

Diagnostic and Functional Dependencies of Credibility

Zbigniew J. Gackowski
California State University, Stanislaus

ZGackowski@csustan.edu

Abstract

This is an inquiry into the “actionable credibility” of information values in informing and how it depends on indirect attributes of information quality. Empirical survey-based research by Wang and Strong (1996) ignores the multidimensional aspects of credibility. Studies that ignore them may produce unreliable results. Most publications discuss quality attributes as independent factors. This paper identifies and describes the ignored dependencies with regard to credibility. To yield research results of a more lasting validity, one must go beyond empirical studies. This inquiry continues the development of a theoretical model of operation quality requirements of data and information values as proposed by Gackowski (2004, 2005a, 2005b).

Keywords: Operation quality of information, actionable credibility of informing

Introduction

In operations, *actionable credibility* is the critical level (percentile) of credibility which triggers a state transition in operations – compels to react. Depending on the **relevance** and **significant materiality** of the received information of uncertain **credibility**, when it exceeds the threshold percentile, it compels decision-makers to act, cease to act, and/or change the manner of acting. **Actionable credibility** is the most complex direct primary universal quality requirement of information values. In the empirical survey-based study by Wang and Strong (1996) about how data users perceive quality, BELIEVABILITY (here labeled credibility) of data was ranked the highest. As most of similar research, it ignores the multidimensional aspects of credibility and its complex diagnostic and functional dependencies on pertinent *indirect* quality attributes. The latter are treated on equal footing with the *direct primary* and *direct secondary* quality requirements, as defined in the impact-focused universal taxonomy of operation quality requirements (Gackowski, 2004).

This paper identifies and describes these interdependencies. It demonstrates how empirical studies that ignore them produce unreliable results. In research, to attain more than situation-specific

Material published as part of this journal, either online or in print, is copyrighted by the publisher of the Informing Science Journal. 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 and payment of a fee. Contact Editor@inform.nu to request redistribution permission.

practical improvements in quality, one needs to identify the qualitative dependencies first. Before embarking on quantitative studies, a qualitative model of diagnostic and functional dependencies of credibility on indirect quality attributes (for brevity later called **qualities**) of information should be developed. This paper:

- Identifies the factors that affect credibility in informing

- Defines credibility of values as a function of many indirect quality attributes
- Maps the identified diagnostic and functional dependencies of credibility into entity-relationship (ER) diagrams, which
- Suggest, a logical economic sequence of examining the qualities affecting credibility

For focused reading, key terms in paragraphs are in **bold** font, emphasis in *italics*, highest emphasis underlined, definitions in ***bold italics***, and the labels used by Wang and Strong (1996) in ALL CAPS.

Credibility of Data/Information Values in Literature

Operation quality requires a rigorous distinction between **data** and **information** values, which otherwise may not be required. Decision makers and/or acting agents already know some aspects of the situation. Then, ***data values*** represent *known* aspects of reality that are **given**, or **assumed true**. To the contrary, ***information values*** represent *unknown* things, events, and states of the environment that need to be gathered. Of course, once they have been gathered (collected, acquired) and recognized as valid, they become data values for the entities informed (clients/users). Information values always *change* the decision situation itself, and/or the actions necessary to implement the decision made, and/or the operations results. Data values never *change* anything, for they have or should have been already accounted for. If not, it is an omission in modeling of the reality. In routine operations, credibility of data values is not less important but is of less concern to individual entities informed. In indirect informing (Gackowski, 2005 and 2006), data/information values are subject to established procedures to assure their integrity during acquisition, entering, storing, retrieving, and presenting. It is done on behalf of all the data-sharing entities informed.

Credibility of information values is a complex notion of the degree or extent to which they represent the actual reality to users in operations. The Financial Accounting Standard Board (FASB) (1983) and Gleim (2004) use the term “reliability” as “the quality of information that assures that information is reasonably free from error and bias and faithfully represents what it purports to represent.” Wang and Strong (1996) define BELIEVABILITY as “the extent to which data are accepted or regarded as true, real, and credible” (p. 31). Gackowski (2005b) defines “credibility” of information as “whether it is true, whether it can be relied on as true representations of reality.” The definitions overlap. Which label communicates better?

Believability stands more for what is believed than for what it objectively should stand for. It is the weakest term. In scholarly and technical writings, reliability is associated with reliability of systems with a complex statistical theory. Information values are symbolic representations of the changing reality. Hence, **credibility** seems to be the better choice. The term “reliability” will still be indispensable in its natural context as reliability of entire data/information delivery systems (information systems, database systems, data warehouses, and Web-page-based intranet and internet informing systems). Hence, reliability of systems is a separate non-trivial subject.

How credibility is treated in practice and research? Is credibility one out of many independent qualities, does it play any special role in operations, is it an independent factor, or is it a complex function of many other qualities? Wang and Strong (1996) empirically studied how consumers view data quality. In the answers provided by respondents, BELIEVABILITY was perceived as the most important quality attribute and ranked the first by importance and within their “Conceptual Framework of Data Quality” as well. Respondents were provided with a scale of importance 1–10 with no criterion in what respect important. This framework was cited in many publications, among them, by Huang, Lee, and Wang (1999) and Lee, Strong, Kahn, and Wang (2002). From the operations viewpoint of quality (Gackowski, 2005b), the results of ranking importance of

BELIEVABILITY as number one and RELEVANCE as number three contradicts each other. When importance of a value is measured by its impact on operations (materiality in finance), by the principle of its pervasiveness, other qualities of the same value cannot rank higher. Obviously, at the time of the discussed study, this was not known or recognized. Nevertheless, the broadly held explicit or implicit assumption of independence of quality attributes lives on.

BELIEVABILITY was assessed on equal footing with many of its contributing factors, such as accuracy, objectivity, timeliness (meant then as currency), traceability, reputation, variety of data sources, etc. Wang and Strong (1996) identified 179+ data quality attributes, later reduced them to 118 for subsequent factor analysis. Finally, they compressed them to into 20, more manageable, parsimonious, and important. Among the 118 labeled or partially described data quality attributes, about 23 are somewhat related to credibility.

The operations approach to quality (Gackowski, 2005b) identifies credibility as a direct primary universal quality requirement. On one hand, as a necessary or mandatory one, it is of highest importance, but equal with all other primary qualities. On the other hand, it is the fourth or fifth one in the diagnostic sequence of their examination. The diagnostic and functional dependencies of credibility on indirect qualities are the subject of this inquiry. It resulted in an analytical model and mapping of these interdependencies. The goal is to develop an algorithm for an intelligent information quality analyzer.

