What constitutes information integrity?

S. Flowerday
Nelson Mandela Metropolitan University
Port Elizabeth
South Africa
flowerdayfamily@gmail.com

R. von Solms
Nelson Mandela Metropolitan University
Port Elizabeth
South Africa
rossouw.vonsolms@nmmu.ac.za

This research focused on what constitutes information integrity as this is a problem facing companies today. Moreover, information integrity is a pillar of information security and is required in order to have a sound security management programme. However, it is acknowledged that 100% information integrity is not currently achievable due to various limitations and therefore the auditing concept of *reasonable assurance* is adopted. This is in line with the concept that 100% information security is not achievable and the notion that *adequate security* is the goal, using appropriate countermeasures. The main contribution of this article is to illustrate the importance of and provide a macro view of what constitutes information integrity. The findings are in harmony with Samuel Johnson's words (1751): 'Integrity without knowledge is weak and useless, and knowledge without integrity is dangerous and dreadful.'

Contents

1 Introduction

The auditing profession has adopted the concept of *reasonable assurance*. This concept requires that the auditors perform enough work to obtain reasonable assurance that the information found within the financial statements of a company is free from materiality and is a *fair or faithful representation* of that company's

financial position. It is argued by critics who oppose the concept of reasonable assurance in favour of absolutes, that this is a way for auditors to reduce legal liability. Houck (2003:16), however, contends that this is unfair and asserts that 'reasonable assurance is a commonsense, cost-effective concept that enhances the integrity of financial reporting'.

This article addresses the research problem of 'what constitutes information integrity', as this is an important issue facing companies today. Moreover, *information integrity* is a pillar of *information security* and is therefore required in order to have a sound security management programme. However, it is acknowledged that 100% information integrity is not currently achievable due to various limitations and therefore the auditing concept of *reasonable assurance* is adopted. This is in line with the concept that 100% information security is not achievable and the notion that *adequate security* is the goal, using appropriate countermeasures.

The main contribution of this article is to illustrate the importance of and provide a macro view of what constitutes information integrity. Based on an extensive and critical literature analysis, the findings have emerged from a logical argument.

Moreover, to understand information integrity, it is necessary to first attempt to clarify what data and information are as well as to understand quality, even if at a basic level. This then enables the discussion to move forward.

2 Data, information and knowledge

It appears at times that the words *data* and *information* are used 'loosely' and some use these words interchangeably. Additionally, because there does not appear to be consensus on definitions for these words, particularly for the word *information*, an attempt is made to clarify their meanings solely for the purpose of this article and for the context in which information is viewed.
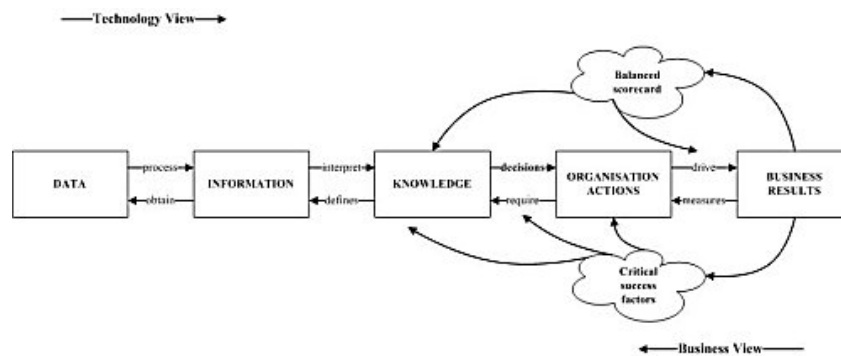
2.1 Data processed

Within the information systems and information technology (IS/IT) arena, it is accepted that *input (data) – process – output (information)* is agreed upon (O'Brien 2000; Oz 2002). The word data is the plural of the Latin word *datum*, though data commonly represents both singular and plural forms. According to O'Brien (2000:27), the word data refers to 'raw facts or observations, typically about physical phenomena or business transactions'. Oz (2002:8) clarifies the word information within the IS/IT context and states: 'Information can be raw data or data manipulated through tabulation, addition, subtraction, division, or any other operation that leads to greater understanding of a situation.'

The accepted paradigm of how information is viewed for the purpose of this research appears in Figure 1, even though some alternative views are discussed. Figure 1 is also known as Venkatraman's DIKAR model (data, information, knowledge, action, results). If the perception that d*ata processed* becomes information and *information interpreted* becomes knowledge is the generally accepted view within the IS/IT field, then it is important that *decisions*, based on knowledge, *direct* the company and ultimately determine the company's business results. As shown in Figure 1, there is a counter flow starting with the company's external environment that feeds back until it is captured as data and is again prepared for processing.

When the paradigm in Figure 1 is applied to stock in a company, the process could start with the data that were collected regarding the amount of stock stored at warehouses. The pure numbers by themselves are data. When processed (totalled), the data become information. Based on this information, interpreting the value of the stock becomes knowledge as to the value of the assets stored. Depending on the value of the assets, decisions can be made to increase sales or stock or, for example, apply for financial loans (bank overdraft), using the stock as collateral for these loans. As mentioned, this can help determine the company's actions and it affects the overall business results. This linear approach clearly illustrates the importance of having *accurate* and *reliable* figures (data) at the start of the process.

