Quality of Informing: Credibility Provisional Model of Functional Dependencies

(Research in Progress)

Zbigniew J. Gackowski
California State University, Stanislaus

ZGackowki@.csustan.edu

Abstract

This is an inquiry into the "actionable credibility" of information values in informing and how it depends on other attributes of data and 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 these attributes as independent quality factors. This paper identifies and describes the ignored dependencies. 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 operations quality requirements of data and information values as proposed by Gackowski (2004, 2005a, 2005b). The results are presented for *challenge, critique,* and *discussion*.

Keywords: Operations quality of data/information, actionable credibility of informing

Introduction

Within the subjects of informing science, "actionable credibility" is the most complex universal direct primary quality requirement of data or 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. It ignores, however, the multidimensional aspects of credibility and its complex functional logical and/or mathematical dependencies on the *indirect* quality attributes. The latter are treated on equal footing with the *direct primary* and *direct secondary* quality requirements, as defined in what is hoped to be the impact-determined universal taxonomy of operations quality requirements (Gackowski, 2005b).

This paper identifies and describes these dependencies. It also demonstrates how empirical studies that ignore them produce unreliable results. To attain more than situation-specific practical improvements in quality, a shift is needed toward first identifying at least the qualitative dependencies. This inquiry is a step in that direction. Before embarking on quantitative studies, a provisional qualitative model of functional dependencies of credibility on other indirect quality attributes of data and information is presented for *discussion, critique,* and *challenge*.

Other contributions of this paper are as follows:

- A discussion of the terminological alternatives to actionable credibility of data/information (D/I) values such as reliability and believability

- An identification of (one hopes) most of the factors that affect credibility in informing

- Qualitative definition of credibility of D/I values as a direct function of indirect quality attributes of the first order, such as variability of independent D/I sources, their source-specific credibility, mapping quality, and their presentation credibility to users. The latter are also functions of other indirect quality attributes of the second and subsequent orders.

- A mapping of these identified dependencies of credibility in an entity-relationship (ER) diagram

- Identification of the logical sequence of examining the quality attributes 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

Operations 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 acquired. Of course, once acquired and accepted, they become data values for the 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. In routine operations, credibility of data values is not less important but is of less concern to individual internal data/information client/users. In indirect informing (Gackowski, 2005a), data/information values are subject to established procedures that assure their integrity during acquisition, entry, storing, retrieving, and presentation. It is done on behalf of all the data-sharing clients/users.

Credibility of data/information values is a complex notion of the degree or extent to which they represent the actual reality to users of those values in purposive operations. 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 a data/information value 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 in reality. It is the weakest term. In scholarly and technical writings, reliability is associated with reliability of systems by means of a complex statistical theory. Data and information values are symbolic representations of reality. Hence, **credibility** seems to be the most appropriate 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). Reliability of systems is a separate non-trivial subject.

How is credibility 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 quality attributes? Wang and Strong (1996) empirically studied how consumers view data quality. They found BELIEVABILITY to be the most important quality attrib-

ute (1 in 10 on their scale of importance and 1 in 20 within their "Conceptual Framework of Data Quality." Respondents were provided with a ranking scale of 1–10 but with no criteria of importance [in what respect important?]). This framework was cited in many publications, among them, Huang, Lee, and Wang (1999) and Lee, Strong, Kahn, and Wang (2002). From the operations viewpoint of quality (Gackowski, 2005b), the results of ranking by importance of BELIEVABILITY (1 in 10) and RELEVANCE (3 in 10) contradict each other. When importance of a D/I value is measured by its impact on operations, by the law of pervasiveness, other attributes of quality of the same value must be of the same importance and cannot be ranked differently. Obviously, at the time of the study, this principle was not recognized. Nevertheless, the broadly held explicitly or implied false assumption of independence of quality attributes lives on.

BELIEVABILITY was ranked on equal footing with all of its contributing factors, such as accuracy, objectivity, timeliness (meant as currency), traceability, reputation, variety of data and data sources, etc. Wang and Strong identified 179 data quality attributes, later reduced them to 118 for subsequent factor analysis, and finally compressed them into 20 relatively well-defined data quality attributes. 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 tentatively universal quality attribute, hence of highest importance, but as the fourth one in the logical sequence of their examination. The functional dependencies of credibility of data and information values on indirect quality attributes are the main subject of this inquiry. Its purpose is to develop a provisional analytical model and a mapping of at least most of these interdependencies. The long-term goal is to develop an algorithm and an intelligent information quality analyzer.