Credibility as Function of Indirect Qualities

The model identifies the *place* of credibility among other quality requirements, defines credibility of information, and the factors that *directly and indirectly* affect credibility. It suggests an economical *sequence of examining* the contributing factors, provides a simplified overview of the suggested diagnostic examination of credibility (Figures 1, 2, and 3), and a schematic entity-relationship diagram of the identified diagnostic *prerequisite dependencies* of credibility on the identified indirect qualities of the *first* order (Figure 2a), the *second, and subsequent* orders (Figures 2b, and 3).

Credibility among other Operation Quality Requirements

Presently, the theoretical model of operation quality of information (Gackowski, 2005b) comprises 14 essential components:

- The impact-focused universal taxonomy of all identified and not-yet-identified quality requirements, where **actionable credibility** is classed as a **direct primary (necessary) universal quality requirement**
- The definitions and descriptions of the five or six universal operation quality requirements of information, and
- The definitions of several universal principles, to which quality requirements are subject such as:
 - The principle of **relativity** implies that **actionable credibility** is determined by the **purpose, circumstances, and criterion of effectiveness** of operations
 - The principle of **pervasiveness** of materiality of factors implies that **materiality of actionable credibility** cannot exceed materiality of the affected factor
 - The principle of **equivalency** of a lost data value and an unavailable information value implies that **actionable credibility** applies equally to **data and information** values

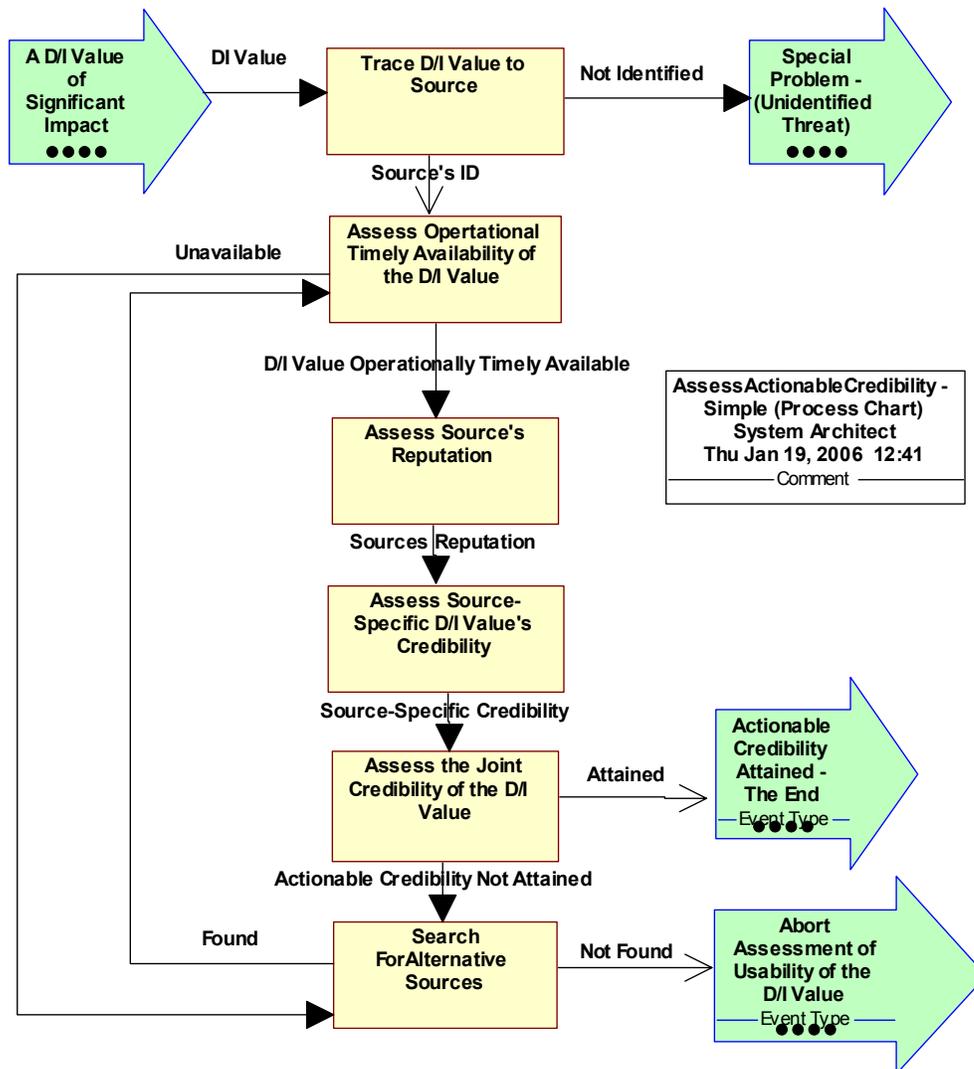


Figure 1. Examine Credibility of a Data/Information Value - Process Chart

- The principle of **operational usability** of values requires **actionable credibility** be one of the four five direct universally necessary quality requirements for any usable value.
- The principle of **degradation** of decision situations implies the following: If **actionable credibility**:
 - Is *certain*, jointly with other primary quality requirements, decision-makers deal with a **deterministic** situation in the area affected by the value
 - Is only *probable* (the most likely case), decision-makers deal with a **stochastic** situation in the affected area
 - Is *not met*, decision-makers **game** in the affected area even when ignoring the factor, for instance a threat.

