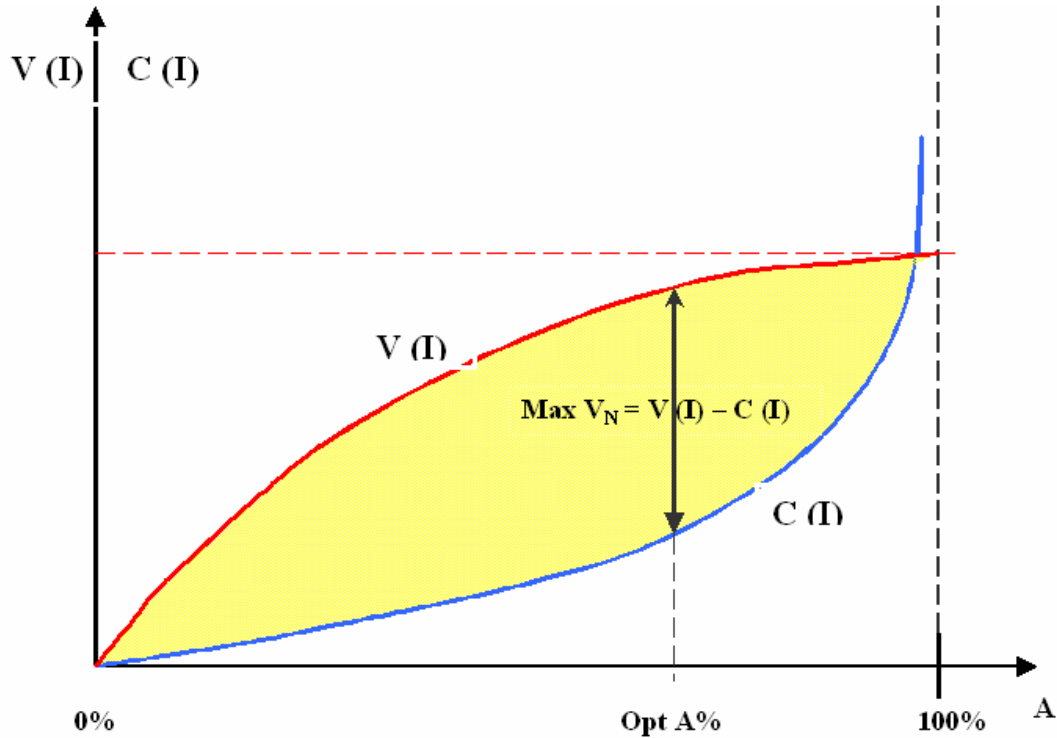
Murkier, however, is the qualitative aspect of completeness of data/information. In cognition and research there is a general rule that considerations of a qualitative nature always precede quantitative considerations. Wild animals and birds are on continuous watch, looking out for anything unusual in their environment. Only after spotting it, they do start focusing their attention on it for a more accurate assessment of its nature, scale and scope. O'Brien's attribute of **scope**, as defined by him, is only an aspect of completeness, thus of secondary nature. Alter's (2002, p. 165) suggestion on using end user's feelings as the criterion, when it is unnecessary to look for additional information makes it completely subjective to end users' education, training, and experience. Even a more objective suggestion that one should cease looking for additional relevant information when the consecutive attempts render marginally negligible improvements in results is very risky. The critical blow most frequently comes from a danger or direction not identified and recognized in time. Therefore, in its qualitative dimension, completeness is rarely to never fully attainable. In real life situations, in the fight for survival, on a battle field or in global business competition, one can never be certain whether all relevant success factors have been identified and evaluated.

#### (6) Economically accurate and precise

**Figure 2** approximates graphically how the net business utility value of data/information changes as a function of its accuracy and precision. There one can see two graphs plotted as functions of the level of **accuracy and precision A** in percentage points (1 - 100%), for its level affects both the numerator and the denominator of the cost effectiveness ratio.

The first graph represents the **business utility value V (I)** of **data/information I** as a function of **accuracy A** expressed in percentage points, that is  $V(I) = f(A)$ . There is an assumption that the utility value of data/information of unknown accuracy is equal to zero. First, the graph line of utility value rises relatively fast then it slows down with increasing accuracy until it reaches its full value according to the definition  $V(I) = V_R(D+I) - V_R(I)$ . Close to the end, any increase in accuracy yields a lesser and lesser marginal increase in the data/information's business utility value. The graph is similar to the graph of a logarithmic function.

