DATA QUALITY DIMENSIONS FOR INFORMATION SYSTEMS SECURITY: A THEORETICAL EXPOSITION (INVITED PAPER)

Gurvirender Tejay, Gurpreet Dhillon and Amita Goyal Chin
Department of Information Systems, School of Business, Virginia Commonwealth University, Richmond, Virginia, USA

Abstract: Data is an important asset used for various organizational activities. Poor data quality could have severe implications for information systems security in organizations. In this paper, data is viewed as embodied in the concept of signs. This paper identifies dimensions of data quality by using semiotics as a theoretical basis. We argue that the nature and scope of data quality dimensions changes as we move between different semiotic levels. An understanding of these changes is essential for ensuring information systems security.

Key words: Data quality, information quality, knowledge quality, semiotics, information systems security.

1.   INTRODUCTION

Information systems (IS) security has become a core business processes in organizations (Trček, 2003). It not only deals with the computer systems but also the environment in which information is created and used (Armstrong, 1999). As Dhillon (1995) argues, "IS security can be viewed in terms of minimizing risks arising because of inconsistent and incoherent behavior with respect to the information handling activities of organizations." This has led to an increased concern associated with the

protection of organizational information assets (Dhillon & Backhouse, 2000). As Feldman & March (1981, pp.177) argue:

> There is no institution more prototypically committed to the
> systematic application of information to decisions than the
> modern bureaucratic organization...Information is not simply
> a basis for action. It is a representation of competence and a
> reaffirmation of social virtue.

Data is a fundamental information asset. Dhillon & Backhouse (2000) acknowledge that the primary focus of IS security is on data. However, the authors argue for IS security research to also address the changing organizational context in which data is interpreted and used.

Various organizational activities require data in some form or manner. Wang *et al.* (2001) give a grim warning that important corporate initiatives are at a risk of failure unless data quality is seriously considered and improved. Redman (2001) argues that poor data quality would impact setting strategy, its execution, ability to derive value from data, and ability to align the organization. It would also affect decision-making capability, trust between organizations, customer satisfaction, and employee morale. These factors are in fact epidemic to the success of IS security initiatives in an organization. The argument of this research is that we need to understand the inter-dependent nature of data and information to identify the relevant dimensions of data quality. The use of these dimensions would be effective in solving the data quality problems, which in turn would bolster the IS security efforts of an organization.

This paper is organized into five sections. Following a brief introduction, section two presents the notion of signs and discusses the relationship between signs, data and IS security. It provides a basis for the use of semiotics to data quality. This section also explains semiotics and its main branches. Section three presents a semiotic interpretation of data quality dimensions based on the research literature. Section four provides a discussion on the lessons learned from the semiotic analysis of data quality. Finally, section five presents the conclusions.

2.   SIGNS, DATA AND IS SECURITY

Peirce (as quoted in Falkenberg *et al.,* 1998, pp.51) defined sign "as something which stands to somebody for something in some respect or

capacity." It is comprised of three constituent elements: the symbol (signifying), the referent (signified) and the concept or the idea of significance. Data is considered to be meaningful symbolic constructs, which are a finite "arrangement" of symbols taken from an alphabet (Falkenberg *et al.,* 1998). In other words, the above statements indicate that data is comprised of symbols, which are constituent elements of signs. As suggested by Stamper (1973), signs include the numerical and alphabetical characters, words, sentences, messages and the actions that are interpretable. That is, signs are the basis of data. This allows the use of semiotics as a theory for analyzing data quality dimensions. Semiotics is the process of analyzing signs (Liebenau & Backhouse, 1990). It is concerned with the properties of signs (Stamper, 1973). Semiotics helps in understanding how a sign is created, processed, and used.

The inherent complexity of data calls for using semiotics as a theoretical basis. Data has various explications depending upon the nature of its use. It is utilized at different levels of granularity and frequency. The same data is employed in a large variety of decision tasks and is also shared among multiple stakeholders - data providers, decision-makers, and data custodians. These myriad complexities need to be acknowledged so as to address the problems associated with data quality. Semiotics may be used as a framework to study complex problems where the function and structure of the problem do not lend to an adequate explanation (Nake, 2001).

Semiotics elucidates the intricacies associated with a sign as it moves from the physical world, where it is created, to the social world of norms, where a meaning is attached to it. Semiotic analysis can be used to understand the technical, formal and informal systems of an organization (Liebenau & Backhouse, 1990). Dhillon (1995) conceptualizes IS security in terms of these three sub-systems. Dhillon argues the technical level involves technology and system security measures, while formal rules and procedures address the IS security issues arising at the formal level. At the informal level, pragmatic concerns are paramount towards the development of security culture and environment (Dhillon, 1995).