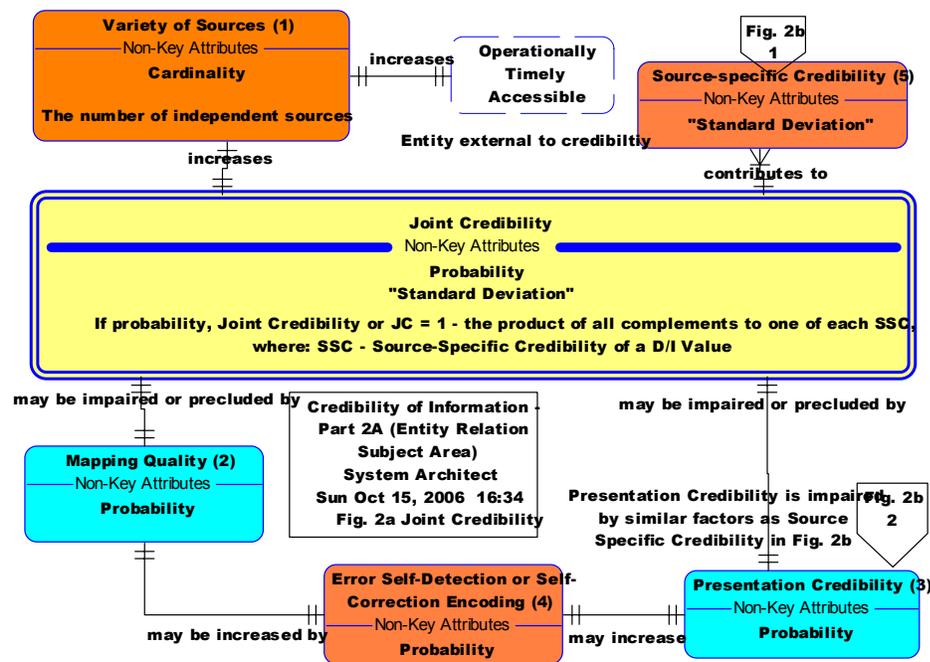


Figure 2a. Joint Credibility as a Function of Required Indirect Quality Attributes
(Entity Relationship Diagram - A Provisional Qualitative Model - Part A)

Credibility of Information in Operations

Credibility of a value means whether it is true—whether one may rely on it. The adjective *true* means consistent with reality. Complete credibility is rarely-to-never attainable. It may be Boolean (true, false) or its degree may be measured by the probability of its veracity $\langle 0, 1 \rangle$. While probing for veracity, entities informed face dramatic options, less with data of well-established roles and more with unknown recently acquired information. The big question is whether they received **disinformation**, **misinformation**, or **valid information**. To this end, outright disinformation must be excluded first and the degree of misinformation assessed next. **Disinformation intentionally** misinforms. **Misinformation unintentionally** misrepresents reality. **Valid information faithfully** represents reality.

All the indirect quality requirements, which affect actionable credibility, will be discussed below. As much as feasible they will be illustrated with examples. Some prefer a single scenario. As it becomes obvious later, this is impossible. There are groups of universally necessary indirect quality requirements. When only one of them cannot be met, the remaining ones need not to be examined. This fact renders it impossible to find one scenario, in which all of them should be examined.

Indirect Factors Directly Affecting Credibility

Affecting factors may improve or impair credibility. Only the **variety** or the **number of independent sources** increases the degree of credibility of a value. It implies that within a society or any group of individuals, the focus should be on creating and maintaining a viable variety of alternative sources of information. By doing so, one increases not one but two direct primary quality requirements: **operational timely availability** and **joint credibility** of information. Lack of access to alternative sources and corresponding communication channels makes one vulnerable to bias, disinformation, and disruption of communication. Because credibility is rarely-to-never

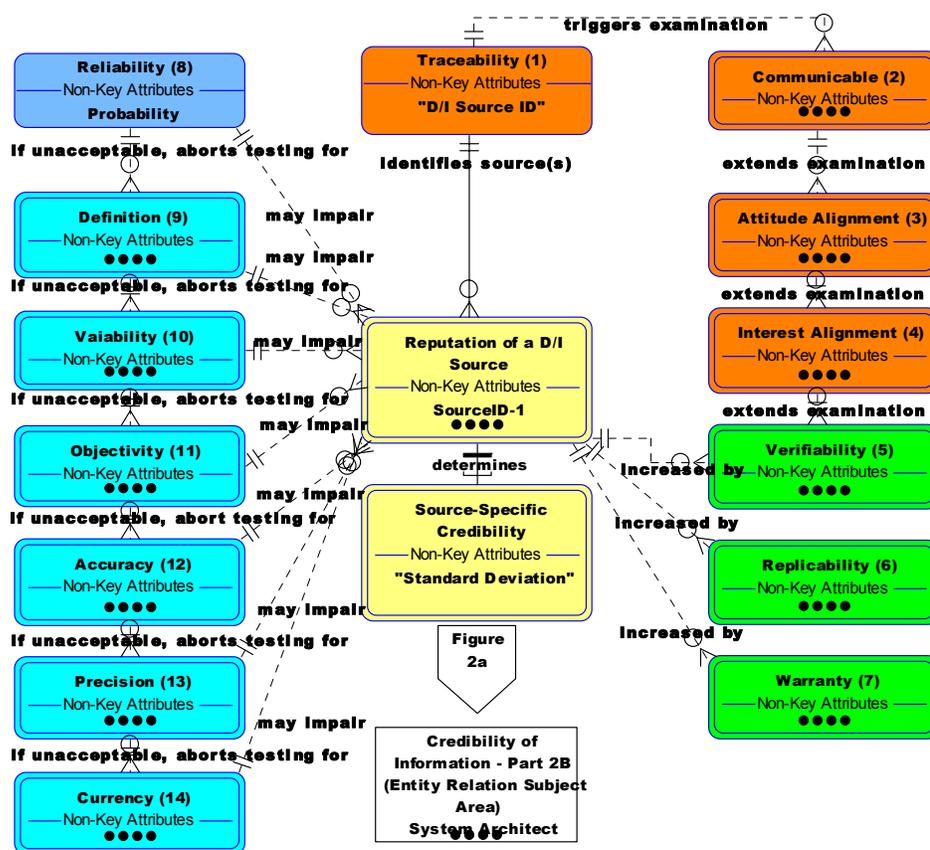


Figure 2b. Reputation of a D/I Source - Diagnostic and Functional Dependencies on Required Indirect Quality Attributes (Entity Relationship Diagram - A Provisional Qualitative Model - Part B)

fully attainable, in cases of grave consequences, one should *avoid relying* on a single source. A single source, witness, intelligence report, etc leaves decision-makers very uncertain. In science, all measurements must be replicable. In courts, corroborating evidence is indispensable even when the defendant pleads guilty. Violations of this rule led to grave consequences.

Sources of information values may be observers, sensors, processes, tests, etc. Sources may remain unknown despite all efforts to identify them (perfectly anonymous threats, threats under false identity, or only rumors of threats). These are special situations, when assessment of sources is impossible.