Figure 1 Information in context (Venkatraman 1996, in Ward and Peppard 2002:207)

2.2 What constitutes information?

In an attempt to describe what constitutes *information,* Machlup and Mansfield reviewed and edited numerous articles on the subject. This they did from various perspectives, covering many disciplines. Their project was to 'analyze the logical and pragmatic relations among the disciplines and subject areas that are centered on information'(1983:3). The content of this extensive book (Machlup and Mansfield 1983), which is often quoted, almost appears contradictory and confusing (depending which disciplines one subscribes to). However, this also highlights that one can find references and quotations supporting the different stances on the subject.

It is emphasized that the discipline, information theory, is at times referred to by alternate names within the scientific domain, such as 'mathematical theory of communication, communication theory, coding theory, signal-transmission theory, and mathematical theory of information measurement' (Machlup and Mansfield 1983:47). They continue and discuss whether information theory is chiefly about *information* or *communication* or about *signals.* Later they comment that the word *communication* is a possible alternative for the word *information.* However, they clarify that the words also have other meanings (Machlup and Mansfield 1983).

Another view to consider is that of Mesarovic (1983:569), who contends that information cannot be defined without reference to the goal-seeking behaviour of a system (in harmony with systems theory). He explains: 'A system is goal-seeking if its behaviour can be best described with reference to the pursuance of a given goal.' An example he uses to illustrate his point is that a system-dynamic approach would describe this in terms of the car's acceleration and speed. Mesarovic continues by saying that a goal-seeking description would require that 'the driver inside the car be identified and the moving point on the line representing the car on a highway be represented in terms of the strategy the driver uses in steering the vehicle along the road'.

To summarize Mesarovic's view (1983), a goal seeking description, therefore, requires a description of a goal, a description of a strategy and the description of the environmental conditions in which the strategy is being pursued. Mesarovic (1983:570) argues that the concept of information is 'much richer' than most perceive it to be and therefore he makes a bold statement that he 'would actually equate systems theory with information theory in an appropriately wider sense'. Additionally, he extends the disciplines and states that 'information theory and systems theory are one and the same'. This last statement has received some criticism.

Machlup (1983:642) contends that any other meanings for the word *information*, other than the ones he lists, are analogies, metaphors or concoctions, the first being 'the telling of something' and the second 'that which is being told'. He expounds on this and elaborates that there are many different methods, however, 'information is a flow of messages'(1983:643). Machlup does point out that it is not a requirement 'that information be correct and knowledge be true'. In fact, information can be misleading, incorrect and even false or fabricated.

Machlup (1983:645) also emphasizes that information involves at least two parties, 'one who tells (by speaking, writing, imprinting, pointing, signalling) and one who listens, reads, watches'. This is essential as one considers the sending of messages or signals to share information with another. Or, to reiterate, the board of directors who is sending messages or signals to the company stakeholders via the financial statements is sharing information. This is in harmony with the view that Mesarovic and others subscribe to, which is that a system is goal-seeking provided it has a clear objective.

Bovee (2004) presents a non-exhaustive summary of the work by Machlup and Mansfield (1983) (Table 1). Bovee (2004:11) states that 'the list details the implied specializations of the word "information" that occurred in the literature without adjectival modification'.

Table 1 Exclusionary narrowing of the term 'information' by Machlup and Mansfield (Bovee 2004:11)

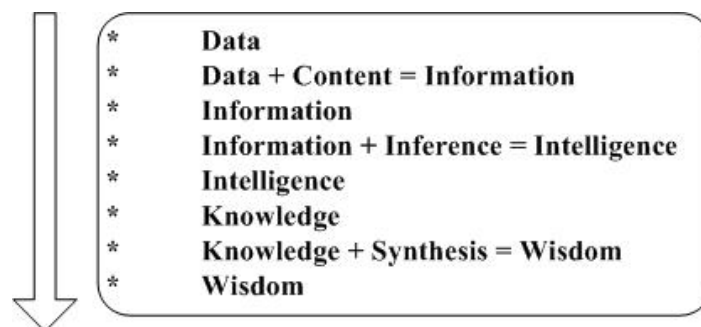| To be 'information' something must… |
| --- |
| |

Be previously unknown
Be previously less assuredly known
Affect the recipient' knowledge stock or structure
Consist of raw, uninterpreted data
Be useful in some way
Be used in decision making
Bear on contemplated, considered or taken actions
Reduce uncertainty
Help identify contextual meaning of words in sentences
Exclude some alternatives to what is predicted in a statement
Change some belief(s)

The fourth line from the top in Table 1 (which has been italicized – '*consist of raw, uninterpreted data*') is in conflict with Figure 1. Machlup (1983:648) proceeds and states that 'there is no need to establish either a hierarchy or a temporal sequence in discussing data and information. Apart from computer systems the two words may be equivalents.' As noted, not everyone agrees with the linear approach that data that have been processed becomes information and information that has been enhanced by experience becomes knowledge. This became clear during the literature research for this article. Some researchers claim that knowledge is also information, depending on the context.

Firestone and McElroy (2003) attempt to clarify distinctions between data, information and knowledge in the context of their research (their focus is knowledge management). They describe *data* as 'the value of an observable, measurable, or calculable attribute' (2003:17). In addition, they state that 'information is frequently data extracted, filtered or formatted in some way'(2003:18). However, Firestone and McElroy also subscribe to the school that data, information and knowledge do not need to be treated as a hierarchy or in a linear fashion. In spite of this, they share the same view as the IS/IT field as to what constitutes data, information and knowledge, but do not necessarily share the same view on the order in which they are produced.

