ELSEVIER

International Journal of Accounting Information Systems
6 (2005) 260–279

INTERNATIONAL JOURNAL OF
ACCOUNTING
INFORMATION
SYSTEMS

IS practitioners' views on core concepts of information integrity

J. Efrim Boritz *

University of Waterloo Centre for Information Systems Assurance, Canada

Received 30 September 2003; received in revised form 20 April 2005; accepted 1 July 2005

Abstract

Based on a review of the literature on data quality and information integrity, a framework was created that is broader than that provided in the widely recognized international control guideline COBIT [ISACA (Information Systems Audit and Control Association) COBIT (Control Objectives for Information Technology) 3rd edition. Rolling Meadows, Il: ISACA, 2000], but narrower than the concept of information quality discussed in the literature. Experienced IS practitioners' views on the following issues were gathered through a questionnaire administered during two workshops on information integrity held in Toronto and Chicago: definition of information integrity, core attributes and enablers of information integrity and their relative importance, relationship between information integrity attributes and enablers, practitioners' experience with impairments of information integrity for selected industries and data streams and their association with stages of information processing, major phases of the system acquisition/development life cycle, and key system components. One of the policy recommendations arising from the findings of this study is that the COBIT definition of information integrity should be reconsidered. Also, a two-layer framework of core attributes and enablers (identified in this study) should be considered.

Keywords: Information integrity; Data quality; Data integrity

1. Introduction

An entity's information assets constitute a significant proportion of an entity's market value (ITGI, 2001) making this a key enterprise asset that needs to be governed effectively.[1] Not only

* Tel.: +1 519 888 4567x5774.
E-mail address: jeboritz@watarts.uwaterloo.ca.

[1] Weill and Ross (2004) identify the following six key enterprise assets: Human, Financial, Physical, IP, Information and IT, and Relationships.

1467-0895/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.accinf.2005.07.001

are investors willing to pay for good governance–Newell and Wilson (2002) report premiums averaging 10–12% in market value when moving from worst to best on corporate governance–but effective governance is the single most important predictor of the value an organization generates from its information and IT asset (Weill and Ross, 2004). Information is increasingly easy to collect and digitize and is increasingly being incorporated in services and products. Information and information technology represent a significant expense in most enterprises; however, it is hard to value or price, has a decreasing half-life and has increasing risk exposure. Senior executives' accountability for the integrity of company financial information under the Sarbanes-Oxley Act of 2002 is a topic of great angst in the business community and a CICA publication aimed at Boards of Directors lists data integrity[2] as one of 20 key issues that Directors should be concerned with (CICA, 2002). The impact of information integrity impairments can be far-reaching and costly in money, time, resources, reputation and customers (Betts, 2001; Redman, 1998; Wang et al., 1995). Small mistakes made by the most well-meaning employee can have a catastrophic effect, propagating errors throughout the organization. For example, Fannie Mae's 3Q 2003 FAS 149 spreadsheet-based calculations understated the value of mortgage loan commitments by $1.3 billion. Fannie Mae attributed this to "human error".[3] A national survey in 2003 of all accredited U.S. medical records managers found that 4–7% of records (depending on region) had significant errors that resulted in over and under-reimbursement of billing claims.[4] Yet a 2001 survey (PricewaterhouseCoopers, 2001) found a dangerous complacency about data management: 2/3 of Boards do not address it; 2/3 place responsibility for it solely on the CIO or IT department; 1/2 of CEOs do not see it as a strategic issue; 1/3 of respondents believe management does not place enough importance on it; and, only 1/3 are very confident about the quality of their own data and even less are very confident about the quality of others' data.

To address information integrity impairments in an organized and rigorous manner requires a comprehensive framework that can be used to guide management risk assessments and control deployment and guide assurance providers on the criteria to be addressed by information integrity oriented assurance services. However, a limitation of today's control and assurance frameworks established by the accounting and auditing professions is that they have focused almost exclusively on financial information. As the focus of information integrity control and assurance efforts expands to other decision-related information beyond financial statements, a need arises for a comprehensive generally accepted definition of information integrity and a control framework linked to such a definition. Thus, one of the objectives of this study is to define and validate a general purpose framework that can be used for controlling and auditing information integrity.

The research approach used in this study involved three stages. First, an extensive review of literature in this area was conducted to identify key attributes of information integrity (ITGI, 2004) and related issues. Then, two focus groups of experienced practitioners were brought together to discuss the documented findings extracted from the literature review. Part of the

[2] The terms data quality, data integrity and information integrity are used in the literature, sometimes interchangeably, sometimes to convey different meanings. In this paper, information integrity is used except when referring to sources that refer to this concept using one of the other terms.

[3] http://www.oliviertravers.com/archives/2003/11/06/billions-hidden-in-spreadsheets/; http://www.louisepryor.com/showTopic.do?topic=41 [accessed December 29, 2004].

[4] Lorence, D.P. The Perils of Data Misreporting. *Communications of the ACM*, November 2003, 85–88.

process required the practitioners to complete a questionnaire addressing key aspects of the information integrity framework such as the definition of information integrity, core attributes and enablers of information integrity and their relative importance, relationship between information integrity attributes and enablers, practitioners' experience with impairments of information integrity for selected industries and data streams and their association with stages of information processing, major phases of the system acquisition/development life cycle, and key system components. Some preliminary results of this step are summarized in ITGI (2004, Appendix D). The third step involved analysing the participants' responses to identify statistically significant findings for inclusion in this paper.

The balance of this paper proceeds as follows. First, the information integrity framework is outlined. Next, the empirical phase of gathering information from practitioners using the questionnaire is described. Then, the findings are summarized, followed by a summary of limitations and brief concluding remarks.