Provisional Model of Credibility as Function of other Attributes

The provisional model of credibility addresses the *place* of credibility among other quality requirements, the *general notion* of credibility of data/information values, the factors *directly and indirectly* affecting credibility of data/information values, and the *examination sequence* of the factors contributing to credibility. In addition, there is a

- schematic presentation of the *examination process* of credibility (Process chart – Figure 1)

- schematic entity-relationship diagram of the identified *functional dependencies* of credibility on other identified indirect attributes of data/information quality of the *first* order (Figure 2b and of the *second and subsequent* orders – Figure 2a)

Credibility among other Operations Quality Requirements and its Role

Presently, the theoretical model of operations quality of data/information values (DQ/IQ) (Gackowski, 2005b) comprises 14 essential components:

- The impact-determined universal taxonomy of all identified and not-yet-identified requirements of operations quality of data/information values

- The definitions and descriptions of the five universal operations quality requirements of data/information values

- An economic sequence of examination of the universal quality requirements
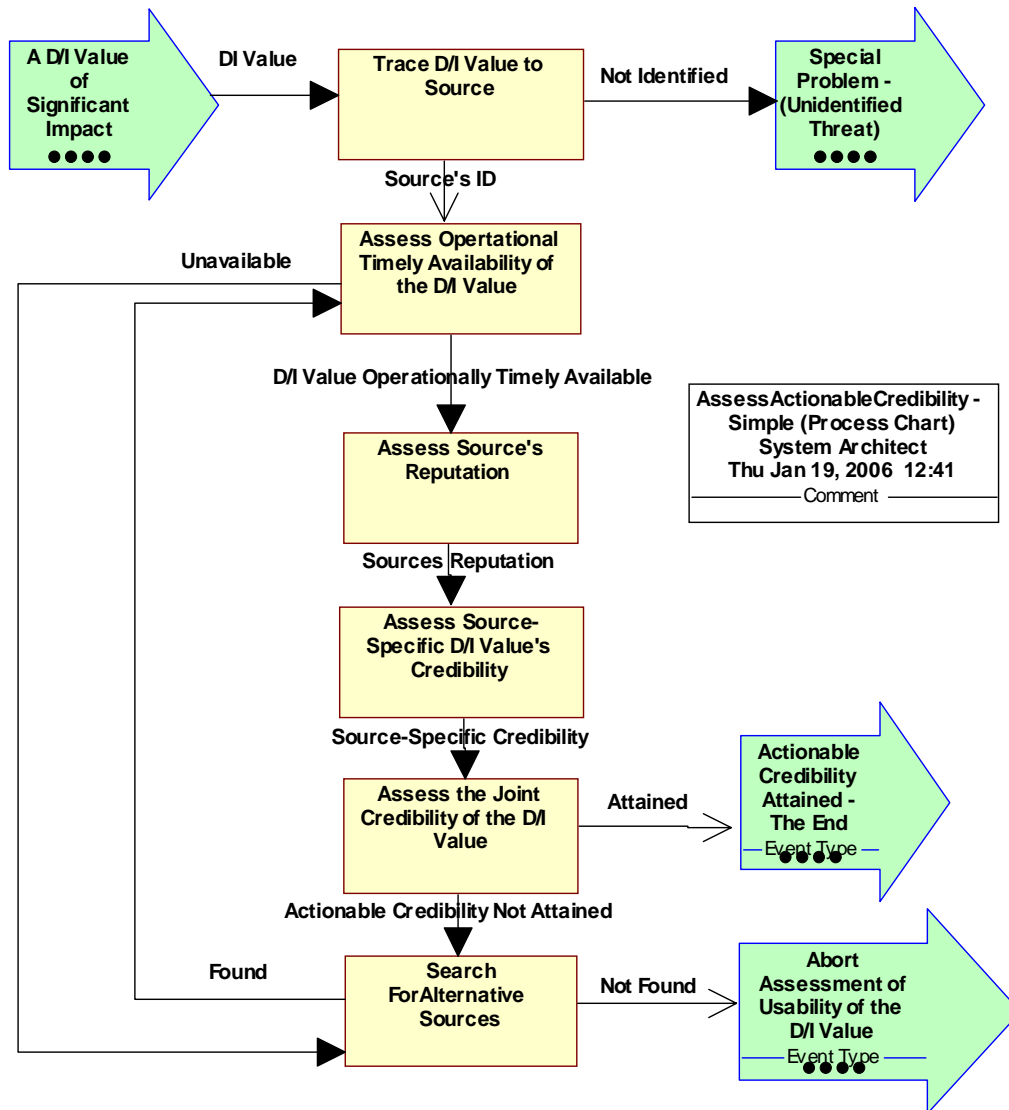
Figure 1. Examine Credibility of a Data/Information Value - Process Chart