Any value acquired from a specific source *inherits* credibility labeled **source-specific credibility (SSC)** from the reputation of its source (a fraction between $<0, 1>$.) The uncertainty related to source-specific credibility or lack thereof is its complement to one ($1 - SSC$). With additional sources, the joint uncertainty declines fast. If many sources generate the same or similar value, the uncertainty $1 - JC$, where JC is the **joint credibility** of the value, is also declining fast as the product of all the source-specific uncertainties (SSC) as follows: $1 - JC = \prod (1 - SSC_j)$ for all j , which implies

$$JC = 1 - \prod (1 - SSC_j) \text{ -- over all independent sources } j \quad (1)$$

Even when all sources are inaccurate, there is possible to obtain more accurate data from them. There is research in progress about a **theory of complementarity** of extracting more accurate data from inaccurate sources through integration (Gelman, 2005).

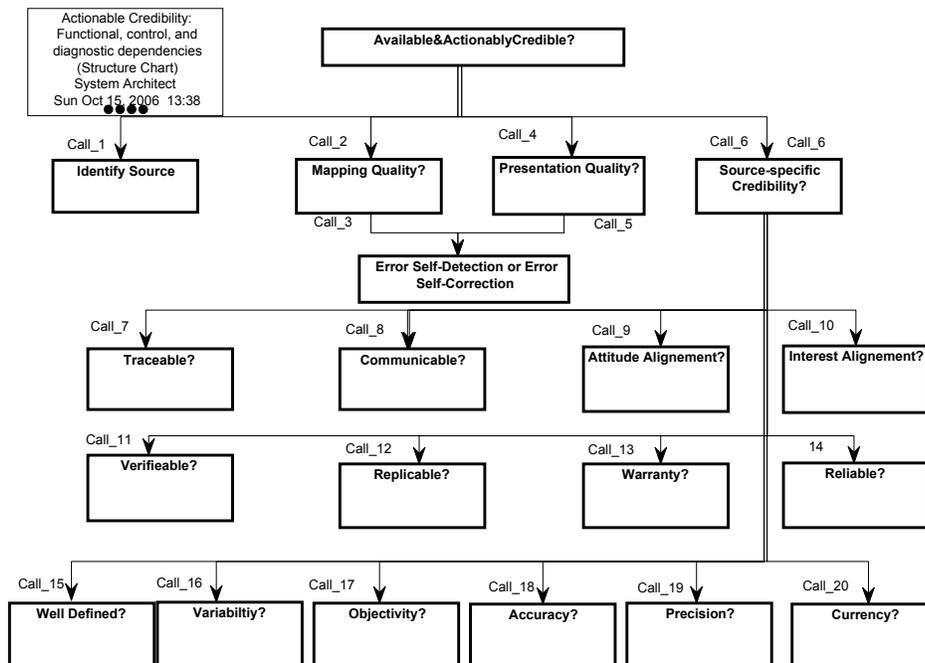


Figure 3 Actionable Credibility: Functional, control, and diagnostic dependencies on indirect qualities (data/information quality attributes)

Actionable credibility of a value is attained when joint credibility (JC) equals or exceeds the threshold (percentile) for a state transition of operations as defined by the policy in effect. It may mean: to act, cease to act, or change the way one acts. **Economical actionable credibility** imposes additional economical requirements, as described by Gackowski (2005b). For instance, the cost of attaining it should not exceed the benefit the value adds to operations. When materiality is very high, one may decide to take preventive measures even when the credibility of such threats is very low and no corroborations are available because the **expected** pay-off is still high. **Expected** materiality of a threat equals its materiality multiplied by the credibility of the threat (its probability). Then the level of actionable credibility may be set low.

The **joint credibility (JC)** of a value gathered from more than one source may be impaired by imperfections in **mapping quality**, as defined by Wand and Wang (1996, p. 92), within the data and information delivery system and **presentation credibility** for entities informed.

Mapping quality (not the label but the notion) of real-world states into corresponding states of information systems was defined by Wand and Wang (1996, p. 92) as a function of four (intrinsic to the design and operations of information systems) quality dimensions: **complete, unambiguous, meaningful, and correct**. Each dimension was precisely defined and explained with regard to the sources and nature of their deficiencies. They were mislabeled as “intrinsic data.” They are intrinsic to the system design and operations. Here, by the principle of relativity, no physical property of any factor is worthwhile considering on its own merit, if it has not acquired relevance and significant materiality from the purpose and circumstances of operations; otherwise, they are not part of the model of operation quality. Mapping quality can be measured by the *probability* of being free from mapping errors represented by a fraction between $\langle 0, 1 \rangle$. This quality aspect is affected by data entry, data definition, and precision of the system. Nevertheless, computers rarely err, and when they fail, they are designed to warn the user, while human data entry is the single source of most of the errors one encounters in computing.

Presentation credibility is a component of presentation quality; it may impair the credibility of the values as beheld by users. Usually, in indirect informing, it is attributed to human factors in the way the values are presented to users/clients and viewed by them. In most organizations, division of labor of data/information acquisition and their use are separated, which is labeled *indirect informing*. In such situations, users must be presented with data or information that is interpretable and understandable for them (legible; in their preferred language, measurement units, conventions, etc); it must be **presentation interpretable** by the *intended* clients/users. Then, it becomes another necessary *but not a universal* quality requirement. Presentation credibility can be measured by the *probability* that it will not be misinterpreted. Data may not be presented in the proper script, language, conventions, measurement units the user operates, particularly in global informing. A dose of medication in grams administered in ounces may be fatal, as it actually happened.

Deficiencies in **presentation credibility** of values impair their overall credibility. The maximum loss of credibility of any value due to impaired presentation quality is the complement to one of the joint product of the probabilities of reliability of its **definition, variability, objectivity, accuracy, precision, and currency**. They all are less than one. The qualities in bold are the indirect attributes of the second and subsequent orders, which first affect the presentation quality and next indirectly the actionable credibility.

The purpose of the model of diagnostic and functional dependencies of the actionable credibility on other indirect quality attributes is to explain the possible intricacies (see Figure 2). Some of the dependencies are even circular.

External Factors Affecting the Information Source's Reputation

The variety of potential sources must be examined with regard to their **external** and **intrinsic** factors of **reputation**. The *external reputation* of individual sources encompasses such *prerequisites* as **traceability** of individual values to their respective sources, **availability of communicable sources, alignment of attitudes, and alignment of interests** between active (providing) informing sources and the entities informed. Each of them may preclude further consideration of a source.