2. Information integrity framework

Integrity means an unimpaired or unmarred condition—entire correspondence of a representation with an original condition (Webster's Third New International Dictionary, 1971). Applied to information, integrity is the representational faithfulness of the information to the condition or subject matter being represented by the information. The FASB's Concept Statement #2 (FASB, 1980) discusses representational faithfulness and identifies attributes of information without specifically linking them to the concept of information integrity. The definition of information integrity provided in ISACA's COBIT (ISACA, 2000) defines it by the three attributes of completeness, accuracy and validity. The CICA's ITCG (1998) includes additional attributes such as authorization, timeliness, consistency and segregation of incompatible functions. An extensive series of studies of data quality at MIT by Wang et al. (1993, 1995) and Wang and Strong (1996) has provided valuable insights into information users' views about data quality, which presumably includes data/information integrity. Table 1 summarizes the key attributes identified in these key sources.

Table 1 suggests that information integrity attributes relate to information reliability, relevance, usability, quality and value. In other words, the concept of information integrity draws on all of these concepts, but is narrower than information quality,[5] falling in the overlapping area of the three major information quality concepts of relevance, reliability and usability illustrated in Fig. 1 from ITGI (2004). This figure suggests that information integrity is the sine qua non of information quality as it would be hard to imagine information having quality in the absence of integrity.

The purpose of this study is to identify and validate a set of core concepts of information integrity to facilitate the development of 1) comprehensive management approaches for addressing information integrity concerns extending beyond financial statement considerations to operational and managerial information, 2) assurance services to provide assurance about all aspects of information integrity extending beyond assurance on financial statement information, and 3) research on causes of information integrity problems and potential solutions to those problems.

[5] For example, relevance includes attributes such as feedback value and predictive value that are not included in information integrity. Usability includes attributes such as perceived usefulness and ease of use.

Table 1
Comparison of core attributes of information integrity and their enablers in this study with key frameworks

ISACA COBIT	CICA ITCG	FASBSFAC#2	MIT research group	Core attributes and enablers in this study (see Table 3)
Integrity				
Complete	Complete	Part of representational faithfulness	Completeness	Complete
Accurate	Accurate	Precision/uncertainty	Accuracy	Accurate
Valid	Valid	Part of representational faithfulness	N/A	Valid
N/A	Authorized	N/A	N/A	Authorized
Effectiveness				
Relevant; pertinent	N/A	Relevant=predictive value, feedback value,	Relevancy; value added	Predictable/dependable
Timely	Timeliness	Timeliness		Current/timely
Correct	N/A	N/A		Correct
Comparable; consistent	Consistent	Consistent		Consistent/comparable/ standards-based
Usable	N/A	Decision usefulness		Understandable/ appropriate granularity and aggregation
Availability				
Available when required	N/A	N/A	Accessible	Available/accessible
Safeguarding (against tampering, loss, destruction)	N/A	N/A	Secure	Secure
N/A	Segregation of incompatible functions	N/A	N/A	
N/A	N/A	Verifiable; neutral	Objectivity	Verifiable/auditable
N/A	N/A	N/A	Believability; reputation	Credible/assured
Compliance				
With laws, regulations and contractual arrangements	N/A	N/A	N/A	N/A
Reliability				
Appropriate for management to operate the entity and to exercise financial and reporting responsibilities	N/A	Reliability= representational faithfulness, freedom from bias/neutrality, completeness, verifiability	N/A	See above
Efficiency				
Optimal use of resources	N/A	N/A		N/A
Confidentiality				
Protection against unauthorized (read) access	N/A	N/A		N/A

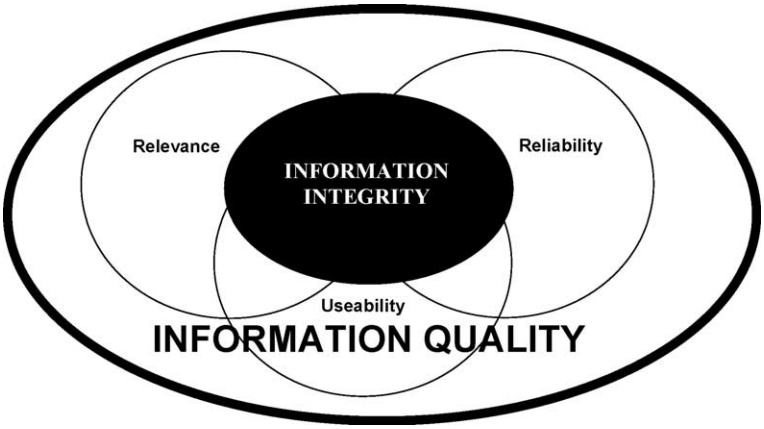

Fig. 1. Relationship between information integrity and other information quality concepts.

To these ends, this paper reports on the views of practitioners gathered through a questionnaire administered during two workshops held in Toronto and Chicago in Spring and Summer of 2003, respectively. The purpose of the questionnaire was to validate an information integrity framework that was developed pursuant to a review of literature conducted under the auspices of the Information Systems Audit and Control Association (ISACA).

An extensive review of literature in this area led to information integrity being defined as the representational faithfulness of information to the true state of the object that the information represents, where representational faithfulness is composed of four essential qualities or core attributes: completeness, currency/timeliness, accuracy/correctness and validity/authorization. These attributes are enabled by seven secondary clusters of attributes or factors, including: security, availability/accessibility, understandability/granularity/aggregation, consistency/comparability/standards, dependability/predictability, verifiability/auditability and credibility/assurance. Table 3 summarizes this framework. The core attributes are the rows of the table while the enablers are the columns. The cells formed by the rows and columns contain the impacts of the enablers on the core attributes, as discussed further in the following sections.

