LOGICAL INTERDEPENDENCE OF DATA/INFORMATION QUALITY DIMENSIONS - A PURPOSE-FOCUSED VIEW ON IQ

(Research-in-progress – IQ Concepts, Models)

Zbigniew J Gackowski
California State University Stanislaus
zgacko@athena.csustan.edu

Abstract: Current MIS textbooks and related publications are overly technology laden. They offer an over-simplified coverage of MIS fundamentals and are deficient in particular with regard to the role of data and information in business environments. A selected few of the most popular MIS textbooks serve as a relevant frame of reference on what is being taught. A product of the Information Quality Programs & Initiatives at MIT (MITIQ Program) serves as a recognized research reference. This paper attempts to address the disparity between what currently is published, what is known from the latest research, and what should be added to the subject, as well. It posits that within the context of business situations, a result-oriented taxonomy of the attributes of data/information quality is possible, and important logical interdependencies among the attributes can be demonstrated. Subsequently a simpler, economical, purpose-focused hence practical sequence for examination of those attributes can be determined. The presented framework complements and accommodates earlier findings, and in addition, overcomes some inherent limitations of PSP/IQ model in AIMQ methodology.

KEYWORDS: Data, information, quality attributes, quality dimensions, taxonomy, purpose-focused view, logical interdependency, and examination sequence

INTRODUCTION

Having taught MIS and CIS courses to business students for several decades, one has come to realize that the current textbooks are overly technology laden. They offer an oversimplified coverage of the MIS fundamentals in general and of the role of data and information in business in particular. Current research supports this view. For instance, Huang et al., [4 p. 4] say "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."

Besides other sources mentioned later, two of J. O'Brien's textbooks are used as the main references: "Management Information Systems" [8], with **six** editions, and his "Introduction to Computer Information Systems" [7], with **12** editions. Due to the exceptionally high number of editions[1], they are considered the most popular ones. Serving as recognized research references are "Quality Information and Knowledge" by K. Huang, Y. W. Lee and R. Y. Wang [4] and "AIMQ: A Methodology for Information Quality Assessment" by Y, Lee, D. Strong, B. Kahn, and R. Wang [12]. Both are products of the Information Quality Programs & Initiatives at MIT – (MITIQ Program).

[1] The number of editions is a simple and reliable indicator of popularity of textbooks and the current trends. It clearly indicates that such textbooks have been on the market for a long time; instructors used and still are using them. They were challenged many times by at least three reviewers before the publisher offered a new edition.

This paper attempts to address the disparity between what currently is being published and offered to students of business administration, what is known from the latest research, and what after a critical inquiry into the current situation could be added to the subject. For a more general examination of the attributes of information quality (IQ), the context of business decision situations is used. One assumes that based on the available data/information, pertinent decisions are made, respective actions are taken, and the results of those actions are measurable or at least identifiable. The suggested framework is based on a purpose-focused view and enables a definition of a hierarchical result-oriented taxonomy of all identified IQ dimensions. In addition, the logical interdependencies identified among them enable a break-through simplification and economy in the sequence of their examination, which is of particular value to practicing analysts. This approach on one hand complements and accommodates earlier findings, and on the other hand, it overcomes the inherent limitations of PSP/IQ model in AIMQ methodology [12].

BACKGROUND

WHAT DO WE FIND IN THE MOST POPULAR MIS TEXTBOOKS?

O'Brien [7 p. 15-16], the author of the two most popular textbooks on MIS, defines

- "Information as data placed in a meaningful and useful context" (glossary), and
- "Information quality as 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 is to examine the characteristics or attributes of information quality." With short explanations, he presents 15 attributes of quality of information within a three dimensional framework: **time, content, and form**. Malaga [6], refers readers to consultants Davenport and Prusak, and suggests six qualities of information: **accuracy, timeliness, accessibility, engagement, application (relevant), and rarity.** Dock and Wetherbe [2] suggest examining: accessibility, timeliness, relevance, accuracy, verifiability, completeness, and clarity. Alter [1, p. 162-8] distinguishes four main factors to information usefulness**: information quality, accessibility, presentation, and security**, which are further subdivided into characteristics that are more specific and illustrated with examples. Within the defined context and limitations, Alter offers the broadest presentation of the subject.

Comment: Those sources, with some exception for Alter, offer *no hints how to analyze the mentioned attributes*. Other problems can be identified immediately, too. For instance, O'Brien mentions value, but does not define it. He enumerates three dimensions: **time, content, and form**, and within them lists many attributes for consideration. If it is the recommended sequence, does time come before content? Does accuracy take precedence over relevance, as other authors claim, as well? Yes, he adds short comments to the listed attributes, but does not indicate: which are **primary or mandatory, w**hich cannot be met **fully**, which are **optional** or nice to have, where one can tolerate some imperfections without losing much of their utility, and finally which are only **subordinate** aspects of other attributes? *There is no agreement among the authors either on the level of their importance, or on the sequence of their consideration, or on the completeness of the list of attributes of data/information quality.*

WHAT DOES EMPIRICAL RESEARCH OFFER?