- The definitions of the seven universal principles, to which all data/information operations quality requirements are subject

Actionable credibility is defined as the degree of credibility of a data/information value at which the user or decision-maker is willing to take action. Subsequently, within the set of the identified principles,

1. the principle of **relativity** of all operations quality requirements of values of all symbolic representations (data, information, knowledge) *implies* that *actionable credibility* is also determined by the *purpose and circumstances of operations* they are used for.

2. the principle of **pervasiveness** of the significant impact that a data or information value exerts on operations *determines* also the *importance of actionable credibility* in relation to all other operations quality requirements associated with the same value.

3. within the **universal taxonomy** actionable credibility is classed as a *direct primary universal quality requirement* for any task-specific data/information value.
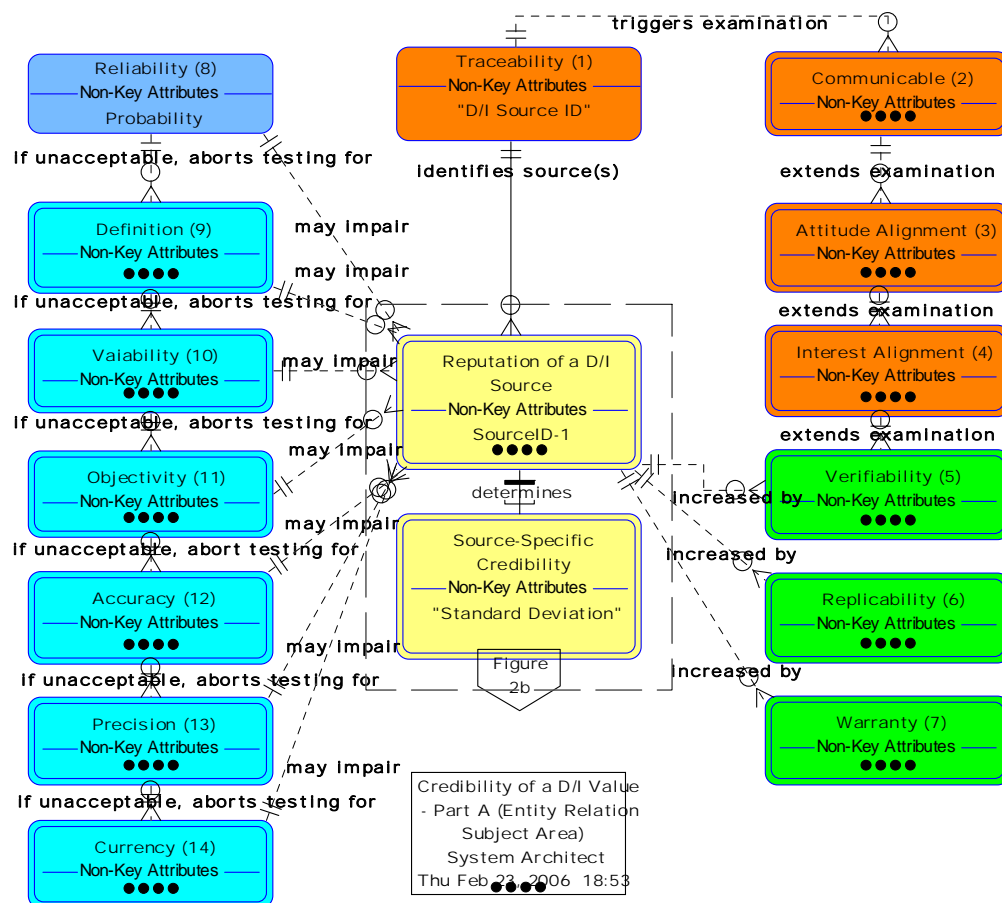
**Figure 2a. Reputation of a D/I Source -
Functional Dependencies on Required Indirect Quality Attributes
(Entitiy Relationship Diagram - A Provisional Qualitative Model - Part A)**