Table 1. Semiotic levels

Semiotic level	Description
Pragmatics	Concerned with relationship between signs and behavior in context
Semantics	Study of meaning of signs
Syntactics	Study of logic and grammar of sign systems
Empirics	Study of physical properties of sets of signs used to transmit messages

Semiotics analyzes signs at four levels: empirics, syntactics, semantics and pragmatics (Table 1).

2.1   Empirics

Empirics is the study of the physical properties of signs. It is concerned with the signals used to code and transmit the message. It describes the physical characteristics of the medium of communication (Liebenau & Backhouse, 1990). Empirics involves the study of communication channels and their characteristics such as sound and electronic transmission. At an empiric level, signs are viewed as a collection of signals that need to be transmitted from one place to another through the communication channels. At this level, we do not consider what the signals are portraying or what they mean. Data quality at this level is concerned with different data types being generated. Ensuring that the integrity of data at an empirical level is maintained results in good communication and network security.

2.2   Syntactics

Syntactics is the study of logic and grammar of sign systems. It is concerned with the structures rather than the content of signs. Here the focus is on the physical form, regardless of any statistical properties. Syntactics allows for "the constructions of formal rules and the means by which they interrelate" (Liebenau & Backhouse, 1990). It studies the relation of signs to one another and how complex signs originate from simple ones. Syntactics level informs us about how signs signify. Formalism is the key term at this level. At a syntactic level, data quality is ensured through concentrating on the formal and structural aspects. Security at a syntactic level assures a high level of integrity in the formalisms and correctness in system specification.

At a syntactic level, information is viewed only from its material aspect – the structure. A set of signs when formulated together in a structure and governed by certain formal rules, becomes data. In other words, data is the mass of facts that have been created according to a certain structure. As such, we can argue that the type of signs operating at the syntactics level is actually data.

2.3   Semantics

Semantics is the study of the meaning of signs. It deals with the purported relationships between signs and what they signify (Stamper, 1985). Signs become useful only when they indicate certain action. For the interpretation of sign we need to ascertain the meaning of what is expressed in a message. A particular sign would imply different meaning to different individuals in a particular social setting. Semantics is concerned with the explication of various meanings associated with a sign. It deals with "the connections that agents make between the signs that they use and their behavior and actions" (Liebenau & Backhouse, 1990).

Semantics ascribes meaning to the data from the syntactics level. When meaning is attributed to the selected data in a particular context it becomes information. Checkland & Howell (1998) define information as structured data that has contextual meaning. Avison & Fitzgerald (1995) consider information to have meaning that comes from selecting data, summarizing it and presenting it in such a way that it is useful to the recipient. Meaning is appropriated when we draw from our knowledge and apply understanding to information. We deal with only information at the semantics level. Ensuring good quality at a semantic level therefore moves from maintaining quality of data to that of information.

2.4   Pragmatics

Pragmatics is concerned with the relationships between signs and behavior in a given context. It concentrates on the shared assumptions underlying human understanding. The pragmatic structures are the beliefs, expectations, commitments, intentions and communication patterns of people (Dhillon, 1995). The focus of pragmatics is to grasp the intentions of the sender of the message. The emphasis is also on the context of intentional use. At this level, we need to understand the usage of signs by people. This addresses the reason for the signs' signification.

The pragmatic level involves information that has intentional use. At the earlier levels, the concept of signs has been acted upon in terms of providing structure and attributing meaning to make it useful. This matured form is knowledge. Information becomes knowledge at the moment of its interpretation (Miller, 2002). Nonaka & Takeuchi (1995, pp. 58) suggest a relationship to knowledge:

...information is a flow of messages, while knowledge is created
by that very flow of information anchored in the beliefs and
commitment of its holder. This...emphasizes that knowledge is
essentially related to human action.

In this section, we have explained the theoretical concepts of semiotics. Semiotics helps us understand and traverse the fine course between the business world and the physical world. The discussion on the four levels of semiotics informs us that the pragmatic level deals with knowledge, semantic level is the information domain, syntactic level is associated with data, and the empiric level involves communication channels. The next section discusses the data quality dimensions at the four semiotic levels.

3.   DATA QUALITY DIMENSIONS

In this section, the semiotic analysis of data quality dimensions is presented. The data quality dimension, as defined by Wang & Strong (1996), is a set of data quality attributes that represent a single aspect or construct of data quality. Table 2 summarizes the analysis of dimensions that have been proposed in the research literature.