Traceability of values means they can be unambiguously attributed to specific sources. It is a prerequisite for examining their source-specific credibility (SSC) particularly important for repetitive operations. The more important, valuable, dangerous, litigation prone, subject of personal accountability and responsibility, and/or vulnerable to criminal activities a value is, the more the responsible managers or commanders preserve its traceability and transparently document its handling (audit trails). "Information systems are designed so that every financial transaction can be traced. In other words, an audit trail must exist that can establish where each transaction originated and how it was processed. Aside from financial audits, operational audits are used to evaluate the effectiveness and efficiency of information system operations" ("Information Systems," 2005)

Availability of communicable sources to entities informed is a necessary prerequisite for examining their reputation. Regardless of whether informing is passive or active, there may be no viable communication channels available, or they may be vulnerable to disruption, tapping, and/or interference. This may render some of the potential sources unsuitable or unreliable due to intermediary circumstances, which preclude their further consideration.

Particularly in active informing, the reputation of sources and the implied subsequent source-specific credibility may be severely affected by a variety of factors of biological, personal, psychological, sociological, economical, and political nature. All of them may become consciously and subconsciously reasons for disinformation and bias in communication—called here **align-**

ment of attitudes and interests. Both may objectively or subjectively, nevertheless effectively improve or impair the reputation of sources. They are of paramount overriding importance. They may even preclude some of them from becoming legitimate and admissible sources of information.

Alignment of attitudes frequently plays such a strong role that it may override even obvious conflicts of interest. Their intensity may vary from an irreconcilable armed life-and-death conflict through neutrality to a love-to-death affair. It may play a role at the personal, family, tribal, ethnic, religious, racial, and national levels, and, in the future, even at the interplanetary level. The resulting disinformation and bias may simply aim at distraction of the targeted entity informed, gaining advantage of or control over it.

With regard to **alignment of interests**,

- The very first question to be asked is what kind of alignment of interest exists between the source and the client, whether there is any open (declared), implied, or only a potentially adversary conflict of interest, even only an association of the source with entities that may have conflicting interests with the entity informed (client.)
- A similar question must be asked: whether there is any objective divergence or disparity of interests. The purpose of such a question is to assess the likelihood of disinformation or of intentional bias in the information provided.
- A milder question is whether there is any history of outright disinformation of general nature, such as fabrication of news or intentional bias in their presentation. It may come from greed, longing for vainglory, etc., as it can be easily observed in mass media, among politicians, performers, marketers, and journalists.

Intrinsic Factors Affecting Source's Reputation

The *intrinsic reputation* of sources depends on their **reliability** in yielding the value of interest with a **source-specific credibility**. Some of the factors improve it other impair it.

Improving Factors

Verifiability, replicability, and warranty increase the objective and perceived reliability of a source.

Verifiability takes various forms. One of them is **accreditation** of the source by a reputable professional body, which periodically verifies the source's eligibility for its accreditation status. Another form is a **periodical audit**, as it applies to public corporations, if the source is subject to a legally mandated audit. Similarly, it is when the source is **bonded** or **legally responsible** for the information service it provides.

Replicability may be another form of assurance in a credible manner of quality when the source enables a replication of the same test, for instance, by preserving the specimen for additional testing later when doubts may arise or if test results are challenged.

Warranty is even a stronger form of quality assurance. It is particularly convincing when combined with bonding. Usually it assures that seriously observed procedures are in place. Warranty offered indicates how much the client's risk is diminished in case of defects in information quality.

Impairing Factors

Reliability in yielding the value of interest with a source-specific credibility plays a significant role mainly in **routine operations**. Counter-intuitively, as never fully attainable, the reliability

factor can have only an impairing effect on reputation, even when very high. For instance, semi online stock quotes provided via Internet by MSN-NBC frequently blackout likely due to overload of requests.

Besides of reliability, Figure 2 shows many regularly observed impairments in: definition, variability, objectivity, accuracy, precision, and currency, which affect both the **reputation of data/information sources** and the **presentation credibility** of values. There occurs, however, a circular dependency. The source specific credibility depends on the source's reputation, but again the source's reputation hinges on credibility of the yielded values.

Uncertainties associated with the **definition** of what specific data values represent in the real world automatically decrease by this much the final credibility of the concerned value. For instance, gross national product (GNP) should be computed the same way by all countries, but it is not so. The same value in accounting might be different dependent on the selected accounting method.

Similar uncertainties result from **variability** of values within samples. For instance, the sales of a multinational company in US dollars may be subject to exchange-rate variations of as much as 1% per day. In statistics, the best measure of variability is the *standard deviation*.

Loss of **objectivity** (free from bias) may happen in the process of data/information acquisition due to the approaches and methods used in selecting the primary sources; measuring and observation points; measuring instruments; and, finally, when collecting, processing and presenting data. The resulting bias may be either unintended due to ignorance or introduced intentionally. The results of distortions may be significant, hence deceptive, and damaging to the source's reputation. To rectify the bias and compensate for it may require substantial additional resources. Whether their use is economically justified, can be assessed only when the materiality of objectivity is known. As in other cases, this will affect the direct secondary requirement of **economically actionable credibility** (Gackowski, 2005b). In reporting news and political statements, objectivity is in very short supply for it is always tainted by their purpose.

In all situations, one encounters some loss of **accuracy**, meant as free of errors, among them random errors. Inaccuracy of data or information values is one's complement of accuracy. Accuracy equals one minus inaccuracy. A typical gross measure of **inaccuracy** and the uncertainty related to it is the **error rate**. This equals the number of values in error divided by the total number of values gathered. A more useful measure of inaccuracy due to different kinds of errors is the **expected cost of dealing** with their consequences. It equals the product of the number of values, the probability or frequency of each type of error, and the average cost of dealing with each type of error. Such measure of **inaccuracy** provides the clients with a better idea of how serious are the consequences due to each type of error. One may reduce many of them by using check digits, error self-detection codes, error self-correcting codes, etc., which are here indirect quality attributes of the third order. Use of barcode readers considerably reduces many types of errors, except for completeness. Clients, users of information systems, even business systems analysts, need not be experts in using them, but they should be taught to recognize the need for such provisions. All statistic and measurement data should always be accompanied by the respective average error rates.