4. the principle of linear logical precedence of direct primary quality requirements, which determines the **economic sequence of examination**, where actionable credibility is the _fourth universal quality requirement_ of any single task-specific value tested.

5. the principle of **equivalence** of a lost data value and an unobtainable information value implies that the importance of _actionable credibility applies equally_ to data and information values of equivalent impact on operations.

6. the principle of the **operational usability** of values _implies_ that actionable credibility is one of the four universal mandatory quality requirements that must be jointly met to make any data/information value operationally usable. Usability increases **effective operational completeness** of task-specific clusters of such values and thus makes them **operationally useful.** (An activity-specific cluster of indispensable operationally usable values to be effectively operationally complete must consist of one or more task-specific benefit-adding values and all its/their mandatory usable companion values and attain effectiveness sufficient to trigger a transition of the operations state [to act or not].)

7. the principle of **degradation** of decision situations with the decline of quality of a usable data or information value _implies_ what follows: If actionable credibility
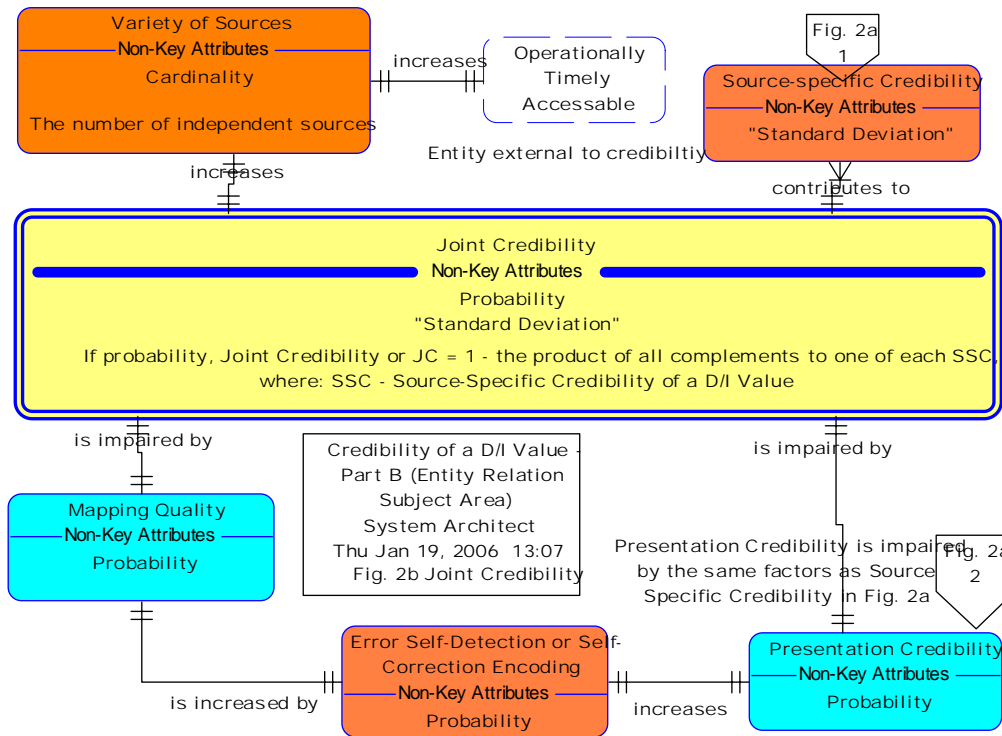
**Figure 2b. Joint Credibility as a Function of Required Indirect Quality Attributes
(Entity Ralationship Diagram - A Provisional Qualitative Model - Part B)**

- is *certain,* jointly with other mandatory quality requirements, decision-makers deal with a **deterministic** situation in the affected area.