On the other hand, the second graph represents the **procurement cost C (I)** of **data/information I** as a function of **accuracy A**, that is  $C(I) = f(A)$ . Again, usually one may assume that the cost of information of zero accuracy is equal to zero; one can get it free as a gossip or rumor, for instance. At the beginning, the graph line rises slowly with increasing accuracy, then the rise accelerates, and before the end, the rises becomes steeper and steeper to reach infinity, whenever one attempts to attain 100% accuracy. In mathematics, this kind of rise is referred to as **asymptotical**. Hence, the first conclusion is that, when one pushes too hard for increased accuracy, the **procurement cost C (I)** becomes *prohibitively high*. Attaining higher levels of accuracy requires end-users to incur ever-higher costs of research, measurement, additional observations, expensive instruments, etc.



**Figure 2: Optimum level of Accuracy A as function of Net Utility Value  $V_N$  of Information I**

Before reaching 100% accuracy, the steep rise in cost and marginally slower rise of utility value causes both graph lines to intersect. In contrast to the prevailing initial perception of business students, in business, one never gets rich or enriches others by incurring costs equal to the value of results. The optimum level of accuracy in the business environment lies where both graphs are the furthest apart; this is the point where the *net utility value of data/information*  $V_N(I)$  reaches its maximum. One can count on *maximum business benefits* from using data/information only at its *optimum level of accuracy*.

Finding this optimum is not easy, but the truth is that it lies somewhere between a low and high level of accuracy. Whenever information technology professionals tempt end users with higher accuracy than they had before, they should ask bluntly: What will be the additional business benefits and at what additional cost? When one has no indication that increased accuracy leads to higher cost effectiveness, forget it.

It is a kind of irony that in section after section of MIS courses, most students enter the course with a deep conviction that accuracy is the most important of all attributes of information quality. Only after one or two sessions of brainstorming and analytical comparisons, they arrive with reluctance and in disbelief at a very different conclusion. On one hand, the accuracy of data/information, meant as free from errors, should be considered as an aspect of its credibility during the third step. On the other hand, the examination of the accuracy of data/information, meant as adequately precise in its representation, should be postponed until the very end of the examination of all the primary attributes of data/information quality.

**(7) Ease and effectiveness of use** (a collective name for all secondary or auxiliary attributes)

Ease and effectiveness of use of data/information is a collective name for most of the quality attributes of secondary nature. Usually they are strongly related to format. For instance, depending on its format, the end-user may read, interpret, comprehend, analyze, draw conclusions, and act upon it either slower or faster. Under this category, one may list **clarity, consistency, order, media used, level of summarization, user-preferred type of presentation such as text, graph, diagram, picture**, etc. Deficiencies with regard to those aspects would rarely preclude the use of the affected data/information. Those aspects, however, may increase or decrease its value or its procurement cost, for both subsequently affect the expected cost effectiveness of data/information. Within this dimension, also O'Brien's **frequency** of informing may be considered. For instance, whether less frequent reporting does not impair taking effective actions or counter measures.

## Conclusions

Table 3 summarizes the most important conclusions pertaining to the discussed attributes of data and information quality viewed from the perspective of business decision situations under consideration.

1. It is possible to establish a logical and subsequently an *economic sequence of examining at least the primary attributes of data/information quality*: **interpretability, relevance, credibility, timeliness, completeness**, and **economic level of accuracy and precision**.
2. All remaining attributes may be examined as ones of secondary or auxiliary nature listed under the collective name **ease and effectiveness of use**.
3. The attributes of **interpretability, relevance, and timeliness** *are mandatory and must be met unconditionally*.
4. **Interpretability, credibility**, the economic level of **accuracy** and **precision** of data and information do not add value but affect its procurement cost.
5. **Credibility and completeness** *are never fully attainable*; hence, one should try to increase them until their marginal cost effectiveness does not fall below the acceptable level. Even so, big surprises must too be expected and contingency provisions established.
6. One should try to optimize the economic level of **accuracy** and **precision**.
7. All aspects of **ease and effectiveness** of data/information **use** are of auxiliary or secondary nature, however, they may change its procurement cost, hence, its cost effectiveness.
8. On its own merit, the amount of information does not add value. The **amount of data/information** should be considered only from the viewpoint of whether it is significantly relevant, can be delivered on time, and adds value to an information system project in a specific business situation.