In contrast to the argument in the paragraph above, however, and in agreement with Figure 1, the Barabba-Haeckel Framework is discussed by Barquin (2000). This framework is described as a continuum that starts with data and goes through stages until it is eventually wisdom. Figure 2 is used to illustrate this framework.

Figure 2 Barabba-Haeckel framework (Barquin 2000)

English's (1999) model of the relationship between data, information, knowledge and wisdom is in harmony with both Figures 1 and 2. English explains that useable data allow meaning to be extracted from them, resulting in information. This information in context allows one to determine its significance, which results in knowledge. The action based on knowledge results in wisdom.

Nevertheless, it was not the focus of this research to attempt to provide a 'perfect' understanding of the various views on what constitutes data, *information* and knowledge as this was beyond its scope. The purpose was merely to illustrate that there are different views on the subject. However, there is agreement that data do influence and have an impact on information. In this article it is argued that information does influence the decisions taken by decision makers and the results of these decisions affect the company and its performance (as shown in Figure 1). Tuomi (1999) adds that, when meaningful information is used in context, it becomes knowledge and is used to *make predictions*.

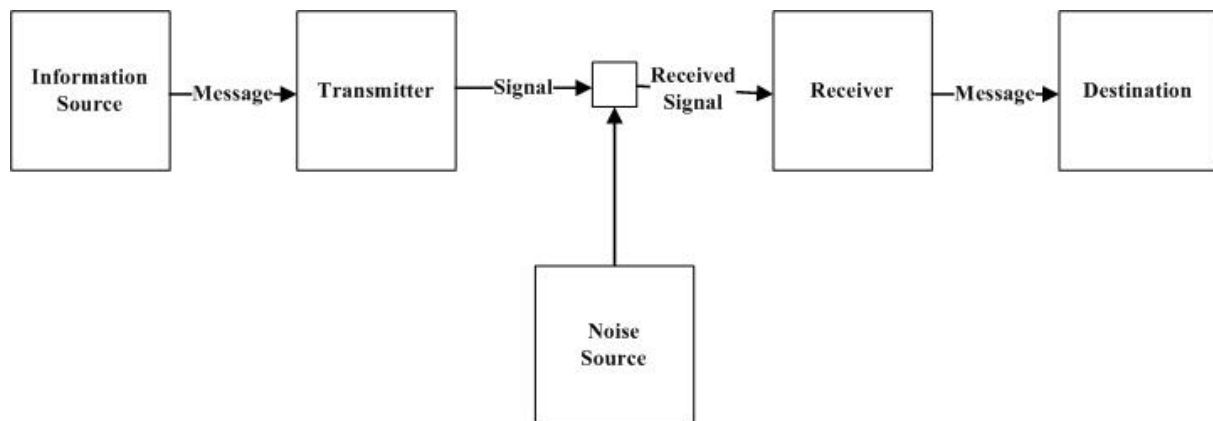**2.3 Information and communication**

It is beneficial to clearly link the information found within a company's financial statements to communication theory. Taking this view, which some may argue is conceptual, is however in line with Mesarovic's view described above. This is done to help reduce uncertainty between parties, which is important to emphasize, as uncertainty affects behaviour. The financial statements are viewed as a means of communication between the board of directors with the various company stakeholders.

Shannon and Weaver (1949) focused on communication and technology, specifically the sending and receiving of *messages*. The work of these researchers has formed the basis of many other research projects and papers. However, Shannon's previous work in 1948, which formed the core of the 1949 publication, is also of interest because it addresses the amount of uncertainty reduction that can occur when one receives a message from a *finite* set of possible messages. In addition, it is comparable to the directors sending messages to the various company stakeholders via financial statements on the financial condition of the company.

Shannon's theorem included a proposed graph and formulas arguing that the maximum possible *uncertainty reduction* or *entropy* occurs when both symbols (one representing uncertainty and one entropy) are equi-probable. In other words, when a system has two possible outcomes, the uncertainty or entropy in the system is maximized when both outcomes have the same probability. Information entropy or Shannon's entropy, as it is occasionally called, is an important point when one researches the 'condition' of information. However, for this article it is not necessary to duplicate too much of Shannon's work, but rather to focus on the work he contributed in the area of uncertainty reduction associated with the receiving of messages.

In Figure 3 there is a diagram illustrating Shannon's general communication system. Even though Shannon refers to a technical system, proving his theory by using a logarithmic approach, the principles he proves are applicable. As noted previously, there is a school that argues, for example, that Shannon's notion of information refers only to the capacity for transmission. However, applying information as described in Shannon's work together with Mesarovic's research, provides a wider view when one considers a goal-seeking system.