Huang et al [4 p. 13], unless specified otherwise, use the term "information" interchangeably with "data." After they reviewed three approaches used in literature and in business practice to study information quality (IQ) **(intuitive, system, and empirical)**, they decided to use [4 pp. 33-34] a system definition anchored in an ontological, logical foundation, and an empirical definition derived from the information consumer's perspective. Later, based on the previous research, Lee, Strong, Kahn, Wang developed AIMQ: A Methodology for Information Quality Assessment [12].

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 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 is derived. Huang et al [4 pp. 39-40] (see Table 1) identified four potential representation deficiencies with regard to four intrinsic (i.e., system-oriented) information quality dimensions (complete, unambiguous, meaningful, correct), associated them with two

Dimensions	Nature of Associated Deficiency	Source of Deficiency	Observed Information Problems
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.

Table 1 Intrinsic Information Quality Dimensions and Observed Problems
(Source: Wand, Wang [11] and Huang et al., [4 p.41])

sources of deficiencies (design and operation failure), and with some observed information problems.

Comment: There is a fundamental terminological difference 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. It is a formidable challenge for students and instructors to develop a consistent presentation of the subject.

There is a problem with **completeness** as defined in Table 1. Despite professing the ontological approach, the Huang et al, 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 mostly frequently unattainable. 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. 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.

Other dimensions such as **unambiguous, meaningful, correct** are defined within the strict context of mapping, but that constitutes only the first 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, other serious distortions cannot be entirely avoided. Only, by careful design of proper organizational procedures, decision-making procedures, and proper checks and balances, they can be minimized to some degree. Hence, they should not be ignored when discussing the subject.

B. The empirical approach from information customer perspective

Here, 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 either use or not use information [Huang, et al., 4 pp. 42-43]. Garvin [3] goes even further and says "high quality means pleasing the consumers, not just protecting them from annoyances."

Huang, et al., [4 p. 44], using qualitative analysis, examined 42 information quality projects from three leading-edge data-rich organizations that are leaders with regard to attention to information quality. Each project served as a mini case, and was analyzed using the quality dimensions listed in Table 2.

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

Table 2 Information Quality Categories and Dimensions (Source: Wang, Strong [10])

The 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. Thus, they developed a framework (see Table 2) with four **information quality categories (intrinsic, contextual, representational, accessibility)**, and with two or more associated **information quality dimensions (attributes).**

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 requirements 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. [4, 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, and improving information quality in a continuous cycle.

Comment: There is not doubt that this first-class empirical study constitutes enormous practical progress. It should find its proper place in MIS and CIS textbooks and replace the current rather peace-meal, *eclectic enumeration of attributes of information quality with no guidelines on how to examine them*. Now, however, we face 18 attributes of information quality derived from the above research, and 18, but different, listed by O'Brien (15) plus three other mentioned by other authors.

The first inconsistency 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 is listed separately 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 due to operational problems within the boundaries of the mini case. Even within the contextual category, the purely empirical approach neglects the difficult strategic aspect of information completeness – 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, **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., 4, p. 34].

C. AIMQ: A Methodology for Information Quality Assessment [12]

Probably the broadest overview of academics' and practitioners' views on IQ dimensions is in [12]. The authors admit: "Despite a decade of research and practice only a piece-meal, ad-hoc techniques are available for measuring, analyzing and improving IQ." They claim: "We developed a methodology called AIM Quality (AIMQ), that provided a rigorous and pragmatic basis for IQ assessment." The foundations of AIMQ methodology are a model, and a set of IQ dimensions, which covers aspects of IQ that are important to information consumers. For defining the IQ concepts, and ensure complete coverage, the authors use the four categories of information quality (intrinsic, contextual, representational, and accessibility) derived from empirical study of information consumers preferences. The first essential component of the methodology is the PSP/IQ model, which considers four situations derived from the combination of two factors: whether one deals with an information product or information service, and whether one is concerned to meet specifications or expectations of information users.

Comment: It is a strong model within the confines of TQM principles; however, one should not ignore its inherent limitations. It is limited to products or services, and to given specifications or preferences of information users. These limitations are substantial, when one realizes at least some of the consequences:

- Products or services are not identical with purposes, goals, and objectives of business entities.
- Specifications provided by a contracting entity may be sacred to a contractor, but they may be substantially deficient in meeting the actual business purpose.
- Preferences of information users within a business entity may deviate considerably or even be in conflict with business purposes of the entity they serve or work for.