2.1. Distinction between core attributes and enablers

Core attributes of representational faithfulness are the minimum criteria that must be satisfied for a given information item or set to be judged as possessing representational faithfulness. In other words, all are necessary, but none are sufficient by themselves to warrant the label. In contrast, the enablers are not themselves characteristics of representational faithfulness, but they can help realize it. This distinction becomes clearer when representational faithfulness is viewed as a degree of achievement rather than an absolute quality. Because of the limits of information processing systems, perfect completeness, currency, accuracy and validity are not achievable. Thus, representational faithfulness is subject to some degree of imperfection, with the tolerable degree of imperfection being defined differently in different domains and contexts.

A numerical entry for a credit limit that is not authorized is not representationally faithful; a monthly sales figure that omits 1 week of sales but is otherwise accurate is not representationally faithful; and so forth. The enablers can help humans and software to assess the degree of representational faithfulness possessed by an information item or information set so that it can be brought within an acceptable range of imperfection. These issues are discussed further below.

2.1.1. Core attributes

2.1.1.1. Accuracy/correctness. Accuracy asserts that the information corresponds to reality (English, 1999); i.e., what is represented in the information system corresponds to a real world object or event with some degree of precision. For example, if the database states there are two cars on the lot but there are actually zero cars on the lot, then the database is inaccurate. If the database states that a policyholder is single, but she is married, then the database is inaccurate. While there are subtle distinctions between accuracy and correctness (i.e., an item can be accurate but not correct) these terms are considered as synonyms in this study. The concept of information accuracy is also linked to neutrality (lack of bias) in the way subject matter is represented. This concept is considered to be subsumed under verifiability/auditability.

2.1.1.2. Completeness. Accuracy by itself is insufficient to convey the full dimensionality of the requirements for representational faithfulness which requires completeness of information in both space and time. Thus, there is a fundamental trade-off between completeness and accuracy because measurement and processing limitations of information processing systems will prevent 100% real-time completeness, especially for subject matter that changes frequently. This, in turn, prevents 100% accuracy. For example, if there are three cars on the lot, two cars in the database, and one car in a receiving transaction that has yet to update the database, then a process that ensured processing completeness would contribute to database accuracy as well. In other words, every discussion of accuracy is also a discussion of completeness, and every discussion of completeness is also a discussion of accuracy. The degree of completeness that is achieved sets the upper bound on the degree of accuracy that is achievable.

2.1.1.3. Currency/timeliness. It must be accepted that absolute completeness and accuracy are impossible or impractical to achieve. The policyholder who was once single is now married. The auto dealer who had two cars available for sale now has none. Information currency is affected by real world changes over time (as well as by information processing delays) with a commensurate impact on information accuracy (Bolour et al., 1982). Since time is continuous, completeness and accuracy must be understood in a context that defines acceptable limits for information currency, hence its accuracy. For example, if certain information, such as cash receipts, is only used to update accounts receivable on a weekly basis, then accounts receivable could be considered accurate if it was missing a day's worth of transactions. However, if information, such as airline reservation transactions, is used to update available seat inventory in real time, then seat inventory would be considered unacceptably inaccurate if a day's worth of transactions were omitted.

As presented here, processing timeliness and information currency are really aspects of information completeness, which in turn, determines the degree of accuracy that information possesses; however, because of their unique relationship to the dimension of time and the change that time engenders, it is useful to identify currency/timeliness as separate attributes of information integrity.

Time Stamping: Given the foregoing discussion, it is important to recognize that completeness, currency and timeliness of processing are pre-requisites for a meaningful focus on information accuracy. There must be understood tolerances for information omissions and delays in the volume and timeliness of processing. Since the tolerances for information integrity may differ among stakeholders, it may be impractical to set standards for information currency or processing timeliness that meet the most stringent user requirements. Instead, various forms of

time stamping can provide useful information to enable stakeholders to assess the limitations of information integrity on this dimension. When information is enhanced by time stamping, its degree of accuracy is more understandable and more verifiable.

2.1.1.4. Validity/authorization. Representational faithfulness of information about intangible objects implies that the information is valid in ways other than correspondence with an original physical condition. The concept of validity means that information represents real conditions, rules or relationships rather than characteristics of physical objects. In a general context, conditions, rules or relationships are valid if what they purport is true. In a business context, conditions, business rules or relationships are established or approved by parties with the delegated authority to do so. Thus, transactions are valid if they were initiated and executed by personnel or systems that have been granted the authority to do so and if approvals are authentic and within the scope of the authority granted to the approver(s). For example, if the credit limit assigned to a customer reconciles to the company's rules and procedures used to set credit limits, the credit limit would be "valid." Thus, the concept of validity includes elements of both accuracy and authorization. A validation process may therefore require an investigation of an individual item, a relationship between one item and another item, or a relationship between an item and a business rule, policy or standard (Agmon and Ahituv, 1987).

2.1.2. Enablers—critical success factors

2.1.2.1. Security (as distinct from Confidentiality). Physical and logical access controls and safeguards over information in motion and at rest are required to protect information against acts of nature and intentional malicious acts such as unauthorized creation, modification, or destruction, as well as inadvertent errors that could compromise its integrity.

Another aspect of security involves protecting the confidentiality of information; that is, protecting it against unauthorized viewing or dissemination. While confidentiality is an important aspect of security, it is conceptually different from representational faithfulness. And, although some security controls serve to simultaneously protect information against threats to representational faithfulness and confidentiality, this study does not include the confidentiality objective of security in its use of the term as an enabler of representational faithfulness.