Insufficient **precision** in representation of reality directly compromises credibility of values. For numerical data, **precision** is measured by the **number of significant digits** used, and for pictures and images, by the **number of dots per inch** (dpi). This unit is commonly used to describe the precision of printers, computer screens, scanners, etc. These measures can be converted into corresponding *standard deviation*. There is a trap associated with accuracy and precision. Generally, these attributes are overrated (Wang, Reedy, & Kon, 1996). *Unchecked* efforts to increase the level of accuracy and/or precision of any value can become counterproductive. The ultimate de-

termination of the indispensable and economically justified level of any of them strongly depends on the materiality of the factor (Gackowski, 2005b). Classic examples in this respect, is insufficient precision of aerial photographs and over-reliance on them. It led, and continuously leads to serious overrating of damages inflicted by bombing.

Currency of data/information values means they are sufficiently up to date. Wang and Strong (1996) labeled it as timeliness and defined, as “the extent the age of the data is appropriate for the task at hand,” (p. 32). The label “timeliness” is used there in conflict with the terminology used by FASB (1983) and subsequently by CPAs for timely availability. This causes unnecessary confusion. The frequency of updates should be optimized. Insufficient frequency and too frequent updates are detrimental to cost effectiveness. If the volatility of the variable is known and **random**, the error, which can be converted to relative error, can be estimated. The standard deviation δ (**Sigma**) of the variable v measures the average error or deviation from the mean. The $-n*\delta < v < + n*\delta$ is the confidence interval. There is a probability associated with it that the variable v will not deviate more from its mean or fall outside of the interval. In total quality control (TQM) there are two well-known levels of confidence used to assure quality. 3δ with 99% confidence and 6δ used by Motorola with a 99.99966% confidence, which is only about 3 times per million, hence very unlikely. In other words with currency, one may be very confident that that the error or deviation from the mean will not exceed $\pm v*(6*\delta + a*t)$, where a - the slope of the current trend of the changing variable and t - the elapsed time since the last update of the variable. Nevertheless, significant abrupt changes to both patterns of behavior of the variable, which are represented by δ and a will ruin such an estimate.

Each factor *directly* and *independently* impairs the final level of credibility of the affected value, and *none of them compensates* for the losses of credibility caused by other factors. Maximum benefits from using information can be attained only at the *optimum levels* of definition, objectivity, accuracy, precision, and currency. Finding this optimum is not easy; it lies somewhere between the low and high levels. Whenever information technology professionals tempt clients with higher accuracy, precision, or currency than they had before, they should ask bluntly, “What will be the additional benefits and at what additional cost?” When one has no indication that their increased level leads to higher cost effectiveness, forget it. After the previous explanations, one may easily see, to the surprise of many, that the *examination of their economic level should be postponed nearly until the very end*, after all the necessary quality requirements have been examined and assessed in the first place.

Diagnostic Sequence of Examining Operation Quality Requirements

Usability of values must be tested for a *one-time* or a *repetitive* use. In the first case, future availability is of no interest. The client/user may still be interested in the one-time actionable credibility of the value. With anonymous threats, one does not know the source, but its credibility for preventive action is still of highest interest. For the sake of simplification, this inquiry focuses on repetitive operations. Then, **traceability** of data or information values to their sources is a universal prerequisite that enables assessment of their reputation and, subsequently, their source-specific credibility. Values instrumental in operations must undergo a systematic examination of their operation qualities including their actionable credibility (see Schema 1), which render values usable.

Schema 1. A simplified schema of examining unknown information for usability: the context of testing for operational availability and actionable credibility

GatherUsableInformation // This module tests unknown information values for planned tasks

```
While (unknownString ≠ endofString)
  V = Read (); // V – a variable, which takes on the unknownString;
  If (not V.Interpretable ()) continue; // Do the same. Stop further testing and get a string
  If (not V.Relevant ()) // Underlined names indicate separately defined modules
    Then continue; // Skip the rest as irrelevant
    Else taskID = v.getTaskID (); // Identify the relevant task, which needs the value of V
  If (not V.Material ()) continue; // Skip the rest if not of significant materiality
  If (not V.Available&ActionablyCredible ()) continue; // Skip the rest, acquire another
  string
  // See Figure 3 for details of testing for availability and credibility of an information
  value
  Task.setTaskID (V); // Value V is assigned to the identified task
  If (taskID.Complete ()) CompletedTasks.add (taskID); // The completed tasks added to
  list
EndWhile // End of loop. Its execution continues until endofString has been detected
EndGatherUsableInformation // end of procedure that shows the context of testing for
//operational availability and actionable credibility of a specific
// information value V, which from now on has become a usable data value of V.
```

Figure 1 shows a process chart, which illustrates steps for assessing **operational availability** and **actionable credibility** of information values that were declared as interpretable, relevant, and of significant materiality in operations. One should continue testing for effective task-specific operational completeness of data, which is a separate issue.

ER-Diagrams of Functional Dependency of Credibility

Figure 2 shows the diagnostic dependencies of credibility of any value on indirect quality requirements of the first, second, and subsequent order. It takes the form of an entity-relationship (ER) diagram. To facilitate its presentation, Figure 2 is divided into Part A (Figure 2a) and B (Figure 2b.)

Figure 2a presents an entity-relationship diagram, showing how the **joint credibility (JC)** of an information value depends on all the **indirect quality attributes** of the first order (**variety of sources, mapping quality, presentation quality, and source-specific credibility [SSC]**).

Figure 2b presents an entity-relationship diagram showing how the **reputation of a source** lends source-specific credibility to the yielded values. The source's reputation depends on:

1. **External** preconditions (**traceability, availability of communicable sources, alignment of attitudes and alignment of interests**) that enable or preclude its further examining, and some:
2. **Intrinsic** qualities, which:
 - a. Improve the source's reputation such as **verifiability, replicability, warranty,** and

- b. Impair its reputation by being always deficient (probability less than one) such as **reliability, definition, variability, objectivity, accuracy, precision, and currency** of the values yielded by the source.

The label **actionable credibility** represents the cluster of interdependent quality attributes related to credibility. Each of them is an indirect contributing factor of the first, second, or subsequent order. The required level of actionable credibility differs from task to task. An extreme example illustrates the extraordinary precautions taken before surgical procedures. One must assure a very high credibility of the patient's identity and location of the specific organ to be operated on. At least three independent tests are performed before a surgical procedure is conducted. Hence, in operations, there is nothing intrinsic to the data and nothing non-contextual in accuracy and other indirect quality requirements.