Rational and purpose

Using a purpose-focused view, this paper attempts to formulate a framework that eliminates the limitations inherent PSP/IQ model, defines a hierarchical result-oriented taxonomy of IQ attributes/dimensions, which leads to an economical sequence for examining them. The main purpose is to help focus analysts' attention on what should be examined first from the business purpose viewpoint. Efforts spent on developing more accurate assessment metrics are secondary to the importance of a stronger, more general, qualitative framework for assessing information quality.

THE SUGGESTED FRAMEWORK

One way the human mind deals with a complex reality is by building a symbolic model of that reality. Such a model should adequately reflect that reality. One uses **computer data**, **information,** and **knowledge** in business to represent the **business reality** with which management must deal.

The conceptual framework for a more general rational analysis of the attributes of data/information quality presented here refers to the context of **business decision situations.** It requires four steps.

1. One assumes that a relatively complete qualitative cause/effect diagram, known also as a fishbone diagram, can be created to identify the major factors impacting the expected business results. These depend on meeting at least some consumer expectations and information end user preferences, but the latter are never the primary purposes of business entities. One also assumes that, based on the available data/information, pertinent decisions are made, subsequent actions taken, and the respective results are measurable or at least identifiable.
2. One conducts an impact analysis and an evaluation of the relative strength of the factors identified before. This enables a quantitative ranking of their relevance, which facilitates checking for their completeness.
3. Based on the previous considerations one develops an informational model of the decision situation under consideration. Now, one can take inventory of what is already known, given or available about different aspects of the model. This constitutes the *data component* of the model. Anything that is not represented in the data model is unknown, must be gathered and acquired by proper intelligence. This represents the *informational component* of the model. The known and the unknown aspects can be ranked by their impact on the operational outcome by any agreed measures, such as net income after taxes, retained earnings, return on investment, return on equity, cost effectiveness of services, etc.
4. Now, within a well-defined frame of reference, one embarks upon a truly purpose-focused examination of every information data/information item with regard to its practical usefulness.

A good analytical example of this approach is desirable to elucidate all of its components. This, however, would exceed the acceptable size of this paper.

USEFULNESS OF DATA AND INFORMATION

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

- For *general education* purposes any message, (which may consist of many 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 entities* only **data/information** that change the results of their operations are useful.

Hence, in a business environment the **usefulness of data/information** should be of foremost interest to end users and managers, it should be the focus discussions on **information quality**. *Currently it is not.* MIS textbooks and other publications rarely, if at all, cover fundamentals of how to define usefulness of data/information in business environment and articulate the attributes of data/information quality that determine it or contribute to it. In this presentation is assumed that the *single most important cumulative measure* of ***usefulness of data/information, information service, or product*** offered is its **expected cost effectiveness** assessed from the viewpoint of the purpose of the business entity it serves.

ATTRIBUTES OR DIMENSIONS OF DATA/INFORMATION QUALITY

Most textbooks and the empirical research list under different names many attributes or dimensions of information quality for consideration. The major question is, however, how to examine those attributes in real life situations. Which of them: affect the business results directly or indirectly, are primary or secondary, mandatory or optional, should be examined first, or are not fully attainable and one must make trade offs?

Taking into account the sources referred to in this paper, one faces about 25 attributes of data/information quality or their equivalents: dimensions, factors, or characteristics. A critical rational inquiry into the plethora of those attributes leads to an insight that a ***hierarchical result-oriented taxonomy*** can be defined as follows (see Table 3):

- One can subdivide them into direct attributes and indirect or subordinate attributes. Changes to the ***direct attributes*** directly affect the results of business operations, while ***indirect attributes,*** as the name suggests, determine/contribute to the direct attributes, hence indirectly affect the results.
- The direct attributes can be further subdivided into primary and secondary attributes. Changes of the ***primary attributes*** result in *qualitative* changes to the decision situations under consideration, while changes to ***secondary attributes*** only *quantitatively* change the business results.
- Within the primary attributes, one must distinguish the mandatory versus the desirable ones. The ***mandatory primary attributes*** constitute non-negotiable requirement pertaining to each individual data/information item. If any of them is not met, the corresponding item must be excluded from further examination. The ***desirable primary attributes*** cannot be mandatory only desirable for by practical or cognitive constraints they are rarely-to-never attainable.

Categories of Data/Information Quality Attributes			
Direct			Indirect
Primary		Secondary	....
Mandatory	Desirable	....	....
....	....	....	....