Figure 3 Schematic diagram of a general communication system (Shannon 1948)
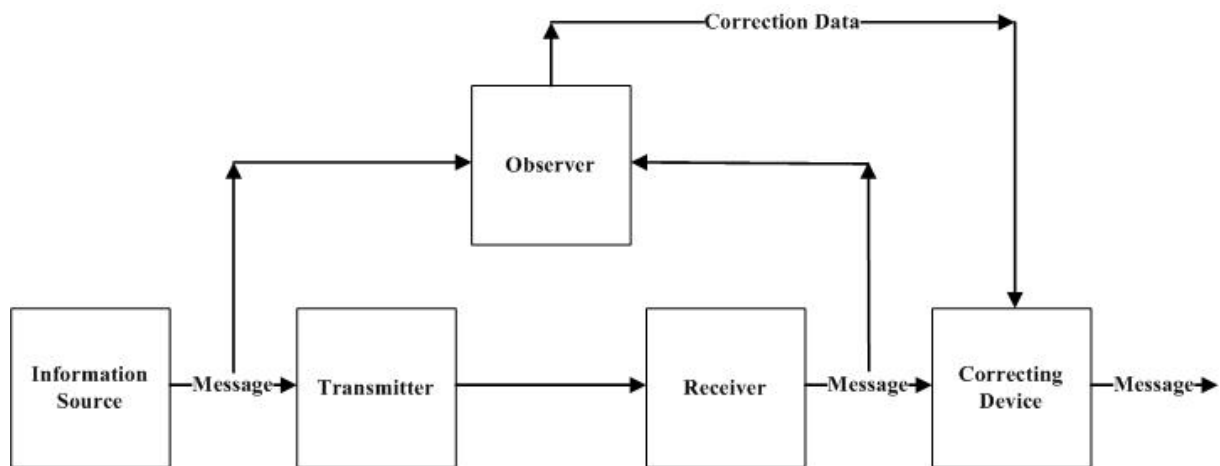

Shannon explains the parts of his diagram as follows: The *information source* produces a message or a sequence of messages to be communicated to the receiver. Then the *transmitter* operates on the message in some way so as to produce a signal suitable for transmission over the channel. The *channel* is merely the medium used to transmit the signal from the transmitter to the receiver. The *receiver* ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal. Finally, the *destination* is the person for whom the message is intended. Shannon also discusses the 'effect of noise in the channel'.

When one considers Figure 3 and contemplates on the board of directors sending messages via the financial statements (the *channel*), one can observe the ease at which the 'noise in the channel' can corrupt the message with either intentional or unintentional errors within the financial statements. The *information source* is the company producing a *message*, which is *transmitted* via financial statements to the *receiver*. The receiver could be any of the company stakeholders relying on the *message* found within the company's financial statements.

Shannon proposes a solution to help counteract the effect of the noise in the channel. This is illustrated in Figure 4 where 'an observer who can see both what is sent and what is recovered (with errors due to noise)' is put in place. 'This observer notes the errors in the recovered message and transmits the data to the receiving point over a "correction channel" to enable the receiver to correct the errors'(1948:21).

Figure 4 Schematic diagram of a correction system (Shannon 1948)

Such an observer, within the context of this research project, could be an auditor. It is important to note the change in the process (Figure 3 versus Figure 4), which allows the *observer* access to the message, both at the *information source* and after the *receiver* has received the message to 'make a comparison' and report on the *condition* of the information. The *correcting device* can be likened to a corrective control or controls that ensure the message is a *faithful representation* of the original message.

3 Concept of information quality

Information is an important commodity and it is not difficult to reach a point where one enters a stage of information overload due to the abundance of this commodity. Having said that, the management of information is a specialized discipline and decisions are based on the available information. Poor quality information has contributed to lost productivity, failed companies and low consumer confidence (English 1999; Wang and Strong 1996). Poor quality information has also caused political controversy and high-profile disasters (Fisher and Kingma 2001).

3.1 Quality decisions

Decisions are affected by information. If the information lacks quality, this has a natural effect on the outcomes of the decisions and, therefore, the decisions ultimately lack quality. To ensure fact-based decision making, one would require assurances as to the condition of the information when the decisions are made. If decisions are made in real time, based on real-time information, one would require real-time assurances.

It is therefore important to take note that 'high quality decisions are expected to lead to more productive actions, quicker problem solving, and better organizational performance' (Jung 2004:166). In other words, a director can perform his or her duties more effectively and the principles of good governance can be applied. Jung continues and points out that to make high quality decisions, it is crucial to have access to information that is relevant and complete on which to base decisions rather than just having an enormous quantity of information. Consequently, where decision makers experience information quality problems, companies can end up taking unnecessary risks by accepting impractical ideas and making errors in interpretation, or ignoring important ideas (Jung 2004).

Since risk is intrinsic to governance, a board of directors needs to ensure that the risks are mitigated to an acceptable level. This of course includes the risks associated with the board's decisions based on corporate financial information. Owing to many boards failing in this duty and the increase in corporate debacles, many influential 'regulations' have been imposed on companies enforcing that the information held within the company's information systems receive a higher priority.

Compliance to the Sarbanes-Oxley Act, Basel II Accord and other 'regulations' calls for heightened internal controls over the financial processes, implying an increased focus on IS/IT to provide a secure and auditable infrastructure. The current quality control standards, for example ISO 9001, concentrate on quantitative controls. These place the emphasis for process development on the controlling, documenting and monitoring of the work performed rather than the qualitative aspect. These are watershed regulations for companies as they mandate a rigorously controlled environment in which information systems operate and a high standard of *quality information* is created, documented and stored.

Accordingly, the philosophy of total quality management (TQM) is that continuous improvement must be applied to all quality standards. Thus, if one considers risk management, data and information, quality standards

are aspects to be considered. The reason is that poor information can cause a company to miss its strategic objectives due to the decisions based on that information. The company's reputation and the director's credibility could be severely damaged by the results of decisions based on poor quality information.