Table 2. Semiotic analysis of Data Quality dimensions

Semiotic levels	DQ Dimensions	Seminal work
Empirics	Accessibility	Delone et. al. (1992), Goodhue (1995), Miller (1996), Wang et al. (1996)
	Timeliness	Ballou et al. (1985), Caby et al. (1995), Fox et al. (1994), Goodhue (!995), Hilton (!979), Miller (1996), Wang et al. (1996), Zmud (1978)
	Locatability	Goodhue (1995)
	Portability	Caby et al. (1995)
	Security	Miller (1996), Wang et al. (1996)
Syntactics	Accuracy	Ballou et al. (1985), Caby et al. (1995), Delone et al. (1992), Doernberg et al. (1980), Fox et al. (1994), Goodhue (1995), Hilton (1979), Miller (1996), Wang et al. (1996), Zmud (1978)

Semiotic levels	DQ Dimensions	Seminal work
	Appearance, Comparability, Freedom from bias, Precision, Redundancy, Uniqueness, Usable	Delone et al. (1992)
	Arrangement, Readable	Zmud (1978)
	Clarity, Ease of use, Presentation	Goodhue (1995)
	Coherence, Format	Miller (1996)
	Compatibility	Goodhue (1995), Miller (1996)
	Composition, Flexibility, Robustness	Caby et al. (1995)
	Conciseness	Delone et al. (1992), Wang et al. (1996)
	Consistency	Ballou et al. (1985), Caby et al. (1995), Fox et al. (1994), Wang et al. (1996)
	Correctness	Wand et al. (1996)
	Ease of operation, Objectivity	Wang et al. (1996)
	Integrity	Brodie (1980)
	Level-of-detail	Caby et al. (1995), Goodhue (1995)
Semantics	Ambiguity	Doernberg et al. (1980), Wand et al. (1996)
	Believability, Understandability	Wang et al. (1996)
	Content, Informativeness	Delone et al. (1992)
	Factual, Reasonable	Zmud (1978)
	Interpretability	Wang et al. (1996), Caby et al. (1995)
	Meaningful	Goodhue (1995), Wand et al. (1996)
	Reliability	Brodie (1980), Delone et al. (1992), Goodhue (1995), Zmud (1978)
	Validity	Miller (1996)
Pragmatics	Appropriate amount of data, Reputation, value-added	Wang et al. (1996)
	Appropriateness	Caby et al. (1995)
	Completeness	Ballou et al. (1985), Caby et al. (1995), Doernberg et al. (1980), Fox et al. (1994), Miller (1996), Wang et al. (1996), Wand et al. (1996)
	Relevance	Delone et al. (1992), Hilton (1979), Miller (1996), Wang et al. (1996)

Semiotic levels	DQ Dimensions	Seminal work
	Importance, Sufficiency, Usefulness	Delone et al. (1992)

3.1 Empiric dimensions

Empirics deals with the problems in which data is used repeatedly. This level is concerned with establishing means of communication and data handling. The focus is on different data types being generated and their risk of being transmitted erroneously. The dimensions operating at the empiric level include accessibility, timeliness, security, portability and locatability.

Accessibility implies that data should be available, obtainable or retrievable when needed. This view is supported by Delone & McLean (1992), Goodhue (1995), Miller (1996), Wang & Strong (1996) and Chin & Becker (1997). Timeliness is concerned with the age of data, whether data is current. It is achieved if the recorded value is not out of date. Zmud (1978) and Hilton (1979) were the early proponents of timeliness as a dimension. Ballou & Pazer (1985), Caby *et al.* (1995), Fox *et al.* (1994), Goodhue (1995), Miller (1996), and Wang & Strong (1996) also support this dimension. Security as a dimension was proposed by both Miller (1996) and Wang & Strong (1996). Description of this dimension involves keeping data secure and restricting access to it. It involves protecting data from people and natural disasters. Portability and locatability as dimensions were supported by Caby *et al.* (1995) and Goodhue (1995) respectively.

The dimensions operating at the empiric level are associated with the problems of medium of communication rather than data itself. Unavailability of communication channels would lead to the problems of accessibility, timeliness and portability. Unavailable channels would hinder the data from being accessible and updated as per the requirements. On the other hand, unauthorized access to the communication channel would lead to the problems of security. In short, the empiric level deals only with network communication.

3.2 Syntactic dimensions

Syntactics deals with the forms and structures of data. It is concerned with the physical form rather than the content of data. The data quality

dimensions operating at the syntactic level include accuracy (Chin & Becker, 1997), appearance (Delone & McLean, 1992), arrangement (Zmud, 1978), clarity (Goodhue, 1995), coherence (Miller, 1996), comparability (Delone & McLean, 1992), compatibility (Goodhue, 1995; Miller, 1996), completeness, composition (Caby *et al.*, 1995), conciseness, consistency, correctness, ease of operation, ease of use (Goodhue, 1995), flexibility, format, freedom from bias (Delone & McLean, 1992), integrity, level-of-detail (Caby *et al.*, 1995; Goodhue, 1995), objectivity, portability, precision (Delone & McLean, 1992), presentation (Goodhue, 1995), readable (Zmud, 1978), redundancy, robustness (Caby *et al.*, 1995), uniqueness (Delone & McLean, 1992), and usable (Delone & McLean, 1992).