Diagnostic versus Functional Dependencies

There are at least two basic types of dependencies, which should be examined in any field of investigation: **diagnostic** and **functional** dependencies.

Diagnostic dependencies of factors pertain to the sequence of their examination. Two major criteria are used: **prerequisite** and a subordinate criterion **ease of examination**.

- The **prerequisite criterion** is relatively easy to follow when testing factors pair by pair.
- **Ease of examination**, however, depends on some prior experience with factors, which are not prerequisite to each other, but are at least perceived of different level of difficulties, which one faces when examining them.

Functional dependencies of factors reveal the immediacy of affecting something, whether directly or indirectly via an intermediary chain of other functional dependencies of the second and subsequent orders.

Figure 3 shows both the functional dependencies of the direct primary universal quality requirement of **actionable credibility** on indirect qualities, which affect it and the suggested diagnostic sequence of examining them; whether **actionable credibility** (its percentile) determined by the policy or doctrine in effect, has been attained.

The chart or diagram (Figure 3) shows **two levels** of functional dependency of availability and actionable credibility and **20 steps** in the diagnostic sequence of their examination marked by the call numbers. This sequence also illustrates the control flow of determining whether actionable credibility has been attained. In case, there are any preset minimal acceptable levels of quality of the indirect factors and a distribution of how frequently they are not met, one may modify the diagnostic sequence so that the most frequently factors are checked first for compliance. This may improve the economy of the diagnostic sequence by sooner quitting the diagnostic sequence, when something necessary could not be met.

The latter deficiencies form a chain of flow-down pervasiveness of quality impairments. If, at any step, a quality requirement cannot be met, testing for requirement compliance further flow-down is redundant. Examination should be aborted. An impairment of a quality requirement incurred at any prerequisite step cannot be compensated at any flow-down step. In reverse order, any quality impairment even at the very end of the chain affects directly (without any intermediaries) the source-specific credibility (SSC) of the concerned value. For instance,

- Ambiguous **definition** of values introduces uncertainties that cannot be repaired by any of the remaining quality attributes within the diagnostic sequence

- High variability cannot be improved by better OBJECTIVITY and other aspects listed further flow-down
- Suspect OBJECTIVITY cannot be repaired by ACCURACY, **precision**, or **currency**. Hence, source-specific credibility of yielded values suffers
- Low ACCURACY (burdened with errors) cannot be improved by higher **precision**. Again, source-specific credibility of presented values suffers
- Insufficient task-required **precision** (for instance, low resolution of a picture) may render it useless, and excellent **currency** will not help either. Source-specific credibility suffers again
- Insufficiently frequent updates make data values questionable with regard to their **currency**, and then, even satisfactory OBJECTIVITY, ACCURACY, and **precision** is wasted. Source-specific credibility of yielded values will suffer again.

Economy requires a proper diagnostic **sequence** of their **examination**. The numbers associated with names of the indirect quality attributes indicate the sequence. The quality attributes are entities in the ER-diagrams in Figure 2, and the corresponding modules for testing them in Figure 3.

The Operation Model of Credibility versus the Conceptual Framework of DQ

Wang and Strong (1996) conducted an empirical survey-based study of how consumers view data quality and empirically developed a Conceptual Framework of Data Quality—Table 1. There, BELIEVABILITY (ranked 1 in 20) plays a prominent role as an intrinsic (emphasis added) attribute of data quality.

Table 1 Information Quality Categories and Dimensions (Source: Wang and Strong, 1996)

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

Quote: “*Intrinsic DQ includes not only accuracy and objectivity, which are evident to IS professionals, but also believability and reputation. This suggests that, contrary to the traditional development view, data consumers also view believability and reputation as an integral part of intrinsic DQ; accuracy and objectivity alone are not sufficient for data to be considered of high quality. This is analogous to some aspects of product quality. In the product quality area, dimensions of quality emphasized by consumers are broader than those emphasized by product manufacturers. Similarly, intrinsic DQ encompasses more than the accuracy and objectivity dimensions that IS professionals strive to deliver. That finding implies that IS professionals should also ensure the believability and reputation of data. Research on data source tagging is a step in this direction*” (Wang and Strong, 1996, p. 16).

Comment 1: Within the realm of operations, there is no room for intrinsic data qualities. **Quality** of something is defined here by the states of its attributes that qualify it to play a significant role in operations. By the principle of relativity of operation quality, all qualities are contextual. The required levels of accuracy or objectivity are task-specific, hence contextual. Maybe it is not evident to IS professionals, but it is clear to anyone managing or commanding operations. The idea of quality attributes intrinsic to data values should be abandoned. Actually, they were mislabeled. The researchers clearly derived them from the deficiencies in system design and operations; hence, they are intrinsic to them. If the data or information values are shared by different applications, the required levels of any quality dimension should be **minimaxed**.

Comment 2: Credibility (BELIEVABILITY) is rightly of high concern to data users. It is bothersome, however, when obvious interdependencies between objectivity, accuracy, reputation, traceability, timeliness (meant as currency), and believability are ignored. They are discussed as independent quality dimensions when they are only contributing factors, as evident in the operational model of credibility.

Comment 3: In the light of this inquiry, all of the data-quality attributes grouped under the label “*INTRINSIC DATA QUALITY*” melt down to a single highly contextual (*not intrinsic to the values*) direct primary universal quality requirement of **credibility**. It is rarely-to-never fully attainable with a long list of not less than **twenty** contributing factors (indirect quality requirements) as shown in Figure 3.

Comment 4: Wang and Strong (1996) postulate that “*IS professionals should ensure the reputation of data. Research on data source tagging is a step in this direction*” (p.16). From the perspective of operation quality, IS professionals are neither competent nor possess means to ensure the reputation of data. The maximum one can expect from them is to ensure that the data entrusted by users adhere to the agreed format and standards, and that they will not be corrupted during storing, processing, and retrieving, and that they will adhere to the agreed presentation standards. In the Conceptual Framework of Data Quality, Wang and Strong, however, for trivial statistical reasons deliberately dropped TRACEABILITY despite its exceptional importance.

Similarly, COST-EFFECTIVENESS of data values was dropped. One can notice here the glaring poverty of empiricism in scientific research, how empirical studies conducted without a sufficient qualitative model can lose even the most important and pervasive quality requirements. Statistics is useful with uncertain quantities but weak in discovering qualitative dependencies, whether diagnostic or functional.