3.2 Data quality – the foundation of information quality

Ward and Peppard (2002) succinctly point out that the challenge companies have is to ensure that information is of the highest quality possible. As noted, information quality in part is based on data quality. By studying the research of Wang *et al.*, as Wang appears to be one of the foremost researchers in the IS data and information quality fields, it became evident that the attributes or dimensions of data and/or information *quality* are not agreed upon. However, a model of what is perceived to be the dimensions (as named by Wang) of data quality, is proposed by Wang and Strong (1996) and illustrated in Table 2.

Table 2 Data quality categories and dimensions (Wang and Strong, 1996)

Data quality category	Data quality dimensions
Intrinsic DQ	Accuracy, objectivity, believability, reputation
Accessibility DQ	Accessibility, access security
Contextual DQ	Relevancy, value-added, timeliness, completeness, amount of data
Representational DQ	Interpretability, ease of understanding, concise representation, consistent representation

Even though Table 2 assists in clarifying what constitutes data quality, it is inconclusive as there appear to be many different views on this topic. Furthermore, to gain a 'clearer' understanding of the attributes of information quality, a table (Table 3) summarizing the work of seven research groups was created for comparison. Table 3 is evidence that there appears to be disagreement on what constitutes information quality or at least disagreement on the use of words and terminology. Wang and Strong clearly present the largest number of attributes with both Bovee, Srivastava and Mak (2003), and ITGI (2004) with only four each.

Table 3 Terms used to describe the attributes of information quality

Information quality attributes	Ward & Peppard (2002)	Wang & Strong (1996)	Eckerson (2002)	Bovee *et al.* (2003)	Wand & Wang (1996)	ITGI (2004)	IASB & FASB (2006)
Accessibility		X	X	X			
Accuracy	X	X	X		X		
Appropriateness	X						
Believability		X					
Comparability							X
Completeness	X	X	X		X		
Concise		X					
Confidence	X						
Consistency		X	X		X		
Flexibility					X		
Integrity			X	X		X	
Interpretability		X		X			
Materiality							X
Objectivity		X					
Relevance		X		X		X	X
Reliability	X				X	X	X
Security		X					
Timeliness	X	X	X		X		
Understandability							X
Usability						X	

Validity			X				
Value-added		X					

When one studies Table 3, it becomes clearer that the words have semantic meanings and this appears to cause confusion. From observation, it does not appear that these researchers disagree in any major way as to what attributes constitute information quality. The disagreement lies more in the use of the words they have chosen and their terminology in general. For example, the words *accuracy* and *integrity,* which are recognized as crucial attributes of information quality, are used inconsistently. Only four of the seven research groups list accuracy as an attribute, as shown in Table 3. However, integrity is listed by three of the seven and *one* research group refers to both accuracy *(correctness, exactness, precision, truth)* and integrity *(truth, reliability, completeness, wholeness)*. There is a subtle difference in the meanings of these two words, but one can see how they are at times used interchangeably.

Wand and Wang (1996) conducted research on what constituted information quality and produced a table listing their results. *Accuracy* was cited more than any other attribute in importance. However, interestingly, their table is called 'Notable data quality dimensions' (1996:92) not *information* dimensions. Yet they conducted research as to what constitutes information quality, but substituted the word *data* for *information* when naming their table. Additionally, they did add that *'there is no exact definition for accuracy'* (1996:93). This may highlight why there are so many different terms used to describe information quality.

To continue with the example of *accuracy*, Bovee *et al*. (2003) emphasize that accurate information generally relates to whether or not the information corresponds sufficiently with its tangible or conceptual real world. If one accepts the definition of accuracy of Bovee *et al*., one can see why some researchers have used the word *integrity* in the place of *accuracy*.

An interesting output of Redman's (1998) research is that he argues that a typical impact caused by inaccurate data is that of 'increased organizational mistrust' (1998:82). Even though Wang and Strong (1996:32) do not refer directly to trust or mistrust in their table, they do refer to *reputation* along with *accuracy* as dimensions of Intrinsic DQ (Table 2). When defining reputation, they state that it is 'the extent to which data are trusted or highly regarded in terms of their source or content'.

Redman's research (1998) also claims that 1–5% of all data fields are erroneous. This is supported by a survey conducted in the USA in 2003 of accredited medical records managers, who found that 4–7% of their financial records had significant errors. Accordingly, these either resulted in over or under reimbursements of billing claims (Boritz 2005:261).

The information in Table 3 clearly illustrates that there is no agreement among researchers on the use of words and terminology describing the attributes (or dimensions) of data and information quality. However, regardless of the words used to describe information quality, Boritz (2005:262) highlights an important point that 'it would be hard to imagine information having quality in the absence of integrity'.
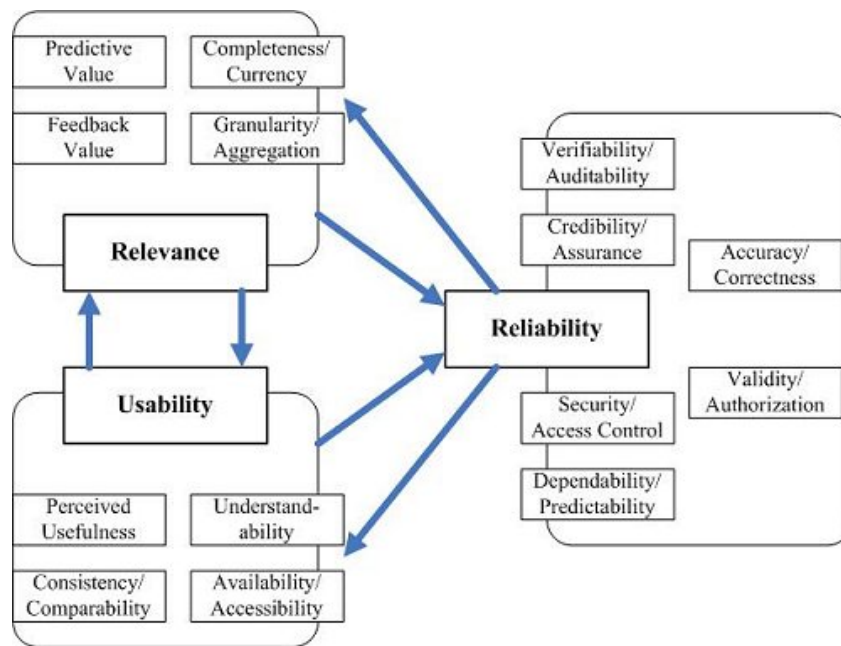
4 Information quality and the attribute of integrity