Accuracy dimension is concerned with the conformity of the recorded value with the actual value. It is a widely accepted dimension of data quality. Accuracy implies that data is correct, flawless, precise, reliable and certified free of error. Zmud (1978), Hilton (1979), Doernberg & Ziegler (1980), Ballou & Pazer (1985), Delone & McLean (1992), Fox *et al.* (1994), Goodhue (1995), Miller (1996), Wang & Strong (1996) and Caby *et al.* (1995) support accuracy as an important dimension of data quality in their research.

Completeness as a dimension has been supported in the literature, with Doernberg & Ziegler (1980) and Ballou & Pazer (1985) being the earlier advocates. It implies that data must be of sufficient breadth, depth, and scope for the task at hand. Completeness involves recording of all values for a certain variable and is concerned with loss of data. Fox *et al.* (1994), Caby *et al.* (1995), Miller (1996), Wang & Strong (1996), and Wand & Wang (1996) also support this dimension.

Conciseness represents well-presented, succinct, and compact representation of data. This view is advocated by Delone & McLean (1992) and Wang & Strong (1996). Consistency is achieved if the representation of the data value is the same in all cases. It involves continuous representation of data in same format, compatibility with previous data, and consistent representation. Ballou & Pazer (1985), Caby *et al.* (1995), Fox *et al.* (1994) and Wang & Strong (1996) are the proponents of consistency as a dimension of data quality.

Wand & Wang (1996) describe the dimension of correctness as arising from garbling (operational failure), when data derived from the information

system, which does not conform to those used to create these data. Ease of operation, as advocated by Wang and Strong (1996), implies that data is manipulable, easily integrated, customized, and usable for multiple purposes. This dimension is similar to the dimension of flexibility as proposed by Caby *et al.* (1995). Flexibility is addressed if data is adaptable, flexible, extendable, and expandable. Miller (1996) describes format as concerned with how the data is presented to the customer.

Integrity as a dimension of data quality was proposed by Brodie (1980). It is a measure of correctness and consists of semantic and physical integrity. Semantic integrity measures consistency and completeness with respect to the rules of the description language. Physical integrity measures the correctness of implementation details. Wang & Strong (1996) identified objectivity as another important dimension. It is concerned with data as being unbiased and impartial.

The dimensions discussed above cover the logic, grammar and structural aspects that compose data. In order to resolve the data quality issues, future research should concentrate specifically on the above discussed dimensions while proposing approaches for improvement.

3.3   Semantic dimensions

The dimensions operating at semantic level are associated with information rather than data. Information is the selected data that has been attributed meaning in a particular context. Semantics is concerned with meaning. It deals with the interpretation of data to get at the meaning of communication. The data quality dimensions from the research literature operating at the semantic level include ambiguity, believability, interpretability, meaningful, reliability, understandability, validity, content, informativeness, factual, and reasonablity. Ambiguity arises as a result of improper representation and is when data can be interpreted in more than one way. This dimension was proposed by Doernberg & Ziegler (1980) and Wand & Wang (1996).

Wang & Strong (1996) considered the dimension of believability as concerned with whether data can be believed or regarded as credible. Interpretability simply means that data should be interpretable, that is, it should be both defined clearly and represented appropriately. This dimension is supported by Caby *et al.* (1995) and Wang & Strong (1996).

Goodhue (1995) and Wand & Wang (1996) consider meaningfulness as an important dimension that is concerned with the interpretation of data. The failure of this dimension results in meaningless data.

Brodie (1980) understood reliability as linked to whether data was dependable, trustworthy and could be relied upon. Agmon & Ahituv (1987) addressed reliability in terms of concepts drawn from the field of quality control. Other researchers who support reliability as a dimension include Zmud (1978), Delone & McLean (1992) and Goodhue (1995). Understandability as a dimension was proposed by Wang & Strong (1996). It is concerned with whether data is clear, readable, unambiguous and easily comprehendible. Data is valid when it can be verified as being true and satisfying appropriate standards related to other dimensions (Miller 1996). The dimensions of content and informativeness were proposed by Delone & McLean, while Zmud (1980) considered factual and reasonablity as the other dimensions.

Ambiguity, interpretability and content as dimensions are concerned with the interpretational aspects of semantics. The issue of credibility is an associative characteristic of meaning. Believability, reliability, validity and factual dimensions strive to capture this issue. Meaning ascribed in semantics is addressed by the meaningful, understandability, informativeness and reasonablity dimensions proposed as part of data quality. The dimensions discussed in this section are actually associated with information. Ascertaining the meaning would lead to appropriate interpretation of data. As such, the issue of interpretation, credibility and meaning are linked to the aspects of information. The dimensions discussed in this section do not form part of data quality and should not be included as such.