Table 3. Schema of hierarchical result-oriented taxonomy of attributes of data/information quality

Analyzing the existing logical interdependencies among at least the primary attributes of data/information quality one may arrive at a purpose-focused, simplified, time saving, and more economical examination sequence of those attributes. This offers not only a better understanding of the phenomenon of data or information quality, but also a tangible practical benefit for all who analyze it, whether theoretically or practically. It is a progress in comparison to the known publications of other authors on this subject. It accommodates earlier findings, and overcomes some limitations of PSP/IQ model in AIMQ methodology [12]. It immediately focuses the attention of analysts on what should be considered first and provides them with a reference point to how much attention should be given to each item. Most authors of MIS textbooks do not go beyond eclectic piece meal enumeration of some attributes of data or information quality and the sequence of their enumeration is usually of undefined logic.

PRIMARY ATTRIBUTES (Their changes result in *qualitative* changes to decision situations)

Interpretable (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.** In practice, this 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 mandatory primary 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 must be excluded from any further examination.

Interpretability is contextual, e.g., a more educated recipient, a trained one, or a different receiving device may be able to interpret it. Within the some context, some authors mention legible. Of course, legibility is a mandatory precondition of interpretability. 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 these are nice to have, and they may actually add value. They are, however, either subordinate or contributing factors to other primary attributes or they are of secondary nature. The latter, in this paper, are listed under the collective name **economically ease to use,** except for conciseness, which is an aspect subordinate to relevance.

It may appear that this requirement may not pertain to data. In real life situations, even what generally is known unexpectedly may become unknown due to forgetfulness, misplacement, database corruption, communication lines failure or simply due to loss of proper encryption key.

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

The content of messages and even individual data/information items must be *significantly* ***relevant*** to the decision situation under consideration. This means that it affects the decision situation, and subsequently significantly changes the operational results of the decisions made and the corresponding actions taken. If not, its remaining attributes are irrelevant, too. Hence, relevance is the *second mandatory primary* 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 is to the usage of any specific data/information value, whether its impact is *significant* enough to warren 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 cease its further examination.

O'Brien lists separately **currency**, or **pertinence to the proper time**. After a short reflection, however, one can easily see they are only subordinate 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 some reflection one can clearly see that it constitutes only another 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, for instance when the surprise effect is lost, it also becomes irrelevant. Hence, security issues may become important preconditions 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 aspects *subordinate* to **relevance** of information, hence should be considered while examining relevance.

[2] **Significantly –** above an assumed or acceptable threshold level

Critically timely (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. 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 is not available when needed, it does not make sense to ask about any remaining attributes of quality. According to Huang et al., [4] research, information consumers perceive lack of timeliness as an **accessibility** problem. Such a perception comes naturally, when on relies mainly on computerized information systems.

There is, however, another aspect. With regard to **timeliness** of information one must make a clear distinction between beeing **critically timely** and **economically timely.** The former means that it is sufficiently on time to make a decision and act accordingly. Being **economically timely** means that there is additional time available to improve the decsion making and the preparation of the subsequent acitions aimed at attainment of optimal results. The latter should be analyzed from the viewpoint of cost effectiveness, if possible. Therefore, in this paper it will be discussed after the examination of all primary attributes of data/information, and the proper ranking of all the factors under consideration. Certainly, **critical timeliness** is the third *mandatory primary* attribute of individual data/information items, after relevance. Within the context of timeliness Alter [1 p. 163] mentions age of information. It is of no value, when one does not know how volatile the factor is. Simply, it is a subordinate aspect of accuracy.

* * * * *

This is the last of the mandatory primary requirements of quality of data/information. The sequence of their presentation is determined by the logical relationship where meeting of one requirement 'is a precondition' of examining the next attribute in the sequence. For instance, when a data/information item is not interpretable, it precludes the examination of any other attribute of this item. Once it is interpretable, the only logical question that should be asked is whether the item is significantly relevant. If it is significantly relevant, it makes sense to ask whether it can be critically timely available. To be concerned whether it will be critically credible would be foolish. It is a difficult question to answer, hence one should make sure first that it can become available, when needed. Following this pattern, the examiner minimizes the number of considerations and the time spent on conducting the examination.

Critically credible (Intrinsic category – by Wang, source - by Alter)

Only messages declared significantly relevant and critically timely available are worthy to test their **credibility,** that is whether they are true, whether they can be relied on. The adjective **true** means consistent with reality. While probing a message or a single data value whether it is true, one should examine how credible, believable, or reputable the source is. There is a need of additional distinction, however, whether the data/information is *critically* credible. ***Critically credible***[3] information can be defined for practical purposes as a level of credibility, at which decision makers are willing to take action in response to it. If not, it will be ignored. Consequently, the latter changes the decision situation qualitatively, for at least, with respect to that item the decision maker decides to gamble. Thus, the situation becomes a game, which certainly constitutes a qualitative change.