Boritz (2005) contends that information integrity is not an isolated attribute, but draws on several other attributes (or, as he refers to them, concepts). Boritz assisted the ITGI (2004) in one of the most extensive studies conducted in an attempt to clarify the understanding of information integrity. As noted in Table 3 and Figure 6, the IT Governance Institute (ITGI) presents four attributes *(reliability, relevance, usability, integrity)* of information quality. This research project subscribes to the ITGI view of information quality which appears holistic and thorough.

4.1 Attributes of relevance, usability and reliability

It is important to note that information quality is dependent on all four attributes. However, as noted, the 'Achilles heel' is the attribute of integrity. For how could the information have *relevance*, *usability* and be *reliable* if it lacks *integrity*? Nonetheless, it is also important to note that the remaining three attributes overlap and complement each other when determining if one has quality information. This is shown in Figure 5. Additionally, this figure also shows the sub-attributes of these three attributes and emphasizes that these are not mutually exclusive, but complement and contribute to each other (ITGI 2004).

Figure 5 Relationship among relevance, usability and reliability (ITGI, 2004, p.21)

Brief explanations or definitions are provided for *relevance*, *usability* and *reliability*.

- *Relevance* is referred to by four of the seven research groups in Table 3. According to the ITGI (2004:16), it is 'the information's capacity to make a difference that identifies it as relevant to a decision'. However, the ITGI points out that it is the 'theoretical capacity' of the information that contributes to the objective of information production.
- *Usability* or, as some have referred to it, usefulness, reflects the users' 'perceptions of the practical value of information they believe will help them in completing their work' (ITGI 2004:8).
- *Reliability* is also referred to by four of the seven research groups in Table 3. Accordingly, the ITGI (2004:20) explains that reliability 'reflects the signal-to-noise value of information used in decision-oriented systems', or expounded on, 'the less uncertainty and risk surrounding information' the more reliable the information.

Interestingly, the International Accounting Standards Board (IASB) and the Financial Accounting Standards Board (FASB, based in the USA) advocate that the word *reliability* is replaced with the words 'faithful representation'*,* as in their view the word reliability is widely misinterpreted (FASB 1 June 2005; IASB and FASB 15 May 2006).

4.2 Understanding integrity and its sub-attributes

According to the ITGI (2004), the external oval in Figure 6 encompasses information and the three circles within the oval illustrate attributes of information quality (as illustrated in Figure 5 which provides a more in-depth view of the three circles). The fourth attribute, which is the central oval, represents *information integrity*. The strategic placement of this oval illustrates the significance of integrity to the information's value.

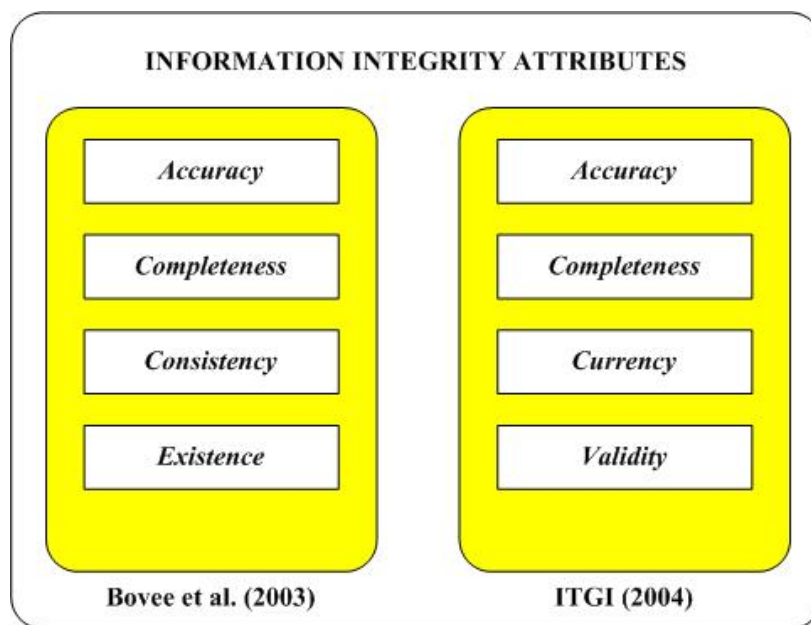**Figure 6** Information integrity (ITGI 2004:3)

As mentioned, it is important to observe the positioning of the information integrity oval in relation to the three

circles (relevance, usability and reliability), thereby overlapping and incorporating the elements of all three. It is also worth noting that information integrity is a narrower concept than information quality although it is a broader concept than data integrity (ITGI 2004).