3.4   Pragmatic dimensions

Pragmatics deals with the use of information by people. It is concerned with the relation between data, information and behavior, in a given context. The data quality dimensions from the research literature at the pragmatic

level[1] include appropriateness, relevance, value added, sufficiency, importance, usefulness, and reputation. These dimensions deal with the pragmatic issues. Intention of use is the underlying focus of these dimensions.

Appropriateness, as a data quality dimension, means that data must be appropriate to the task at hand. Caby *et al.* (1995) and Wang & Strong (1996) are proponents of the appropriateness dimension. Hilton (1979) considered relevancy as an important dimension. This dimension is concerned with the applicability of data to the task at hand. Relevance is a key dimension as if the data does not address its customer needs and when the customer will find the data inadequate. Delone & McLean (1992), Miller (1996) and Wang & Strong (1996) are the other supporters of this view. Wang & Strong (1996) proposed value-added as a dimension that addresses the benefits and advantages of using data. Delone & McLean (1992) also considered sufficiency, importance and usefulness, while Wang & Strong (1996) concentrated on reputation as a valid dimension.

The contextual aspects of pragmatic issues are addressed by appropriateness, relevance and usefulness dimensions. Sufficiency, importance and reputation as dimensions deal with the expectations of the use. Value-added dimension attempts to understand the intention of use. These dimensions are concerned whether data fits the problem task. In doing so, these dimensions are not associated with data quality. That is, these dimensions do not address the issues of data quality. But rather they deal with the issues of knowledge. These dimensions are concerned with the intentional use, that is, how data would be used in relation to the problem at hand. This leads us to the domain of knowledge.

4.   DISCUSSION

In the field of data quality, there is a lack of theoretical basis for identifying, classifying and discussing data quality dimensions. The analysis presented in the previous section allows us to develop an in-depth

---

[1] Although in this paper we argue that at a semantic and pragmatic level, the nature of data changes to information and knowledge, the dominant research literature does not make this distinction.

understanding of the dimensions associated with data quality. Indeed, there are complexities involved with data. Semiotics assists us in this endeavor to unravel and understand the inter-dependent nature of data, information and knowledge. This provides a clear understanding of the dimensions associated with data quality.

Researchers have proposed various solutions and approaches to the problems of data quality by addressing dimensions that actually exist at the pragmatic and semantic levels. A careful analysis of different levels of semiotics would indicate that the pragmatic level is associated with knowledge; semantic level is associated with information, while only the syntactic level is associated with data. In other words, the dimensions operating at the pragmatic, semantic and syntactic levels pertain to knowledge quality, information quality and data quality respectively. In order to improve the quality of data, special attention should be paid to the dimensions operating at the syntactic level. At the same time, dimensions associated with knowledge quality and information quality cannot be ignored. These dimensions should also be acknowledged as part of any equation devised to address the problems of data quality.

Data quality problems range from definition, measurement, analysis and improvement to tools, methods, and processes (Wang *et al.,* 2001). Semiotic interpretation of data quality dimensions addresses the definition, measurement and analysis aspects of data quality, while the improvement aspect is touched indirectly. In terms of measurement, data quality dimensions can also serve as the metrics (Ballou & Pazer, 1985). Wang et al. (2001, pp.12) have used timeliness, security and credibility, among others, as part of data quality metrics. The trouble with such metrics is that it leads us away from the domain of data quality and into the fields of networking, IS security and telecommunications. This problem highlights the need for a sensitive yet deep examination of the dimensions that actually pertain to data quality.

Let us consider the case of a data warehouse. Data warehouse is a repository of data needed for decision making (Ballou & Tayi, 1999). Wang *et al.* (2001), while explaining the European Union Data Warehouse Quality Project, provide the following as data warehouse quality parameters: accessibility, interpretability, usefulness, believability and validation. Given our argument, none of these parameters are related to data (syntactics) quality. These are in fact dimensions of communication (empirics) quality,

information (semantics) quality and knowledge (pragmatics) quality. Although, we are concerned with ensuring the quality of data, our current emphasis is on dimensions that do not directly impact data quality. In doing so, we are setting up for disappointment in data quality efforts.

The semiotic analysis of data quality dimensions assists us in solving the measurement issues related with data quality. The dimensions of data quality exist at the syntactic level and may be objectively measured, to an extent, as compared to those of information quality. Information quality dimensions like informativeness, meaningful or believability would be problematic to measure as these are subjective in nature. Their measurement cannot escape the subjective valuation by users.