Let us not forget that sufficient assurance of credibility is sometimes impossible or at least not cost effective. Full credibility is rarely-to-never attainable. Hence, critical credibility is only a ***desirable primary attribute***. If critical credibility cannot be assured, other considerations are unnecessary. Many authors, within the scope of the term credibility discuss the issues of bias and/or accuracy and precision of presen-

3 A tragic example of what means to be critically credible are the many warning signs and openly declared threats, known before the 9/11 terror attack, but ignored as insufficiently credible, for they were nearly beyond imagination.

tation of information. Because they are scalable and never fully attainable, they should be examined only later after proper ranking of all the significantly relevant factors while examining their completeness.

Acceptably complete (Contextual category - by Wang, content dimension - by O'Brien)

Completeness of data/information pertains to the *totality* of all identified significantly relevant factors. By definition, it cannot be an attributed to *individual* data/information items. Once one arrived at a set of significantly relevant, critically timely available, and critically credible data/information items, one can embark upon testing their completeness with regard to the entire decision situation under consideration.

The problem of completeness of data or information is more complex than it appears on its surface. Completeness is strongly related to relevance, for relevance is the primary attribute of information quality that lends significance to each data/information value under consideration. One must distinguish at least two types of completeness: **operational completeness** and **congnitive completeness**.

Within the context of decision situations, ***operational completeness*** measures the degree to which the significantly relevant data/information values are available. Operational completeness may be measured in percentage points [1 - 100%] as the ratio of the sum of all results that can be attributed to the corresponding relevant data/information available and the sum of results attained. Now, one can perform a cursory completeness check. In real-life situations, usually, some residual operational results will remain unaccountable. This means it is not possible to attribute them to any previously identified factors. They may be used as a relative or absolute measure how incomplete the impact analysis is. **Figure 1** illustrates the general interdependence between **relevance** and **operational completeness** of data/information items.

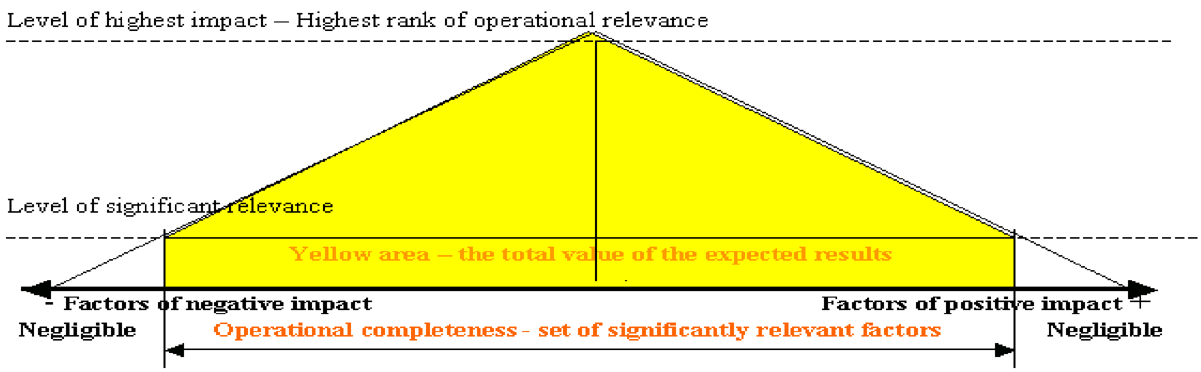

Figure 1. The relationship between significant relevance and operational completeness of data or information items pertaining to the corresponding factors that determine and contribute to the total value of the expected results in a decision situation under consideration.

Murkier, however, is the qualitative or **cognitive** aspect of **completeness** of data/information in a decision situation under consideration. In cognition and research, there is a general rule that qualitative considerations 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. The critical blow most frequently comes from a danger or direction not identified and recognized in time. Therefore, the **qualitative or cognitive completeness** is a only a ***desirable primary attribute*** of all the significantly relevant data/indormation items in specific decision situations. It cannot be considered mandatory, for it is rarely-to-never fully attainable. In real life situations, in the fight for survival, on a battlefield or in global business competition, one may never be certain whether all relevant success factors or dangers are identified and evaluated. Prudence, however, requires, whenever possible, to gather more information in order to assess all the may-be-not-yet perceivable but potentially possible critical factors for planning of counter measures and contingency provisions. **Figure 2** illustrates the general interdependence between **rele-**

vance of data/information items on all identifiable hypothetical factors pertaining to a decision situation under consideration and the fuzzy notion of their ***cognitive completeness***.