- 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 whether they act or not (threats ignored).

General Notion of Credibility of Data/Information Values in Operations

Credibility of a data/information value means whether it is true—whether one may rely on the value. The adjective *true* means consistent with reality. Complete credibility is rarely to never attainable. If it is Boolean (true, false), its degree is measured by the probability of its veracity (0–1). While probing for veracity of data or information values, users/clients face dramatic options, less with data of well-established roles and more with unknown recently acquired information values. The 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 <u>intentionally</u> misinforms. On the one hand, it may not be clear who the originator is due to omission of contact addresses, when it was originated or updated, what methods of collection or acquisition were used, etc. On the other hand, all the above may be available. Now, however, the client faces two extremes of **deception** with many possibilities between the two. All

indicators of validity are given to legitimize the value, but one or more of them are false, or they all are true to convince and <u>trap</u> the gullible in a desired manner.

Misinformation <u>unintentionally</u> misrepresents reality. It may be distorted at its acquisition, communication, storing, processing, presentation, and interpretation. It may be of lesser or higher materiality, as required and defined by FASB (1983). In operations, it is determined by the impact a value exerts on operations.

Valid information <u>faithfully</u> represents reality. Usually, validity is a function of the reputation of its source when it can be traced back to a responsible originator, where the level of responsibility is adequate to the potential consequences (in operations) of the expected errors, and when the information value is of proven authorship, can be replicated, or confirmed otherwise.

Factors <u>Directly</u> Affecting Credibility

Affecting factors may be classed into those that are positive and those that are negative. It seems that only the **variety** or the **number of independent sources** increases the degree of credibility of data/information values. It implies that whether within a society or only within a private domain, main efforts should be focused on creating and maintaining a viable variety of competitive sources of information. One should always search for alternative or competitive data/information sources to increase the **variety of independent sources** and subsequently increase not one but even two direct primary mandatory quality requirements of any D/I value: **operational timely availability** and its **joint credibility**. Lack of access to alternative sources of D/I values and corresponding channels of communication makes anyone and any organization extremely vulnerable to bias, disinformation, and disruption of any communication. Because credibility is rarely to never fully attainable, in cases of grave consequences, one should try *not to rely* on a single source.

Sources of data or information values may be observers, sensors, processes, tests, etc. They may remain unknown despite all efforts to identify them and due to difficulties locating alternative sources (perfectly anonymous threats, threats under false identity, or only rumors of threats). These are special situations.

Any data/information value acquired from a specific source <u>inherits</u> credibility labeled **source-specific credibility (SSC)** from the reputation of its source. If the value is Boolean (true, false), the uncertainty related to source-specific credibility or lack thereof is its complement to one (1 – SSC). With additional sources, the uncertainty about a data/information value declines fast. If many sources generate the same or similar value, the **joint credibility (JC)** becomes the complement to one of the products of all the source-specific uncertainties according to the following formula:

$$JC = \Pi\ (1 - SSC_j) - \text{over all \textbf{independent sources j}} \qquad (1)$$

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

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

Actionable credibility of a data or information value is attained when joint credibility (JC) equals or exceeds the threshold for a state transition of operations (to act or not) as defined by the policy in effect. Economical actionable credibility imposes additional economical requirements, as de-

scribed by Gackowski (2005b). For instance, the cost of attaining it should not exceed the benefit the value adds to operations. When the level of significant impact 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. Then the level of actionable credibility may be set low.

Mapping quality (not the label but the notion) of the real-world states into the states of information systems (in informing science, the data/information delivery system) was defined by Wand and Wang (1996, p. 92) as a function of four (intrinsic to the design 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. However, they were mislabeled as "intrinsic data quality dimensions." By the law of relativity, no *operations quality requirements can be intrinsic* to data or information values per se. They may, however, be intrinsic to something else; e.g., to system design and operations. Mapping quality can be measured by the *probability* of being free from mapping errors.

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. It becomes another mandatory *but not a universal* quality requirement. Presentation credibility can be measured by the *probability* of not being misinterpreted.