4.1   Implications for IS Security

Semiotics lends itself as a theoretical framework by which we can understand the nature of data, information and knowledge. It also makes us sensitive to the distinctions among the three. The intricacies involved with data, information and knowledge gives rise to impervious gaps that can be epidemic to the success of an organization. As such, careful thought is required to manage these gaps (see figure 1).

Receptivity gap exists between the empiric and the syntactic level. It arises as a result of problems encountered with transmitting signals through the physical communication channels. As we move from data to information, we would encounter the interpretation gap. Interpretation gap is the interplay between data and information. It arises due to lack of appropriate meaning being attached to data. Usefulness gap is encountered when we derive knowledge from information. It arises when the information is construed with inappropriate intention in a particular context.

Receptivity, interpretation and usefulness gap have implications for IS security and organizational issues. Receptivity gap has technical implications; interpretation gap has an impact on operations, while usefulness gap impacts both decision-making and overall strategy of an organization. In terms of IS security, lack of adequate technical controls would result in a receptivity gap. At this level, the observance of principles of confidentiality, integrity and availability is paramount. The presence of this gap indicates technical IS security failure.

Interpretation gap has an impact on operations. In terms of IS security, it can be addressed in terms of the formal controls. Misinterpretation of IS security policies existing in the form of data would have severe implications for an organization. The non-contextual use of data might lead to incorrect allocation of responsibilities. This would produce an exposure of a weak link. Further, misinterpretation of data would also result in poor predicates for the process of decision-making. It would provide us with incorrect operational information about different processes crucial for IS security. As such, interpretation gap would have severe implications on the effectiveness and efficiency of an economic organization.

Usefulness gap has implications for decision-making capability and affects the strategic function of the organization as well. Inappropriate use of information pertaining to IS security in a particular contextual situation would result in poor decision predicates and in turn, a bad decision. The decision-making capability of an organization would be adversely hampered if the information is used in an inadequate manner. It is also important to understand the context of the information provided. The failure to do so would lead to the establishment of IS security objectives or goals that are based on improper (out-of-context) information. This would result in diverting the focus to a wrong set of priorities. Informal controls of an organization are important to check this gap. The principles of responsibility, integrity, trust and ethics, as advocated by Dhillon & Backhouse (2000), would stand in good stead to overcome this gap.

In terms of strategic function, inappropriate information would influence the policy and strategy formulation. The usefulness gap would also affect the implementation and establishment of corporate security policy and strategy. It would result in failure to align the IS security objectives with those of the organization, subsequently impacting the management's focus. As such, usefulness gap would have implications for the existential thrive of an organization.
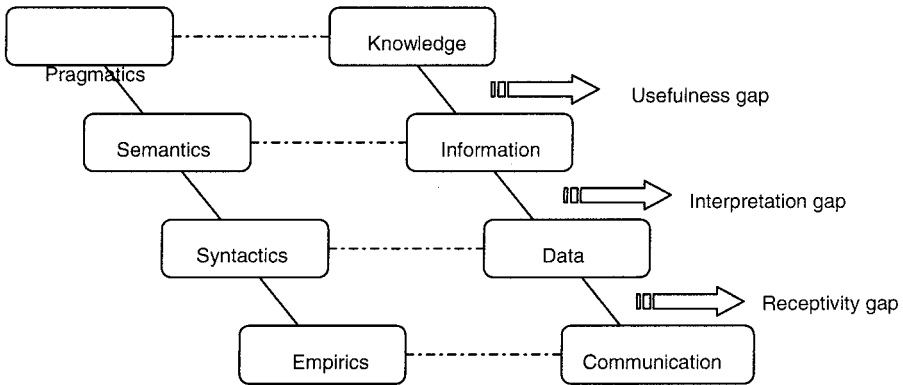
Figure 1. Semiotics, data-information-knowledge and the gaps

5. CONCLUSION

In this paper, data is approached as comprised of signs. We have provided a deeper understanding of data and information. We argue for the need to understand the inter-dependent nature of data, information and knowledge. This would allow us to address the problems of data quality and propose effective approaches towards its improvement. Such an approach would lead to better decision predicates for the process of decision making in modern organizations.

Semiotics is used as a frame of reference to explicate the concepts of data and information to allow critical analysis of dimensions of data quality. This investigation enhances our understanding that the dimensions associated with data quality can only exist at the syntactic level of semiotics. This alleviates the problems of definition and measurement. Further, usefulness and interpretation gaps arise as we attempt to extract knowledge from information and information from data. These gaps have implications for IS security that can be adequately addressed by establishing appropriate formal, as well as, informal controls in an organization.

In terms of future research, it would be valuable to draw implications for organizations based upon the semiotic view of data quality dimensions

presented in this research. These implications might address how to measure, assess, manage and improve the quality of data. Limitations of this research arise from the use of semiotics as a frame of reference. We do not state that semiotics is the only theoretical basis from which to look at the concepts of data and information. However, we certainly assert that semiotics as a theory of information is relevant and rewarding in solving the problems with the use of information. Semiotics encompasses all aspects, the mechanical, the abstract, the philosophical and the human (Stamper 1985).