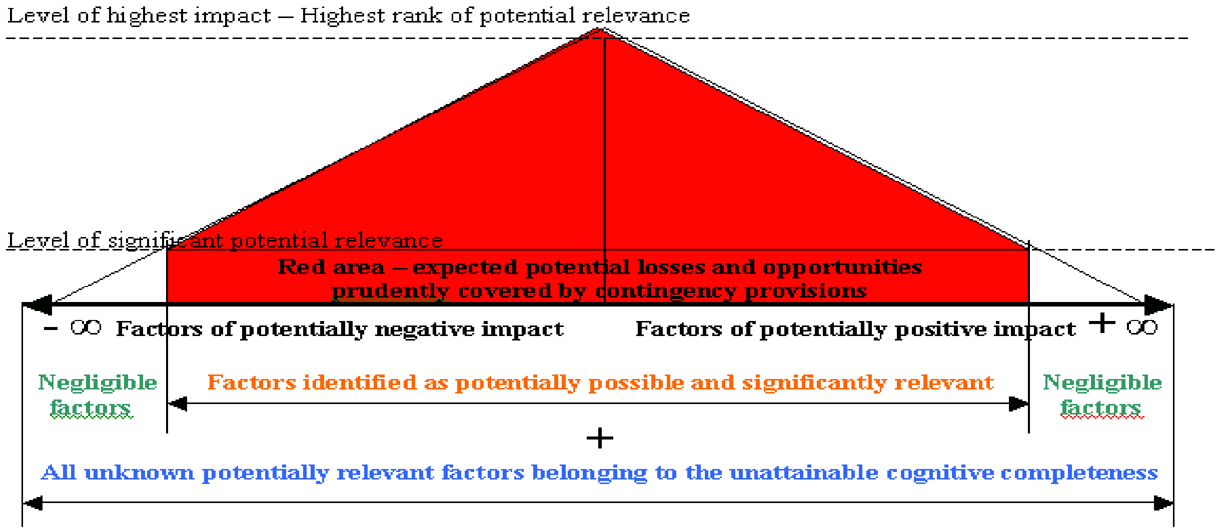

Figure 2. Relevance of data/information items about identifiable potential factors pertaining to a decision situation under consideration and the unattainable notion of their cognitive completeness.

Both Figures (1 and 2) illustrate how quantified or at least ranked relevance of data/information determines both aspects (operational and cognitive) of completeness of the totality of factors pertaining to a specific decision situation. For both are rarely-to-never fully attainable, one must resign oneself only to an ***acceptable*** level of ***completeness*** in both aspects. One may ask how sensitive a model of a decision situation is to the use of any specific data/information value, whether it has a significant impact on results worthy of consideration. The quantified and ranked relevance provides the examiner with a reference scale. It suggests how much attention one should pay to each data/information item in comparison to the remaining secondary attributes of its quality such as optimum level of timeliness, accuracy meant as free from random errors, precision, and finally ease and effectiveness of use. Any changes in this respect may only *quantitatively* change the results and/or the cost of business operations, hence the cost effectiveness of each data/information item used.

* * * * *

This above enumeration and discussion closes the list of the five primary attributes of data/information quality. The first three (interpretable, relevant, critically timely) are the *mandatory primary attributes* and the remaining two (critically credible, acceptably complete) are the *desirable primary attributes*. One must be, however, fully aware that such a term as *critically* timely strongly depends on the way one arrives at decisions. Is the decision made by an individual or by many participants in the process, whether horizontally or vertically? Similarly, what is *critically* credible or *acceptably* complete strongly depends again on the personal traits of the decision maker; is he/she averse to risk, passive, indifferent, hesitant, cautious, prudent, motivated, jumpy, etc. What follows will be a list and discussion of the secondary attributes of data/information quality.

SECONDARY ATTRIBUTES (Their changes result only in *quantitative* changes to business results)

The sequence of examination of the secondary attributes of data/information quality is practically irrelevant for each of the following entries leads only to fine-tuning with regard to the quantitative results of business operations. They can be economically evaluated only after proper ranking of the relevance of the factors represented by those data/information values, hence, after their relevance was ranked and their completeness examined.

Economically timely

Meeting mandatory requirements usually does not add value; it makes the data value only acceptable. Timeliness, however, is also scalable. One may receive the necessary data or information not only on time, but also more or less in advance. The additional time may be used for making decisions with less haste, and/or for better preparation of actions. Hence, one may obtain better results when additional time is available. Additional time may add value. There is no analytical formula, however, in specific situations it may be either possible to calculate or improve experimentally. Whether it is worthwhile, it depends on how much difference in results it will make, and on how much it will cost to accelerate the informing process or increase its frequency.

Economically unbiased

Credibility of data/information discussed before in the aspect of its truthfulness may also be compromised by lack of **objectivity** or **bias** in the data acquisition process due to approaches and methods used in selecting the primary sources, measuring points, observation points, and finally collecting, processing and presenting data. The resulting distortions 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, very deceptive and damaging. To rectify the bias and compensate for it may require engagement of additional substantial resources. Whether it is justified, it can be estimated only when the size of its impact on the results is serious enough.