In the same manner that relevance, usability and reliability have sub-attributes, so does integrity. As with the attributes of information quality there also appears to be disagreement among researchers as to what constitutes information integrity or, more specifically, what the sub-attributes of information integrity are. However, information has integrity if the accuracy, completeness, timeliness, validity and processing methods are safeguarded (ITGI 2004; Carlson 2001; NIST 800-12 Handbook 1995). According to the IT Governance Institute (ITGI 2004:22), integrity means unimpaired or unmarred condition. Applied to information, 'integrity is the representational faithfulness of the information to the condition or subject matter being represented by the information'.

Figure 7 is used to illustrate the sub-attributes of information integrity according to the ITGI (2004) and Bovee *et al*. (2003). These two research groups were chosen because their research appears extensive and thorough within this research domain. In addition their research is more current than most others.

Figure 7 Information integrity attributes

The research of Bovee *et al*. (2003) produced four attributes of information quality and related them to the process of *how* information is created (see Table 3). They describe three of these attributes (accessibility, interpretability and relevance) as *extrinsic* (*immaterial, insignificant, nonessential*) in nature. However, the fourth attribute, *integrity*, they claimed was *intrinsic* (*material, central, essential*) in nature when related to the process of how information is created.

Bovee *et al*. (2003) define the four sub-attributes of *integrity,* which is intrinsic to how information is made, in this way:

- *Accuracy* – This information conforms to the real-world or conceptual items of interest to the user. It is typically considered to be error free.
- *Completeness* – Refers to having all required parts or having enough information for decision-making.
- *Consistency* – Requires that multiple recordings of the values for any of the attributes be consistent across time and space. To be consistent, these values must be the same in all cases.
- *Existence* – This is an important intrinsic element of information used in auditing. If one needs to validate information, Bovee *et al*. (2003) claim that the information would need to meet any tests of existence that there are no false or redundant entities, fields or values.

The ITGI (2004) define the four sub-attributes of *integrity* this way:

- *Accuracy* – The information is a faithful representation of events.
- *Completeness* – 'All information necessary to reflect business activity in accordance with established business rules is captured, processed, stored and reported' (ITGI 2004:29).
- *Currency* – The information is current and timely and within preset definitions of the duration of time in

an information period.

- *Validity* – The information would be considered valid if it is authentic, not duplicated inappropriately, nonrepudiable, and in accordance with specific business rules that define relationships among information items, governing form, content, function, time, source and destination.
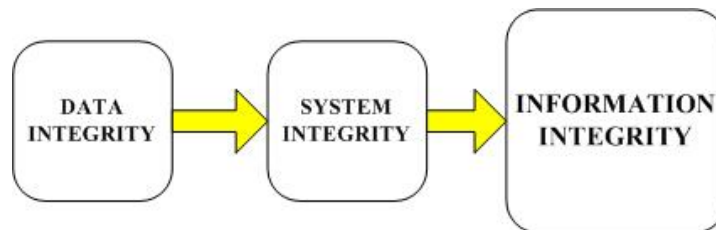
The first two sub-attributes, *accuracy* and *completeness*, are the same for both the ITGI and Bovee *et al*. even if defined slightly differently. However, it appears as if both research groups are saying the same thing. The third sub-attribute, *consistency* or *currency,* has overlapping qualities but a slightly different focus. The fourth sub-attribute, *existence* or *validity,* is not defined too dissimilarly. This once again highlights the semantic meanings of the words and the terminology favoured by the particular researchers.

Interestingly, a few of the other researchers in Table 3 listed some of these 'sub'-attributes or used similar words when describing information quality attributes. Nevertheless, Bovee *et al*. (2003), Boritz (2005) and the ITGI (2004) all emphasize the significance of the attribute of information integrity. Bovee *et al*. (2003:32) define the integrity of information as being 'satisfactory free from defects or flaws'.

4.3 Data integrity + system integrity = information integrity

Figure 8 graphically depicts the process of how information integrity is achieved. This demonstrates that to have information integrity both the *data* and the *system* (including IT infrastructure and operating system) need to have integrity. As discussed, data is considered to be the raw material used to create a finished product ready for use, that is, information. It is important to note that besides the *data*, information integrity is dependent on *system* integrity. In other words, information integrity can be no better than the integrity of the system processing the data or information, although it can be worse (ITGI 2004; Woodroof and Searcy 2001).

Figure 8 Requirements of information integrity

A system demonstrates processing integrity if 'its outputs fully and fairly reflect its inputs, and its processes are complete, timely, authorized and accurate' (ITGI 2004:5). To emphasize the two aspects (Figure 8), a system may have integrity, but if the data it processes lacks integrity at the time the system receives it, then the data can continue to lack integrity when it is transferred to its destination or transformed into information. Transmission integrity is therefore not treated as a separate element, but part of system integrity.

The following is a costly example of a processing error and its consequences when information lacks integrity. Owing to bank error in the currency exchange rate, an Australian was able to purchase Sri Lankan Rupees for AUS $104 500 and then sell them to another bank the next day for AUS $440 258. The original bank's computer displayed the central Pacific Franc rate in the Rupee position. Because of the circumstances surrounding the bank's error, a judge ruled that the Australian man had acted without intended fraud and could keep his windfall of AUS $335 758 (ITGI 2004:7). Another discomforting example (of which this is just one of many) is Fannie Mae's third quarter 2003 FAS 149 spreadsheet-based calculations that understated the value of the mortgage loan commitments by US $1.3 billion. Fannie Mae is a US company providing financial services, specifically mortgages. Fannie Mae attributed this to 'human error' (Boritz 2005:261).