2.1.2.2. Availability/accessibility. For information to be complete, current and timely, it needs to be available and accessible to users in accordance with business specifications and to be retrievable in a usable form when required. Information that is not accessible when needed would not have any practical consequences for users' activities or decisions, except in the negative sense of limiting the quality of the information and users' decisions based on that information (O'Reilly, 1982). For information to be deemed accessible, users need to be able to work with the information in a way that meets their needs (O'Reilly, 1982; Wang and Strong, 1996). Practically, this requires the use of a robust system to provide the information. Such a system needs to be available when needed, enable the users to change the system (i.e., without programming changes) to meet their needs, operate efficiently and effectively, and be able to accommodate users' expanding need for information (Halloran et al., 1978).

Security and availability are complementary in the sense that security aims to restrict (unauthorized) access to information, whereas availability aims to facilitate (authorized) access to information.

2.1.2.3. Understandability/granularity/aggregation. Many factors can contribute to the understandability of information, including user knowledge, skill, training and motivation. In addition, information design choices such as its level of aggregation (or granularity) will affect its understandability, hence, its usefulness for controlling information integrity. For some purposes, highly aggregated information may be called for; whereas for other purposes, very detailed information may be required. Thus, appropriately tailored levels of granularity/aggregation can be enablers of information integrity. A proxy for the understandability of information is its conformity with user-specified requirements.

2.1.2.4. Consistency/comparability/standards. Consistency is the stability of measurement and presentation rules over time or space. Such rules represent standards against which information measurement and presentation can be compared and assessed. Environmental uncertainties perturb information systems and can lead to changes that can adversely affect stability and consistency and, hence, their comparability. Examples of such environmental factors include complexity (e.g., a system incorporates the use of new interfaces with external entities), change (e.g., regulatory changes), IT devices and computer crime (e.g., hacking) (Nayar, 1996).

2.1.2.5. Dependability/predictability. Several similar, but not identical, characteristics are grouped together under this heading, including: dependability, repeatability; stability and predictability. The dependability of information is facilitated by consistency in how information is measured and presented or displayed to decision makers (Kahn et al., 2002), the predictability of information processing and the predictability of the events that the information systems are designed to process information about. Events may be inherently unpredictable, but the information about them need not be. For example, a baseball player may not hit a home run each time at bat because athletic performance is unpredictable, but the information about the baseball player's performance may be dependable because there is a well-defined measurement protocol for observing and recording that information. In other cases, when the events themselves are predictable, then the representational faithfulness of information will be enabled by the predictability of the events. For example, the repeatable behaviour of ocean tides and currents enhances the accuracy of measurement compared with less frequent and less predictable events such as hurricanes.

To the extent that dependability is the result of a consistent measurement protocol or a predictable information process or system, the representational faithfulness of information will depend both on the reliability of the process and the reliability of related change management processes applied to the protocol, process or system (Halloran et al., 1978).

Elimination of information risks can contribute to stability, although it is important to recognize that some types of environmental uncertainty that may contribute to information risks are difficult or impossible to eliminate. Thus, improving the dependability of information processing systems will reduce information risks related to those processes but not the risks that arise from uncertainties related to the events themselves.

2.1.2.6. Verifiability/auditability. Verifiability is the ability of independent observers, applying the same processes and tolerances for completeness, currency, accuracy and validity that are used to produce the information, to replicate substantially the same result. In this study, the concept of verifiability is considered to subsume the concept of neutrality (freedom from bias). In order to verify and communicate information integrity to parties external to the information

process, the core components of integrity need to be complete, objective and measurable. This implies an approved or agreed upon set of processes or measurement rules, otherwise it would be difficult to obtain the measurement consensus that verifiability requires. In a business context, the approved set of processes or measurement rules springs from Board-approved policies and standards and any applicable legal, regulatory or professional requirements. Among other things, these must define the degree of tolerable imperfection in information integrity (in the assurance literature, also termed precision or tolerable error) for the core attributes of completeness, currency/timeliness, accuracy/correctness and validity/authorization, as further discussed in the section on the importance of context.

The intangible nature of information may prevent direct physical observation of information's integrity and, therefore, it may only be verified through an audit. In addition, parties external to the information may only be able to ascertain the integrity of information if they, or their agents, can audit the various facets of information.

Auditability features make it possible to trace information back to its source and confirm its representational faithfulness. Key auditability features include unique transaction/record identifiers such as a unique document or transaction identification number, creation date and modification date time stamps, a record of the document or transaction source and collection method, record retention and archiving, accessibility information and unambiguous and clearly documented re-computation rules (Winter and Huber, 2000).

2.1.2.7. Credibility/assurance. The intangibility of information may limit the ability of users to assess information to determine whether or not it has integrity (Wang et al., 1993; Richters and Dvorak, 1988). For information to be trusted to possess integrity, there must be evidence that it has been safeguarded against forgery or tampering by unauthorized parties (Winter and Huber, 2000). While verifiability/auditability represent necessary conditions for obtaining assurance about information integrity, credibility/assurance stem from procedures that are actually performed to verify/audit the integrity of the information by gathering evidence about its representational faithfulness.

Table 3 summarizes the relationship between the core attributes of information and their enablers. Table 1 compares the core attributes of information and enablers listed in Fig. 3 and Table 3 with those identified in COBIT, a leading control framework (ISACA, 2000) as well as the research conducted by the Total Data Quality project at MIT. An analysis of Table 1 indicates that currency, timeliness and authorization are not included in COBIT's definition of information integrity, although these concepts are included in other COBIT information criteria. Also, the relationship between several enablers of information integrity and core attributes of information integrity identified in this study are not explicitly considered in connection with the information criteria identified by COBIT. Specifically, enablers such as dependability/predictability, verifiability/auditability and credibility/assurance are not explicitly considered by COBIT as part of its discussion of information criteria, although other frameworks include these concepts.