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 sum of the probabilities of imperfections in **definition, variability, objectivity, accuracy, precision,** and **currency**. Usually the loss is smaller. It is unlikely that all of these problems occur at the same time. Most of them are rarely to never attainable, hence mostly less than 100%. The highlighted quality attributes are indirect attributes of the second order that will be discussed next. Some of them are affected again by other indirect quality attributes. Thus, one deals here with functional dependencies of the second and subsequent orders.

This overview covers only the direct functional dependencies of the **joint credibility (JC)** of data or information values. This is a simplification because most of its direct factors that are indirect DQ/IQ attributes of the first order may again functionally depend on a plethora of other indirect attributes of the second and subsequent orders. The purpose of the provisional model of functional dependencies of actionable credibility on other indirect quality attributes is to explain the possible intricacies. Figure 2b shows most of them. Some of them are even circular.

Factors <u>Indirectly</u> Affecting Credibility

The entire scope of the variety of the potential sources of data/information values must be examined with regard to their **external** and **intrinsic reputation**. The *external reputation* of individual sources encompasses **traceability** of individual data or information values to their respective sources, the **availability of reliable communication channels—communicable sources**, the **alignment of attitudes,** and the **alignment of interests** between active (providing) informing sources and the entities informed.

Traceability of data/information values means they can be unambiguously attributed to specific sources. **Traceability** of D/I values to specific sources is a **mandatory** prerequisite for examining credibility of their sources, which is particularly important for repetitive operations. The more important, valuable, dangerous, litigation prone, subject of personal accountability and responsi-

bility, and/or vulnerable to criminal activities a value is, 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 systems operations" ("Information Systems," 2005)

Similarly, **availability and reliability of communication** channels between the **communicable** source and the entity informed is a mandatory prerequisite for further examination of 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 and/or interference. This may render some of the potential sources unsuitable or unreliable due to intermediary circumstances, which precludes their further consideration.

In active informing particularly, the reputation of D/I sources and the implied subsequent source-specific credibility of the values provided 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 all types of communications—called here **alignment of attitudes and interests**. Both may decrease or increase the reputation of any source. They are of paramount overriding importance and determine the source-client specific bias and disinformation. 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 overrides 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 or at gaining a partial or even absolute control over the entities informed.

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 conflicts of interest, even only an association of the source with entities that may have conflicting interests with the 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 asses the likelihood of disinformation or intentional bias in the information values provided.

- A milder version of the same types of questions is whether there is any history of outright disinformation of general nature, such as fabrication of new or intentional bias in their presentation that does not necessarily target a specific client. It may come from greed, longing for vain glory, etc., as it can be easily observed in all mass media.

The *intrinsic reputation* of data/information sources depends on their **reliability** in yielding the value of interest with a **source-specific credibility.** The latter depends on the offered **verifiability**, **replicability**, or **warranty** that <u>increases</u> the source's reputation, while any imperfections in **definition, variability, objectivity, accuracy, precision,** and **currency** decrease it. The six imperfections affect the same way the **presentation credibility** in indirect informing; however, only **definitions** and **objectivity** are subject to bias and are less likely to outright disinformation. It is a circular dependency. The **source-specific credibility** of yielded values depends on the source's **reputation**, but, again, the source's **reputation** hinges on the **credibility** of the yielded values.

Reliability in yielding the data/information value of interest with a source-specific credibility plays a significant role only in **routine operations**.

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 some tests. For instance, this could be achieved by preserving the specimen for additional testing later when doubts may arise or if tests are challenged.

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

Uncertainties associated with the **definition** of what specific data represent in the real world automatically decrease to the same degree the final credibility of the concerned value.

Factors Impairing Reputation of Sources and Presentation Credibility

Figure 2 shows that impairments in definition, variability, objectivity, accuracy, precision, and currency affect both the reputation of data/information sources and the presentation credibility of the data or information values provided; hence, they are discussed separately. 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.

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 engagement of substantial additional resources. Whether it is economically justified can be assessed only when its impact on operations is known. As in other cases, this will affect the direct secondary requirement of **economically actionable credibility** (Gackowski, 2005b).

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 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., 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.