**Table 3 Priorities, attainability, and the potential of adding value by the primary and secondary attributes, dimensions or characteristics of information quality**

Primary and mandatory attributes, when attainable, must be met unconditionally	Logical examination sequence of attributes of data/information quality ↓	<b>Name, description and comments to attributes of data/information quality</b>	
		<b>1</b>	<b>Interpretable, recognizable, legible</b> (whether possible)
		<b>2</b>	<b>Significantly relevant</b> (Pertinence, currency, conciseness, engagement are of mandatory nature, but may also add value depending of the sensitivity of the decision model to the specific information item)
		<b>3</b>	<b>Credible (true, accurate – free from error)</b> (Reputable source, objective, not misinformation, never fully attainable)
		<b>4</b>	<b>Timely accessible or available</b> (A mandatory requirement, may be met with shorter or longer time available for making decision and taking action, hence it may increase or decrease the cost/effectiveness of the service)
		<b>5</b>	<b>Complete</b> (The degree to which relevant data are available and their marginal cost/effectiveness is acceptable, hence, it adds value; never fully attainable, hence surprises should be expected)
		<b>6</b>	<b>Economically accurate and precise</b> (One should try to identify its approximate optimum for the highest level of cost/effectiveness of data/information)
Change procurement cost, hence cost effectiveness	<b>7</b>	<b>Ease and effectiveness of use (a collective name for the <u>secondary</u> or <u>auxiliary</u> attributes of information quality)</b> (Usually strongly related to format such as: clarity, consistency, order, media used, frequency, whether detail or summary, application of user-preferred type of presentation such as text, graph, diagram, picture, etc.; may change cost effectiveness of data and information)	

## References

Alter, S. (2002). *Information systems – Foundation of e-business* (4<sup>th</sup> ed.). Prentice Hall.

Davenport, T. & Prusak, L. (1997). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Dock, V. T., & Wetherbe, J. C. (1988). *Computer information systems for business*. West Publishing.

- Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65 (6), 101-109.
- Huang, K., Lee, Y. W., Wang, R. Y. (1999). *Quality information and knowledge*. NJ Prentice Hall PTR. (Quality Programs & Initiatives at MIT - [MITIQ Program](#)).
- InSITE 2004. (n.d.). Retrieved from <http://2004.informingscience.org/>
- Kofler, E. (1968). *O wartosci informacji* (On value of information). Warsaw, Poland: Panstwowe Wydawnictwa Naukowe (PWN).
- Malaga, R. A. (2005). *Information systems technology*. Pearson Prentice Hall.
- Mazur, M. (1970). *Jakosciowa teoria informacji* (Qualitative theory of Information). Warsaw, Poland: Panstwowe Wydawnictwa Techniczne (PWT).
- O'Brien, J. A. (2003). *Introduction to information systems* (11<sup>th</sup> ed.). Mc Graw-Hill/Irwin.
- O'Brien, J. A. (2004). *Management information systems* (6<sup>th</sup> ed.). Mc Graw-Hill/Irwin.
- Post G. V., & Anderson D. L. (2003). *Management information systems – Solving business problems with information technology* (3<sup>rd</sup> ed.). McGraw-Hill Irwin.
- Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39 (11), 86-95.
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems (JMIS)*, 12 (4), 5-34.

## Author's Note

Comments, discussion and critique are encouraged, in particular by e-mail!

## Biography

**Zbigniew J. Gackowski** has extensive experience in industry, public administration, and numerous universities. His teaching and research uniquely bridges the gap between Central European and US experience in Computer Information Systems [Warsaw Polytechnic, The University of Michigan (*Fulbright Research Scholarship* with the ISDOS Project), Purdue University (Visiting Associate Professor), Baruch College (Visiting Professor), CSU, Stanislaus, The University of Melbourne (Visiting Professor)]. He presented papers in Poland, across Europe, United States, and some in the Middle East and South America.

His research has received much recognition. While in Poland he published over 120 research and professional items, among them: 4 books, 6 papers in refereed journals, 12 text books of limited circulation, 10 papers published in professional periodicals, 12 papers presented on national and international conferences in Europe, Asia, and South America. While in the USA he taught at three USA universities and presented 11 research papers of which 9 were published in the proceedings of the *American Society for Information Science (ASIS)*, *Association for Computing Machinery (ACM)*, *Western Decision Sciences Institute (WDSI)*, *Information Systems Educators Conference*, and *International Conference on Information Management*. He is a member of ACM, DSI and a charter member of the *Association for Information Systems* serving globally Information System academics.