As Fig. 2 illustrates, information integrity is enhanced by processing integrity. Indeed, the level of system processing integrity determines the upper limit of information integrity. System processing integrity, in turn, depends on system availability and system security. The information integrity attributes of completeness, currency, accuracy and validity flow from systems reliability attributes of processing integrity, availability and security. Availability contributes to the completeness, currency/timeliness and accuracy/correctness of processing while security contributes to all of these as well as validity/authorization. Security also

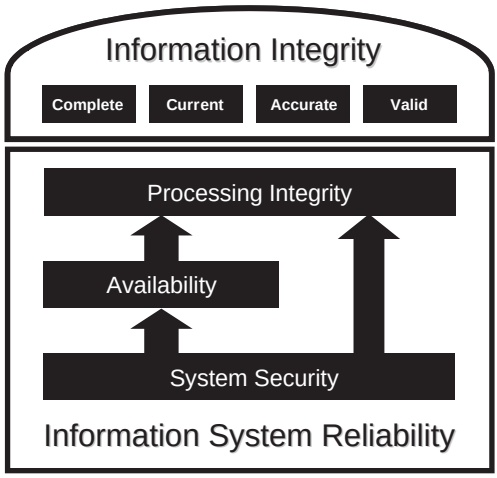

Fig. 2. Relationship between information integrity, processing integrity and system reliability.

contributes to Availability, reinforcing the point that security plays a critical role as an enabler of information integrity.

2.2. The importance of context

Information integrity attributes must be considered in the context of the stakeholders' specific requirements related to the information and the recognition that perfect information integrity is not achievable because completeness, currency, accuracy, and validity are affected by delays in data recognition, processing and utilization that, however small, introduce a degree of information impairment into all information processing functions. Thus, the standard for information integrity is not 100% representational faithfulness, but rather, representational faithfulness within accepted tolerances established in consultation with users of the information, parties responsible for maintaining the integrity of the information and assurance providers who are charged with confirming the integrity of information. The tolerances or materiality guidelines that are established must take into account the sensitivity of the information and the requirements of the user decision-models that are served by the information.

3. Method

3.1. Questionnaire

Since the core concepts described in the previous section were derived from a review of the professional literature, a questionnaire was administered to experienced IS practitioners to validate the importance of these concepts. The concepts were presented to two groups of practitioners using a workshop format for discussion and comment, but first, they were asked to complete a questionnaire to gather the following information:

1. Relative importance of information integrity attributes and enablers
2. Definition of information integrity

3. Definition of core attributes of information integrity
4. Relationship between information integrity attributes and enablers
5. Experience with information integrity impairments for selected industries
6. Experience with information integrity impairments for selected data streams
7. Information integrity impairments by stages of processing
8. Information integrity impairments by phases of the system acquisition/development life cycle
9. Information integrity impairments by system component.

3.2. Participants

Workshop participants were volunteers who responded to announcements distributed electronically by the Toronto and Chicago chapters of ISACA. Workshop participants were about equally drawn from Toronto and Chicago.[6]

The participants were experienced professionals with an average of 17 years of work experience and an average of 5 years in their current position. About 2/3 of the participants were male and 1/3 were female. The organizations represented were predominantly small to medium size entities; 3/4 had less than 10 000 employees. The most represented sector was the financial services sector, followed by consulting and healthcare. Information systems was the largest area represented, followed by audit. Almost half of the participants also had a formal information systems education as represented by their undergraduate degrees, followed by accounting and other management related fields. As might be expected, half of the participants possessed a CISA certificate, often in combination with other professional certifications. A demographic summary of the participants is provided in Table 2. All in all, the demographics for the participants indicated an experienced and knowledgeable group of professionals whose views about the attributes and enablers presented in this report should be carefully considered.

3.3. Format of the workshops

Pre-reading material was distributed as advance reading. When participants arrived at the workshop, a brief 30-min summary of this material was provided and questions about the purpose and structure of the workshop were answered. This took a further 30 min. Then the participants were asked to complete the questionnaire. This task was done individually by each participant and took about 60 min to complete. After the questionnaire was completed, the data were transcribed into a spreadsheet and displayed to all participants and a discussion ensued. Generally, the discussion centered around similarities and differences in patterns of responses by workshop participants to identify issues or problems with the concepts. Responses to the questionnaire were not changed except in one or two instances when errors were identified. Comments and observations made by the participants were captured by the researchers and are reported in this section of the report. This part of the workshop took approximately 60 min.

After a lunch break, the participants were divided into groups based on the transaction streams that they had self-selected when they were completing the questionnaire section

[6] There were only four significant between-city differences out of the 370 response variables collected; thus, the data from the two workshops were pooled for the analysis presented here.

Table 2
Summary of workshop participant demographics

City	
Toronto—April 8, 2003	15
Chicago—June 23, 2003	13
	28
Area	
Information Systems	11
IS/Management/Accounting/Production	4
Audit	9
Accounting/Finance	2
Sales/Marketing/Other	2
	28
Number of employees in your firm	
1–100	5
100–1000	7
1000–10 000	10
10 000–50 000	6
50 000–100 000	2
	28
Industry	
Financial Services	11
Consulting	8
Health care	4
Energy	2
Public sector	2
Telecommunications	1
	28
Gender	
Male	19
Female	9
	28
College major	
Information Systems	9
IS Management	2
IS Finance	1
Accounting	8
Economics	2
Management/Finance	3
Statistics/Other	2
	28
Graduate degree	
None	21
MBA	5
MS	1
MBA plus MSMIS	1
	28

(continued on next page)

Table 2 (*continued*)

Professional certificate	
None	11
CISSP	2
CISA only	3
CISA plus CA, CMA, CGA, CPA, CIA	12
	28

addressing the participants' experience with information integrity impairments. This part of the workshop took 3 h in total and consisted of several sessions devoted to group discussion and information sharing by all the groups. Upon completion of this part of the workshop, participants were thanked for their contribution and the workshop ended. Some of the raw data gathered during the workshops is reported in Appendix D of ITGI (2004) but did not include the statistical analyses reported here.