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. Any of these measures can be converted into corresponding *standard deviation.* There is a trap associated with accuracy and precision. Generally, these attributes are overrated (Wang et al., 1996). *Unchecked* efforts to increase the level of accuracy and/or precision of any data or information value can become counterproductive. The ultimate determination of the indispensable and economically justified level of any of them strongly depends on the impact the affected value makes on operations (Gackowski, 2005b).

Currency of data/information values means they are sufficiently up to date. It directly affects the final level of credibility. Wang and Strong (1996) labeled it as timeliness and defined it 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 or too frequent updates are detrimental to cost effectiveness. If the volatility of the value is known, its *standard deviation* subtracted from one measures the degree of currency.

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-operations benefits from using data/information values can be attained only at *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 mandatory quality requirements have been examined and assessed.

Examination Sequence of Operations Quality Requirements

Values instrumental in operations must undergo a step-by-step examination of their quality:

1. Signals received declared as **acquisition-interpretable** values,

2. D/I values tested for **relevancy** within task-specific operations,

3. D/I values assessed to determine whether their use **significantly impacts** operations,

4. D/I values checked for **traceability** to their sources,

5. Source checked for **operational timely availability** of values,

6. Assessment and ranking of the identified alternative sources with regard to their capacity to make the value of interest **operationally timely available** and of proper **reputation,**

7. **Actionable credibility** of values tested for attainability;

8. **If not, s**earch for alternative or competitive D/I sources to increase their variety, and, finally,

9. Assessment of **effective operational completeness** of usable values **that triggers state transition of operations** (to act or not to act) has been **attained**.

Figure 1—a process chart illustrates steps 4–8. These steps pertain only to the assessment of **actionable credibility** of data/information values that were declared as interpretable and relevant—in operations "of significant impact." Repetition of steps 1–8 for all members of a task-specific set of usable data/information values enables, in step 9, testing for effective operational completeness.

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 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 mandatory prerequisite that enables assessment of their reputation and, subsequently, their source-specific credibility.

ER-Diagrams of Functional Dependency of Credibility

Figure 2 illustrates the functional dependency of credibility of any data or information value on indirect quality requirements of the first, second, and subsequent order. It takes the form of an entity-relationship (ER) diagram divided into Part A—Figure 2a and Part B—Figure 2b.

Figure 2b represents a provisional entity-relationship diagram, showing how the **joint credibility** (JS) of a data/information value depends on (one hopes) all the **indirect quality attributes** of the first order (**variety of sources, source-specific credibility [SSC], mapping quality,** and **presentation quality)**.

Figure 2a represents a provisional entity-relationship diagram showing how the **reputation of a D/I source**, which lends source-specific credibility to the yielded values, depends on a multitude of external preconditions (**traceability, availability of reliable communication channels - communicable** sources, **alignment of attitudes and alignment of interests**) that enable or preclude its further examination and the more intrinsic ones that increase its reputation (**reliability, verifiability, replicability, warranty**) or impair it due to deficiencies in **definition, variability, objectivity, accuracy, precision,** and **currency**. The latter deficiencies form hierarchies of linear downward directed pervasiveness of quality impairments. If, at any level of their hierarchy, a quality requirement cannot be met, testing for quality-requirement compliance at lower levels is redundant. Further examination should be aborted. An impairment of a quality requirement incurred at a higher level of the hierarchy cannot be compensated at any lower level. In reverse order, any quality impairment at any lower level of the hierarchy affects upward directly the credibility of the concerned value. For instance,

1. Ambiguous **definition** of D/I values introduces uncertainties that cannot be repaired by any of the remaining quality attributes within the hierarchy.

2. High variability cannot be improved by better OBJECTIVITY and other aspects listed below but can be further compromised by any of them.

3. Suspect OBJECTIVITY cannot be repaired by ACCURACY, **precision**, or **currency**. Hence, REPUTATION and the credibility of yielded values will suffer.

4. Low ACCURACY (burdened with errors) cannot be improved by higher **precision**. Again, REPUTATION or credibility of presented values will suffer.

5. Insufficient task-required **precision** (for instance, low resolution of a picture) may render it useless, and excellent **currency** will not help either. REPUTATION or credibility of presented values will suffer again.