Economically accurate

Another problem is inaccurate representation of reality; how **faithful** the mapping from real world states to respective data values was **(complete[4], unambiguous**[5], **meaningful[6]**, and **correct[7]** as accurately described by Wand and Wang [11]**),** and how free it is from random errors. One encounters **random errors** in all measurements and observations. A typical gross measure of **inaccuracy** in this sense is the **error rate**. One calculates it by dividing the number of values in error by the total number of data or information values gathered. In practice, a more useful measure of inaccuracy due to random distortions is the **expected cost of dealing** with the consequences of those errors. One may calculate it by multiplying the number of data/information values by the probability or frequency 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 serious the consequences of each type of error are. One may reduce many of them 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 be experts in dealing with such a situations, but they should be taught to recognize the need for preventive measures.

Economically precise

Finally, inaccurate representation of reality may be due to low precision of data/information values used. For numerical data, precision is measured by the number of significant digits used. Precision of pictures and images one measures by the number of dots per inch. It is commonly used to describe the precision of printers, computer screens, scanners, etc. Insufficient precision of data or information presentation may compromise the results obtained.

There is a trap associated with accuracy of information understood as free from error, and as precise in its presentation. Generally, it is overrated [10]. U*nchecked* efforts to increase the level of accuracy of any-

---

[4] Lost data value for existing real world states.
[5] Multiple states of the real world not mapped to the same state of data value.
[6] Real world state mapped to a meaningful data value
[7] Real world state not mapped to a wrong data value

thing can become counterproductive. The ultimate determination of the indispensable and economically justified level of accuracy of data/information strongly depends on its utility value.

Figure 3 illustrates graphically how the net business utility value of data/information changes as a function of its accuracy. There one can see two graphs plotted as functions of the level of **accuracy A** in percentage points [1 - 100%], for its level affects both the numerator and the denominator of the cost effectiveness ratio.

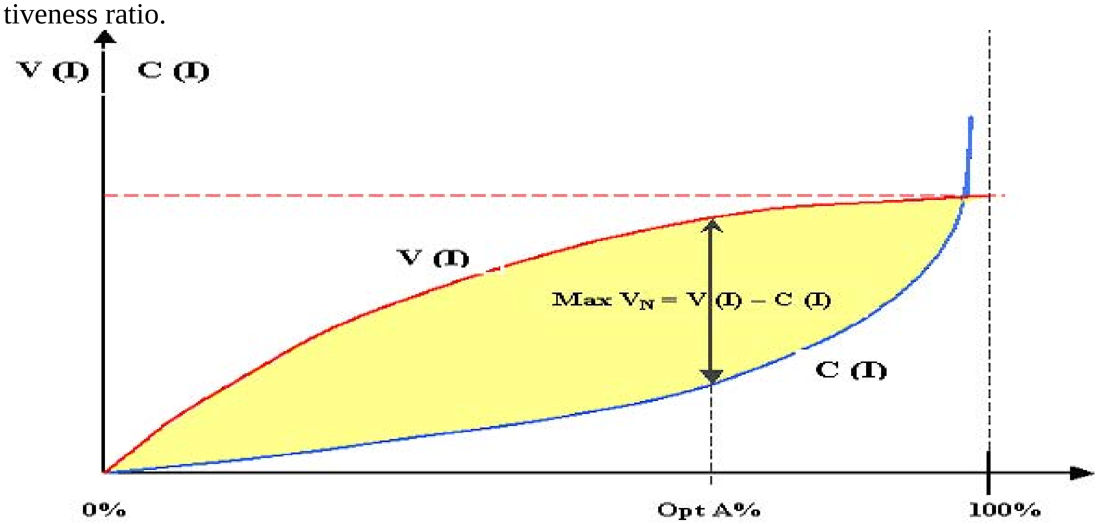

Figure 3 Optimum level of Accuracy A as function of Net utility value V_N of information I

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.

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. One thing is sure, the accuracy of data/information meant as free from bias, errors, and insufficient precision should be postponed *nearly until the very end*.

Economically easy to use

Ease and cost effectiveness of use of data/information is a collective name for all aspects related to its format and mode of delivery. It may affect how fast the end-user may read, interpret, comprehend, analyze, draw conclusions, and act upon it. 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. It may, however, increase or decrease its ease of use and/or its procurement cost, for both subsequently affect the expected cost effectiveness of data/information, which in business environment should be the ultimate determining measure of data/information quality.