4. Analysis and summary of findings

4.1. Definition of information integrity

Information integrity was defined as the representational faithfulness of the information to the condition or subject matter being represented by the information. This definition reflects the dictionary meaning of integrity. About 75% (22 of 28) of the participants agreed with this definition. Those that did not agree had the following comments. One participant felt it is too close to the FASB conceptual framework. Another participant felt that faithfulness is a value-loaded term that can be interpreted subjectively; why not simply use the attributes to define information integrity rather than an overall term such as "representational faithfulness"?[7] Another participant suggested that there should be some qualification such as materiality or mention of a context when presenting the framework.

4.2. Core attributes of information integrity

Similarly, 75% (21 of 28) of the participants agreed with the definition of representational faithfulness using the core information attributes of completeness, currency, accuracy, and authorization. Those that did not had the following comments: the framework is too financially focused; there are too many attributes, only completeness and accuracy are required and they are less subjective than the other attributes;[8] there are too few attributes; some of the enablers should

[7] The rationale for the adoption of representational faithfulness as a synonym for information integrity is that integrity has come to have many meanings in common usage, and the term is often associated with honesty and truthfulness.

[8] The rationale for the addition of currency and validity to the core concepts of information integrity is that many practitioners may not see the time dimension that is implicit in completeness. Currency/timeliness of information is a very significant issue affecting the representational faithfulness of information so it needs to be reinforced. This was a major omission in the COBIT definition of information integrity which led to the omission of controls oriented towards the achievement of this attribute. Also, for many items integrity can only be determined by conformity with business rules. There is no real world physical reference point for establishing the representational faithfulness of much business information. For example, customer credit limits or a preferred supplier list do not describe physical realities of customers or suppliers; they represent business rules that have integrity if they are authorized and otherwise do not.

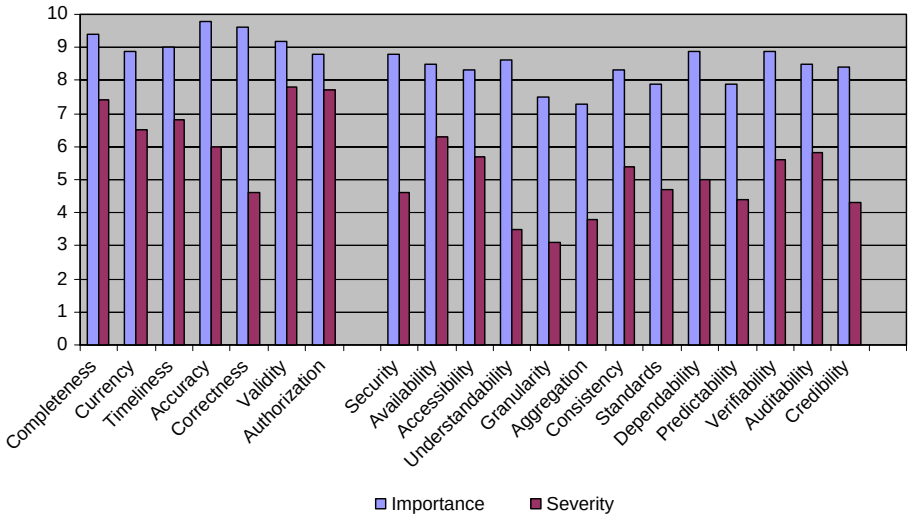

Fig. 3. Graphic summary of practitioners' importance ratings of attributes and enablers and practitioners' severity ratings of impairments experienced.

be included, particularly verifiable/auditable;[9] a context is required to define these attributes or make them objectively measurable.

4.3. Relative importance of information integrity attributes and enablers

The participants were asked to consider the information integrity attributes and enablers and rank them in importance from 0 (completely unimportant or irrelevant) to 10 (absolutely essential). Subsequently, they were asked to identify a data stream with which they had personal experience and rate the severity of observed information integrity impairments where 0 represents no impairment experienced and 10 represents extremely serious impairments exceeding 1% of gross revenues.

As summarized in Fig. 3, the concepts identified had high to very high ratings, indicating broad support for the framework components. All of the attributes and enablers were significantly different ($p < .05$) from 5, the midpoint of the scale, which could be considered to represent a neutral degree of importance.

The severity scores for the enablers were generally lower than those for the primary information integrity attributes. The severity ratings highlight the importance of validity/authorization. Interestingly, security had a lower impairment severity score than several other enablers. This could be due to the effective use of security controls in the organizations represented. The user-oriented enablers of understandability, granularity and aggregation were

[9] A question that can arise is how these attributes differ from financial statement assertions such as completeness, existence/occurrence, valuation/measurement, ownership/incidence and presentation. A key distinction is that financial statement assertions are assertions about an entity's assets, liabilities, equity, revenues, expenses, gains and losses. The attributes discussed in this paper are assertions about information. In this case, the integrity of the financial statements would be assessed against the criteria of completeness, currency/timeliness, accuracy/correctness and validity/authorization. And, the integrity of the financial statements would be enabled by security, availability/accessibility, understandability/granularity/aggregation/, consistency/comparability/standards, dependability/predictability, verifiability/auditability and credibility/assurance.

rated as being less important and being associated with lower severity ratings than all other enablers.