6. Insufficiently frequent updates make data values questionable with regard to their **currency,** and then even satisfactory OBJECTIVITY, ACCURACY, and **precision** is wasted. REPUTATION or credibility of presented D/I values will suffer again.

Economy requires a proper **sequence** of their **examination**, which is indicated by the sequential numbers associated with the names of the corresponding entities.

The label **actionable credibility** seems to represent well 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 might differ 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 about accuracy and any other operational quality requirement.

The Provisional 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 developed a Conceptual Framework of Data Quality—Table 1. There, BELIEVABILITY ranked 1 in 20 plays a prominent role as an intrinsic (?) 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"* (p. 16).

Comment 1: Within the realm of purposive operations, there is no room for intrinsic data quality. By the law of relativity of operations quality, defined as "fit for use," *non-contextual* aspects of quality do not exist. The task-specific required levels of accuracy or objectivity are 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, for the researchers clearly derived them from the viewpoint of system *de-*

sign and *operations*; hence, they are intrinsic to the latter. If the data/information values are shared by different applications, the required levels of any quality dimension should be **mini-maxed**.

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 one can see in the provisional 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 <u>contextual</u> (*not intrinsic*) direct primary universal quality requirement of credibility. It is rarely to never fully attainable with a long list of contributing factors.

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). In their 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/information values was dropped. In empirical studies conducted without a sufficient analysis of the functional interdependencies of quality requirements, even the important and pervasive ones may be lost. Statistics are acceptable and indispensable with <u>uncertain</u> quantities but are very weak in discovering <u>qualitative</u> interdependencies.

The empirically derived Conceptual Framework of Data Quality uses classification criteria that are not disjoint (overlapping), clear, and complete. All of the listed quality dimensions are contextual; hence, such a class contains all of them. Liu and Chie (2002) raised similar objections to the "Conceptual Framework …" in the following way: "... 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).

Thus, the cited empirical research and its results are of exploratory and transient value and may serve well for situation-specific improvement of quality. These improvements are well known and fully acknowledged; however, they are not a reliable rationale in yielding results of a more lasting scientific validity. Empirical studies may serve well for exploring users' preferences, and/or for first confirming or rejecting hypotheses derived from a critical qualitative examination of the subject.

Conclusions

The paper with its provisional model of most (likely not all) of the qualitative functional dependencies of joint actionable credibility of D/I values offers

- an explanation of the terminological advantage of the label **credibility** over **reliability** or **believability** of data/information values

- identification of, it is hoped, the most indirect DQ/IQ attributes, which directly and indirectly affect the joint credibility of data/information values. It may facilitate assessment, whether **actionable credibility** or **economical actionable credibility** are attained.

- the definition of **Joint credibility of data/information** values as a direct function of indirect quality attributes of the first order, such as source-specific credibility, variety of sources, mapping quality of the delivery system, and, finally, the presentation credibility

of these values, if indirect informing is the mode of operation (Part B of the provisional model)

- the definition of **reputation of data/information sources** as a function of many indirect quality attributes of an external nature **(traceability**, **availability of reliable communication channels - communicable** sources**, alignment of attitudes and alignment of interest** between the source and the clients) of a more intrinsic nature **(reliability, verifiability**, **replicability** of values, and **warranty**) that increase the reputation of the source, and which, due to imperfections, decrease the source's overall reputation (**definition, variability, objectivity, accuracy, precision,** and **currency** (Part A of the provisional model)

- an economic sequence of their examination

- a cursory comparison of the proposed **provisional model** with the Conceptual Framework of Data Quality empirically derived by Wang and Strong (1996) with regard to BELIEVABILITY and credibility in particular. The latter model turns out to be inconsistent, incomplete, and lacking clear, disjoint, exhaustive classification criteria and misses important qualities 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.*

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). Operations 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 systems. (2005). Encyclopedia Britannica Online. Retrieved April 22, 2005, from 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

Wand, Y., & Wang R. Y. (1996b). 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, 4 books, and 6 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 15 research papers published in the proceedings of ASIS, ACM, DSI*, Information Systems Educators Conference*, *The Informing Science Institute*, and the *International Conference 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 *the Institute of Informing Science.* His resent research focuses on quality of data and information values in informing.