6. REFERENCES

Abate, M.L., Diegert, K.V., and Allen, H.W., 1998, A hierarchical approach to improving data quality, *Data Quality*. 4(1).

Agmon, N., and Ahituv, N., 1987, Assessing data reliability in an information system, *Journal of Management Information Systems*. 4(2): 34-44.

Anderson, James (2003). Why we need a new definition of IS security.

Armstrong, H.L., 1999, *A Soft Approach to Management of IS Security*. Doctoral dissertation, Curtin University of Technology.

Arnold, S.E., 1992, Information manufacturing: The road to data quality, *Data Base*. 15(5): 32-39.

Avison, D.E., and Fitzgerald, G., 1995, The Field of IST, Private communication to Peter Checkland.

Ballou, D.P., and Pazer, H.L., 1985, Modeling data and process quality in multi-input, multi-output information systems, *Management Science*. 31(2): 150-162.

Ballou, D.P., Wang, R., Tayi, G.K., and Pazer, H.L., 1998, Modeling information manufacturing systems to determine information product quality, *Management Science*. 44(4): 462-484.

Baskerville, R., 1989, Logical controls specification: An approach to information systems security, in: *Systems Development for Human Progress,* H.Klein, and K.Kumar, ed., North-Holland, Amsterdam.

Boland, R.J., 1987, The in-formation of information systems, in: *Critical Issues in Information Systems Research*, R.J.Boland, and R.Hirschheim, ed., John Wiley and Sons, Chichester.

Brodie, M.L., 1980, Data quality in information systems, *Information and Management*. 3(6): 245-258.

Caby, E.C., Pautke, R.W., and Redman, T.C., 1995, Strategies for improving data quality, *Data Quality*. 1(1): 4-12.

Checkland, P., and Howell, S., 1998, *Information, Systems, and Information Systems – Making Sense of the Field,* John Wiley and Sons, Chichester.

Chengalur-Smith, I.N., Ballou, D.P., and Pazer, H.L., 1999, The impact of data quality information on decision making: An exploratory analysis, *IEEE Transactions on Knowledge and Data Engineering*. 11(6): 853-864.

Chin, A., and Becker, S., 1997, Improving decision making using confidence scaling for enhanced data quality, AIS Americas Conference, Indianapolis.

Delone, W.H., and McLean, E.R., 1992, Information systems success: The quest for the dependent variable, *Information Systems Research.* 3(1): 60-95.

Dhillon, G., 1995, *Interpreting the Management of Information Systems Security,* Doctoral dissertation. London School of Economics and Political Science.

Dhillon, G. and Backhouse, J., 2000, Information system security management in the new millennium, *Communications of the ACM.* 43(7): 125-128.

Doernberg. D.L., and Ziegler, D.H., 1980, Due process versus data processing: An analysis of computerized criminal history information systems, *New York University Law.* 1: 55.

Falkenberg, E.D., Hesse, W., Lindgreen, P., Nilsson, B.E., Oei, J.L.H., Rolland, C., Stamper, R.K., Assche, F.J.M.V., Verrijn-Stuart, A.A., and Voss, K., 1998, *A Framework of Information Systems Concepts: The FRISCO Report.* University of Leiden, International Federation for Information Processing, The Netherlands.

Feldman, M.S., and March, J.G., 1981, Information in organizations as signal and symbol, *Administrative Science Quarterly.* 26: 171-186.

Fox, C., Levitin, A., and Redman, T., 1994, The notion of data and its quality dimensions, *Information Processing and Management.* 30(1):9-19.

Funk, J., Lee, Y., and Wang, R., 1998, Institutionalizing information quality practice, *Proceedings of Conference on Information Quality, Cambridge, MA*, pp.1-17.

Goodhue, D.L., 1995, Understanding user evaluations of information systems, *Management Science.* 41(12): 1827-1844.

Guynes, C.S., Prybutok, V.R., and Myers, B.L., 1996, Evolving data quality considerations for client/server environments, *Data Quality.* 2(1): 21-27.

Hilton, R.W., 1979, The determinants of information system value: An illustrative analysis, *Journal of Accounting Research.* Autumn: 411-35.

Huh, Y.U., Keller, F.R., Redman, T.C., and Watkins, A.R., 1990, Data quality, *Information and Software Technology.* 32(8): 559-565.

Jarke, M., and Vassiliou, Y., 1997. Data warehouse quality: A review of the DWQ project, *Proceedings of the Conference on Information Quality, Cambridge, MA*, pp. 102-115.