Comments were gathered from participants related to these items. Some participants questioned the inclusion of usability factors such as understandability and availability/accessibility in an information integrity model. Others agreed with including these concepts because they reflected the user dimension within the information integrity framework. Some participants questioned whether enablers such as granularity and aggregation related to information usefulness rather than integrity, and this is reflected in the comparatively lower ratings that these two items received. The rationale for including granularity and aggregation in the framework is that, in addition to affecting the usefulness of information, the absence of these enablers could detrimentally affect the functioning of decision-making and control processes, thereby affecting information integrity as well. Some participants questioned the value of the enabler "standards," while others defended its inclusion. There seemed to be a consensus that perhaps the enabler should be described as "enforced standards," since a number of participants questioned the value of standards if they were not enforced. Other participants observed that the mere existence of standards should improve information integrity as compared to the situation where there are no standards, even if there was no formal enforcement system. Some participants questioned whether predictability should be an enabler since it represented an inherent attribute of the information rather than an actionable item. However, others countered that an enabler does not need to be an action. Enablers can be inherent properties of the information and the environment.

4.4. Relationship between information integrity attributes and enablers

Participants were presented with a version of Table 3 (with all cells blank) and asked to consider the clusters of enablers listed in the columns and rate their importance to the attributes listed in the rows from 0 (completely unimportant or irrelevant) to 10 (absolutely essential). Table 3 summarizes the questionnaire responses. The pattern of participants' responses is broadly supportive of the anticipated relationship between the primary attributes and enablers. Table 3 indicates that only one cell (the cross between understandability/appropriate level of granularity/aggregation and validity/authorization) contains an average value of less than 5.

Table 3
Relationship between core attributes and enablers

		Secure	Available/accessible	Understandable/appropriate level of granularity/aggregation	Consistent/comparable/standards-based	Dependable/predictable	Verifiable/auditable	Credible/assured
Complete	Avg	7.1	7.9	7.1	7.9	7.5	7.9	7.8
	S.D.	3.8	3.2	3.6	3.3	3.4	3.2	3.1
Current/timely	Avg	5.5	8.8	5.2	6.2	7.3	6.5	6.7
	S.D.	4.0	2.1	3.9	3.9	3.4	3.4	3.6
Valid/authorized	Avg	8.7	5.1	4.1	6.8	6.6	8.5	7.8
	S.D.	3.1	4.2	4.2	3.3	3.9	2.7	3.6
Accurate/correct	Avg	7.5	6.3	6.3	7.9	8.4	8.7	8.9
	S.D.	3.6	4.2	2.9	3.3	3.1	2.6	2.1

Avg=average; S.D.=standard deviation of 28 responses by participants described in Table 2.

4.5. Information integrity impairments by industry

Participants fell into four industry groups: financial services, consulting, health care and other. Participants were asked to identify a data stream with which they had personal experience and relate the information integrity impairments to that stream. Participants were asked to rate the severity of the impairments where 0 represents not experienced and 10 represents extremely serious impairments exceeding 1% of gross revenues. It is interesting to note that there were no significant industry differences identified in the overall importance or severity rankings by industry except for the importance of accuracy, being significantly higher ($p < .05$) for participants in consulting, financial services and health care than those in the "other" category. This may be due to the small sample size. It is noteworthy, however, that a comparison of overall importance ratings with ratings of the severity of observed impairments by industry yielded significant differences between the two sets of ratings for every attribute and enabler under financial services. Generally speaking, the observed impairment ratings were significantly lower than the importance ratings.[10]

4.6. Information integrity impairments by data stream

As mentioned previously, participants were asked to identify a data stream with which they had personal experience and relate the information integrity impairments to that stream. Participants were asked to rate the severity of the impairments where 0 represents not experienced and 10 represents extremely serious impairments exceeding 1% of gross revenues. Participants' choices of data stream clustered around revenues, expenditures such as claims payments, management of customer account data and event capture involving shipping of goods or provision of services. It is interesting that no statistically significant differences were identified in respondents' reported levels of the severity of observed impairments for the various data streams. This may be due to the small sample size. It is noteworthy, however, that a comparison of overall importance ratings with severity of ratings by data stream yielded significant differences between the two sets of ratings for every attribute and enabler in the management of customer account data stream. Generally speaking, the observed impairment ratings were significantly lower than the importance ratings.[11]

4.7. Information integrity impairments by stages of processing

The following stages of processing were presented to participants who were asked to relate impairments to stages of processing,[12] on a scale from 0 to 10, where 0 is no relationship and 10 is an absolutely powerful relationship; *Input* (Data source/transaction initiation; Data collection, preparation and data entry; Data editing and validation); *Transmission* (Communications over public/private networks); *Processing* (Updates to databases, files and tables; Logic applications, computations, and analyses); *Storage* (Intermediate storage in

[10] The two sets of ratings may not be comparable in this way even though they are measured using a similar scale. However, if the scales were comparable, then the across-the-board pattern observed here indicates that severity of observed impairments was comparatively lower than the importance ratings of the attributes and enablers.

[11] Ibid.

[12] Input, processing, storage and output are widely recognized as key phases of processing; e.g., refer to Gelinas et al. (2004). Transmission was added because in network-based systems this phase represents a key information processing phase with special risk and control considerations that should not be overlooked.

databases or other logical storage devices; Back/up and recovery, including off-site storage); and *Output* (Output reporting, abstraction, and summarization; Use of output/Interface to other destination).