The empirically derived Conceptual Framework of Data Quality uses classification criteria that are not disjoint (no overlapping), clear, and complete. All of the listed quality dimensions are contextual; hence, this class contains them all. Liu and Chie (2002) raised similar objections to the Conceptual Framework as follows: “... believability may *inflate* (emphasis added) DQ by double counting dimensions such as accuracy, objectivity, and reputation. In addition, attributes like integrity, credibility, or reputation overlap in meaning with each other and other attributes such as accuracy and objectivity” (p. 303).

Hence, the cited research and its results are of exploratory nature; however, they may serve well for situation-specific improvement of quality. These improvements are well-known and fully acknowledged; nevertheless, they do not yield results of a more lasting validity. Empirical studies serve well in exploring users’ preferences, and/or confirming or rejecting hypotheses derived from a critical qualitative examination of the subject.

Conclusions

The presented paper and the model of the qualitative diagnostic and functional dependencies of joint **actionable credibility** of data and information values:

- Explains the terminological advantage of the label **credibility** over **reliability** or **believability** of data/information values
- Identifies most of the most indirect qualities, which directly and indirectly affect the joint credibility of information values. (It may facilitate assessment, whether **actionable credibility** or **economical actionable credibility** has been attained.)
- Defines the **joint credibility of information** values as a direct function of indirect quality attributes of the first order, such as variety of sources, mapping quality of the delivery system, the presentation credibility of these values, if indirect informing is the mode of operations, and finally the source-specific credibility (Figure 2a)
- Defines the **reputation of data/information sources** as a function of:
 - Many indirect qualities of external nature (**traceability, availability of communicable sources, alignment of attitudes and alignment of interest** between the source and the clients), and
 - Many qualities intrinsic to the source such as **verifiability, replicability** of values, and **warranty** that *improve* the reputation of the source, and other, which *impair* the source's overall reputation (**reliability, definition, variability, objectivity, accuracy, precision, and currency**)
 - Suggests an economic diagnostic sequence of their examination.

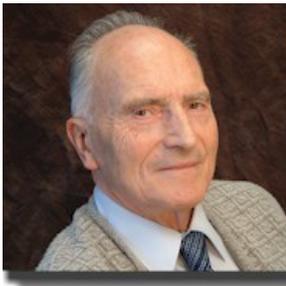
The paper present also a cursory comparison of the **operation quality of information** with the **Conceptual Framework of Data Quality** empirically derived by Wang and Strong (1996) with regard to BELIEVABILITY or credibility in particular. The latter model turns out to be inconsistent, incomplete, lacks clear, disjoint, exhaustive classification criteria, and misses major quality attributes.

References

- Financial Accounting Standard Board [FASB]. (1983). *Accounting standard: Statement of financial accounting concepts*. New York, NY: McGraw-Hill.
- Gackowski, Z. J. (2004). Logical interdependence of data/information quality dimensions—A purpose-focused view on IQ, *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, Cambridge, Massachusetts. Retrieved from http://www.iqconference.org/Documents/IQ_Conference_2004/Papers/LogicalInterdependence.pdf
- Gackowski, Z. J. (2005a). Informing systems in business environment: A purpose-focused view. *Informing Science Journal*, 8, 101–122. Retrieved June 27, 2005, from <http://inform.nu/Articles/Vol8/v8p101-122Gack.pdf>
- Gackowski, Z. J. (2005b). Operation quality of data and information: Teleological operations research-based approach, call for discussion. *Proceedings of the 10th Anniversary International Conference on Information Quality—ICIQ-05*, Massachusetts Institute of Technology (MIT), Cambridge, MA
- Gelman, I. A. (2005). A theory of complementarity of extracting accurate data from inaccurate sources through integration. *Proceedings of the 10th International Conference on Information Quality (ICIQ-05)*, Cambridge, Massachusetts.
- Gleim, I. N. (2004). *CPA Review Financial* (12th ed.). Gleim Publications.

- Huang, K., Lee, Y. W., & Wang, R. Y. (1999). *Quality information and knowledge*. Upper Saddle River, NJ: Prentice Hall (Quality Programs and Initiatives at MIT–MITIQ Program.)
- Information system. (2005). *Encyclopedia Britannica*. Retrieved April 22, 2005, from Encyclopedia Britannica Online <http://search.eb.com/eb/article?tocId=218074>
- Lee, Y., Strong, D., Kahn, B., & Wang, R. (2002). AIMQ: A methodology for information quality assessment. *Information & Management*, 40(2), 133–146. Retrieved from <http://mitiq.mit.edu/Documents/Publications/TDQMpub/2002/AIMQ.pdf>
- Liu, L., & Chi, L. N. (2002). Evolutional data quality: A theory-specific view. *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)*, Cambridge, Massachusetts, 292–304. Retrieved, from <http://www.iqconference.org/iciq/iqdownload.aspx?ICIQYear=2002&File=EvolutionalDataQualityATheorySpecificView.pdf>
- Wang, Y., & Wang R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86–95. Retrieved, from <https://www.crg2.com/iqconference/documents/publications/TDQMpub/WandWangCACMNov96.pdf>
- Wang, R. Y., Reddy, M. P., & Kon, H. B. (1995). Toward quality data: An attribute-based approach, *Journal of Decision Support Systems*, 13, 349–372.
- 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. Retrieved, from http://jmis.bentley.edu/articles/v12_n4_p5/index.html

Biography



Zbigniew J. Gackowski has extensive experience in industry, public administration, and universities. His teaching and research bridge the gap between Central European and US experience in Computer Information Systems (Warsaw Polytechnic, the University of Michigan [*Fulbright Research Scholarship*], Purdue University [Visiting Associate Professor], Baruch College [Visiting Professor], California State University, Stanislaus [Professor], and the University of Melbourne [Visiting Professor]). His research has received much recognition. While in Poland, he published more than 120 items, among them, four books, and six papers in refereed journals, 12 papers presented across Europe, the United States, the Middle East, and South America. While in the USA, he taught at three universities and presented 31 research papers published in the proceedings of ASIS, ACM, DSI, *Information Systems Educators Conference*, *The Informing Science Institute*, and the *International Conferences on Information Quality* at MIT. He is a member of ACM and DSI and is a charter member of the *Association for Information Systems* and a founding member of the *Institute of Informing Science*. His recent research focuses on quality of data and information values in informing.