Kahn, B.K., Strong, D., and Wang, R.Y., 2002, Information quality benchmarks: Product and service performance, *Communications of the ACM.* 45(4): 184-192.

Klein, B.D., and Rossin, D.F., 1999, Data errors in neural network and linear regression models: An experimental comparison, *Data Quality.* 5(1).

Lee, Y., Allen, T., and Wang, R.Y., 2001, Information products for remanufacturing: Tracing the repair of an aircraft fuel-pump, *Proceedings of the 6th International conference on Information Quality,* pp. 77-82.

Lee, Y.W., Strong, D.M., Kahn, B.K., and Wang, R.Y., 2002, AIMQ: A methodology for information quality assessment, *Information and Management.* 40: 133-146.

Liebenau, J., and Backhouse, J., 1990, *Understanding Information: An Introduction.* Macmillan, Basingtoke.

Mackay, H., 1998, *The Good Listener,* Pan Macmillan, Sydney.

Miller, F.J., 2002, I = 0 (Information has no meaning), *Information Research.* 8(1).

Miller, H., 1996, The multiple dimensions of information quality, *Information Systems Management.* 13(2): 79-83.

Morris, C., 1946, *Signs, Language and Behavior,* Prentice Hall Braziller, New York.

Nake, F., 2001, Data, information, and knowledge, in: *Organizational Semiotics: Evolving a science of information systems*, K.Liu, R.J.Clarke, P.B.Anderson, R.K.Stamper, and E.Abou-Zeid, ed., Kluwer Academic Publishers, Boston, pp. 41-50.

Nonaka, I., and Takeuchi, H., 1995, *The Knowledge-Creating Company: how Japanese companies create the dynamics of innovation,* Oxford University Press, New York.

O'Leary, D.E., 1993, The impact of data accuracy on system learning, *Journal of Management Information Systems.* 9(4): 83-98.

Pesche, M., 1999, Boundary bath, *Presented at SCOPE1 workshop,* Vienna.

Pipino, L., Lee., Y.W., and Wang, R.Y., 1998, Measuring information quality, *TDQM-97-04, Total Data Quality Management Program,* MIT Sloan School of Management.

Redman, T.C., 1996, *Data Quality for the Information Age,* Artech House, Boston.

Redman, T.C., 1998, The impact of poor data quality on the typical enterprise, *Communications of the ACM.* 41(2): 79-82.

Redman, T.C., 2001, *Data Quality: The field guide,* Digital Press, Boston.

Shankar, G., Wang, R.Y., and Ziad, M., 2000, IP-MAP: Representing the manufacture of an information product, *Proceedings of the 5th International conference on Information Quality,* pp. 1-16.

Shankaranarayan, G., Ziad, M., and Wang, R.Y., 2003, Managing data quality in dynamic decision environments: An information product approach, *Journal of Database Management.* 14(4): 14-32.

Specht, P. H., 1986, Job characteristics as indicants of CBIS data requirements, *MIS Quarterly.* September: 270-286.

Stamper, R.K., 1973, *Information in Business and Administrative Systems,* John Wiley and Sons, New York.

Stamper, R.K., 1985, Towards a theory of information: Mystical fluid or a subject for scientific enquiry? *The Computer Journal.* 28(3): 195-199.

Strong, D.M., 1997, IT process designs for improving information quality and reducing exception handling: A simulation experiment, *Information and Management.* 31: 251-263.

Trček, D., 2003, An integral framework for information systems security management, *Computers & Security.* 22(4):337-360.

Wand, Y., and Wang, R.Y., 1996, Anchoring data quality dimensions in ontological foundations, *Communications of the ACM.* 39(11): 86-95.

Wang, R.Y., 1998, A product perspective on Total Data Quality Management, *Communications of the ACM.* 41(2): 58-65.

Wang, R.Y., Allen, T., Harris, W., and Madnick, S., 2003, An information product approach for total information awareness, *IEEE Aerospace Conference.*

Wang, R.Y., and Strong, D.M., 1996, Beyond accuracy: What data quality means to data consumers, *Journal of Management Information Systems.* 12(4): 5-34.

Wang, R.Y., Lee, Y.L., and Pipino, L., and Strong, D.M., 1998, Manage your information as a product, *Sloan Management Review.* 39(4): 95-105.

Wang, R.Y., Storey, V.C., and Firth, C.P., 1995, A framework for analysis of data quality research, *IEEE Transactions on Knowledge and Data Engineering.* 7(4): 623-640.

Wang, R.Y., Ziad, M., and Lee, Y.W., 2001, *Data Quality,* Kluwer Academic Publishers, Boston.

Zmud, R., 1978, Concepts, theories, and techniques: An empirical investigation of the dimensionality of the concept of information, *Decision Sciences.* 9(2): 187-195.