These examples draw attention to the fact that both the data entering the system and the system processing the data needs to have integrity. It therefore becomes imperative to have controls in place to ensure this integrity. This stresses the importance of risk management, especially information security, which subsequently plays an important function ensuring information integrity.

5 Conclusion

In this article, the attributes of information integrity are reviewed and examined, as integrity is a problem companies grapple with today. Furthermore, there is a discussion on the linear sequence in which data processed becomes information and information interpreted becomes knowledge. It stresses that decisions are based on knowledge and that they determine the company's business results. This is followed by a rich debate clarifying

the attributes of *information quality* and the attributes of *information integrity*.

It proposes that information quality is essential to a company's success. However, the information cannot have quality if it does not have integrity. For the information to have integrity, both the *data* and the *system* need to have integrity. To have information integrity, a company needs to have a sound system of internal controls with IT controls at its core. Data and system integrity need to be ensured and the reason for this is that information is often found in electronic formats within the company's business processes. The controls need to limit uncertainty and the risks need to be mitigated to an acceptable level. This is especially true when considering a company's financial statements on which managers and stakeholders base decisions.

6 References

Barquin, R. 2000. From bits and bytes to knowledge management. [Online]. Available WWW: http://www.barquin.com/ (Accessed 9 May 2006).

Boritz, J.E. 2005. IS practitioners' views on core concepts of information integrity. *International Journal of Accounting Information Systems* 6:260-279.

Bovee, M.W. 2004. Information quality: a conceptual framework and empirical validation. [Online]. Available WWW: http://www.bsad.uvm.edu/Research/FacPubs/details?author=265 (Accessed 9 December 2005).

Bovee, M., Srivastava, R.P. and Mak, B. 2003. A conceptual framework and belief-function approach to assessing overall information quality. *International Journal of Intelligent Systems* 18(1):51-74.

Carlson, T. 2001. Information security management: understanding ISO 17799. Lucent Technologies Worldwide Services. [Online]. Available WWW: http://www.netbotz.com/library/ISO_17799.pdf (Accessed 1 February 2004).

Eckerson, W.W. 2002. Data quality and the bottom line: achieving business success through a commitment to high quality data. The Data Warehousing Institute. [Online]. Available WWW: http://www.dw-institute.com (Accessed 19 May 2006).

English, L.P. 1999. *Improving data warehouse and business information quality.* New York: John Wiley and Sons.

FASB. 2005. Minutes of the May 25, 2005 Board Meeting: conceptual framework. [Online]. Available WWW: http://www.fasb.org/project/conceptual_framework.shtml (Accessed 19 February 2006).

Firestone, J.M. and McElroy, M.W. 2003. *Key issues in the new knowledge management.* USA: Butterworth-Heninemann.

Fisher, C.W. and Kingma, B.R. 2001. Criticality of data quality as exemplified in two disasters. *Information & Management* 39(2):109-116.

Houck, T.P. 2003. *Why and how audits must change*. New Jersey: USA, John Wiley & Sons.

IASB and FASB. 2006 Conceptual framework – joint project of the IASB and FASB. [Online]. Available WWW: http://www.fasb.org/project/conceptual_framework.shtml (Accessed 9 June 2006).

ITGI. 2004. *Managing enterprise information integrity: security, control and audit issues.* USA: IT Governance Institute.

Jung, W. 2004. A review of research: an investigation of the impact of data quality on decision performance. In: *ACM International Conference Proceeding Series* 90:166-171.

Machlup, F. 1983. Semantic quirks in studies of information. In: Machlup, F. and Mansfield, U. (eds.). *The study of information: interdisciplinary messages*. New York: John Wiley & Sons.

Machlup, F. and Mansfield, U. 1983. *The study of information: interdisciplinary messages*. New York: John Wiley & Sons.

Mesarovic, M.D. 1983. Mathematical systems theory and information science. In: Machlup, F. and Mansfield,

U. (eds.). *The study of information: interdisciplinary messages*. New York: John Wiley & Sons.

NIST 800-12 Handbook. 1995. *An introduction to computer security.* National Institute of Standards and Technology. USA: US Department of Commerce.

O'Brien, J.A. 2000. *Introduction to information systems: essentials for the Internetworked enterprise* (9th ed). USA: McGraw-Hill Companies.

Oz, E. 2002. *Management information systems* (3rd ed). Canada: Course Technology Thomson Learning.

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

Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal.* [Online]. Available WWW: http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html (Accessed 11 November 2005).

Shannon, C.E. and Weaver, W. 1949. *The mathematical theory of communication*. Urbana IL, USA: University of Illinois Press.

Tuomi, I. 1999. Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organisational memory. *Journal of Management Information Systems* 16(3):1003-117.

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.

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

Ward, J. and Peppard, J. 2002. *Strategic planning for information systems.* England: John Wiley & Sons.

Woodroof, J. and Searcy, D. 2001. Continuous audit: model development and implementation within a debt covenant compliance domain. [Online]. Available WWW: http://raw.rutgers.edu/continuousauditing/ (Accessed 14 October 2004).