RESULTS AND CONCLUSIONS

Table 4 summarizes the results and conclusions from this rational inquiry into the logical interdependencies among dimensions/attributes of data and information quality discussed in the referred MIS textbooks and related research, when viewed from a **purpose-focused perspective** in business environments.

1. A **hierarchical result-oriented taxonomy** of data/information quality dimensions was defined with a general demonstration how applicable it can be.
2. The taxonomy above supports a framework for a contextual consideration of the multitude of dimensions of data/information quality identified by different authors. It *overcomes the inherent limitations* of PSP/IQ model in AIMQ methodology, for it is business **purpose-focused** in contrast the orientation on products, services, users' preferences, and requirement specifications.
3. It complements and accommodates the earlier findings[8]. In addition, it enables a simpler and more *economical order* of examination of IQ dimensions by exploiting the **logical interdependencies** among them and using their ranking for providing the analyst with a point of reference to how much attention they deserve.

Since this is still a research-in-progress, at least two limitations can be identified. At the same time, they constitute potentially promising directions for further progress. For instance, elaboration of a *dependency map of IQ dimensions*, which shows explicitly the most important ones among all identified and suggested dimensions, could facilitate considerably a faster and more economical sequence of their examination. Similarly, an extension of the proposed result-oriented taxonomy by combining, incorporating or overlaying it with a *cognitive taxonomy of factors impacting business results* is very promising. It enables to elevate the assessment of IQ from mainly operational level to the *strategic level* of applications in business and administration, including applications related to national security. The results may be offered for presentation at one of the subsequent ICIQ Conferences.

[8] The proposed framework (see Table 4) easily accommodates current findings, including those listed by AT&T and Redman [13], and overcomes the inherent limitations of PCP/IQ model [12, Table 3] in AIMQ methodology.

				Examples of attributes of data/information quality		
				***Direct* Attributes**		***Indirect* Attributes**
Sequence of examination		***Primary Attributes***	***Mandatory* Attributes**	**Interpretable**		Legible, user trained, untrained, educated
				Significantly relevant		Concise, current, admissible, secure, appropriate amount
				Critically timely		Obtainable, accessible, style and mode of decision making, individual or collective
			***Desirable* Attributes**	**Critically credible**		Believable, reputable, decision maker's traits: risk averse, passive, hesitant, cautious, prudent, motivated, jumpy
				Acceptably complete (for totality of factors)		
	Irrelevant	***Secondary Attr.***		**Economically**	**Timely**	Frequency, how much in advance
					Unbiased	Sampling, observation points,
					Accurate (error-free)	Mapping (complete, unambiguous, meaningful, and correct), granularity, age
					Precise	Number of significant digits, dots
					Easy to use	How summarized, detail, text, graph, diagram, picture, media, clarity, order, consistent, homogene, understandable, natural, efficiently encoded

Table 4. Example of a hierarchical result-oriented taxonomy of data or information quality attributes in economical sequence of their examination

REFERENCES

[1] Alter, S, *Information Systems – Foundation of E-Business*, Prentice Hall, 2002.

[2] Dock, V. T., Wetherbe, J. C., *Computer Information Systems for Business*, West Publishing Company, 1988.

[3] Garvin, D. A., "Competing on the Eight Dimensions of Quality," *Harvard Business Review*, 65(6), 1987, pp. 101-109.

[4] Huang, K., Lee, Y. W., Wang, R. Y., *Quality Information and Knowledge*. Prentice Hall, NJ, 1999 (Quality Programs & Initiatives at MIT - MITIQ Program.)

[5] Kofler, E., *O wartosci informacji (On Value of Information)*, Panstwowe Wydawnictwa Naukowe (PWN), Warsaw, Poland, 1968.

[6] Malaga, R. A., *Information Systems Technology*, Pearson Prentice Hall, 2005.

[7] O'Brien, J. A., *Introduction to Information Systems* (Twelfth Edition), Mc Graw-Hill/Irwin, 2003.

[8] O'Brien, J. A., *Management Information Systems* (Sixth Edition), Mc Graw-Hill/Irwin, 2004.

[9] Post G. V., Anderson D. L., *Management Information Systems – Solving Business Problems with Information Technology* (Third Edition), McGraw-Hill Irwin, 2003.

[10] Wang, R. Y., and Strong, D. M., "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems (JMIS)*, 12(4), 1996, pp. 5-34.

[11] Wand, Y., and Wang, R. Y. "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM*, 39(11), 1996, pp. 86-95.

[12] Lee, Y., Strong, D., Kahn, B., Wang. R. "AIMQ: A Methodology for Information Quality Assessment," *Information & Management*, December 2002, Volume 40, Issue 2, pp. 133-146.

[13] Redman, T. C., Data Quality: Management and Technology, Bantam Books, New York, NY, 1992.