Pairwise comparisons between the means of stages of processing indicate 75 of a possible 200 significant differences ($p < .05$)—36 of 70 core attribute pairs and 39 of 130 enabler pairs.[13] The normal approximation to the binomial test for the significance of a proportion indicates that this is significant ($p < .05$). The findings suggest (consistent with the literature) that the input phase is a particularly significant source of impairments in core attributes and enablers. The storage and transmission phases are least associated with impairments.

4.8. Information integrity impairments by SDLC

The following stages of system development life cycle[14] were presented to participants who were asked to relate the impairments to stages of system acquisition/development, on a scale from 0 to 10, where 0 is no relationship and 10 is an absolutely powerful relationship: *Initiate* (Initial proposal, investigation, funding approval and planning; *Design* (Analysis of business function and user interface requirements; Initial conceptual design); *Build* (Detailed design; Acquisition/development (including unit and system testing); Implementation, deployment, acceptance testing, conversion); *Operate* (Operation; Monitoring, checkpoints, feedback loops); and *Maintain* (Maintenance and change management; Learning and improvement, abandonment or destruction).

Pairwise comparisons between means of the stages of processing indicate 47 of a possible 200 significant differences ($p < .05$)—12 of 70 core attribute pairs and 35 of 130 enabler pairs. The normal approximation to the binomial test for the significance of a proportion indicates that this is significant ($p < .05$). In particular, the initiation phase is quite different from the other phases of the SDLC; i.e., it is least likely to be associated with severe impairments in core attributes or enablers. The operate phase is a significant source of impairments associated with enablers. Interestingly, the maintenance phase does not appear to be an unusually important source of severe impairments.

4.9. Information integrity impairments by system component

The following system components[15] were presented to participants who were asked to relate the impairments to system components, on a scale from 0 to 10, where 0 is no relationship and 10 is an absolutely powerful relationship: *IT Infrastructure; Software; Human infrastructure; Procedures; and Data*.

[13] The base of 200 is calculated as follows: 10 pairwise comparisons of transaction processing phases × 20 attributes/enablers (7 attributes and 13 enablers).

[14] While there are numerous life cycle models most include these key phases. For example, CICA (1998) *Information Technology Control Guidelines* identify the following main phases: investigation; requirements analysis and initial design; development and system testing; conversion, implementation and post implementation review; and ongoing maintenance. Gelinas et al. (2004) identify the following main phases: analysis, design, implementation, and operation.

[15] This is based on AICPA/CICA (2001) and IFAC (2004) ISA 315 par 8. COBIT identifies the following components: facilities, technology, applications, people and data. While there is a similarity between these frameworks, the IFAC approach was viewed as more comprehensive. IT infrastructure encompasses COBIT's facilities and technology, software includes both systems and applications software and procedures represents an element not explicitly identified by the COBIT framework.

Pairwise comparisons between the means for system components indicate 33 out of a possible 200 significant differences ($p < .05$)—10 of 70 core attribute pairs and 23 of 130 enabler pairs. The normal approximation to the binomial test for the significance of a proportion indicates that this is significant ($p < .05$). The findings suggest that IT infrastructure is quite different from human infrastructure, software, procedures and data; i.e., it is least likely to be associated with severe impairments in core attributes or enablers. Understandability is an important differentiator between most components, with (lack of) understandability of procedures at the high end of association with impairment severity and (lack of) understandability of IT infrastructure at the low end.

5. Summary of findings, limitations and concluding remarks

Information quality is a key goal of effective corporate governance; however, Weill and Ross (2004) suggest that it is the least understood and most poorly utilized of the key enterprise assets. Information integrity is the sine qua non of information quality. This study posits that information integrity is synonymous with its representational faithfulness. Representational faithfulness is exhibited when information is: complete (within limits established by agreement, policy or regulation); current/timely (within limits established by agreement, policy or regulation); accurate/correct (within limits established by agreement, policy or regulation); and valid/authorized (in accordance with policies, standards and "business rules" established by top management and the Board and applicable laws and regulations established by regulatory agencies or legislative bodies).

In addition, the study finds support for a second layer of attributes represented by the following "enablers" for the core attributes of information integrity: Secure; Available/Accessible; Understandable/Appropriate level of granularity/aggregation; Consistent/Comparable/Standards; Dependable/Predictable; Verifiable/Auditable; Credible/Assured.

This definition is broader than that provided in COBIT (ISACA, 2000) which is a widely recognized international control guideline, but narrower than the concepts of information quality discussed in the literature. Also, the definition of information integrity given here is a broader concept than data integrity, since data is commonly considered to be a "raw material" that is used to create a "finished information product" ready for use by an internal user such as an employee or manager or external user such as a customer, supplier, analyst or regulator. Thus, one would expect a discussion of the attributes of information integrity to be somewhat broader than a discussion of data integrity and to consider the users of the information products. Interestingly, the user-oriented enablers understandability, granularity and aggregation were rated as being both less important and associated with less severe impairments than other enablers.

The questionnaire results provide strong support for both the core attributes and the enablers. For example, currency, timeliness, authorization and security are not included in the COBIT definition of information integrity, although these concepts are included in other COBIT information criteria. Also, enablers such as dependability/predictability, verifiability/auditability and credibility/assurance are not explicitly considered by COBIT in connection with information integrity criteria, but were highly associated by respondents with information integrity, particularly accuracy/correctness.

Interestingly, data stream and industry were not associated with significant differences in respondents' reported severity of observed impairments, although this could be due to the small sample size. However, phases of transaction processing, stages of system acquisition and development and system components were associated with impairments in significant ways. Of

the five transaction processing phases assessed by the respondents (input, transmission, processing, storage and output), the input phase had the highest level of impairments while storage and transmission had the lowest. Of the five phases of system development assessed by the respondents (initiate, design, build, operate and maintain) the system initiation phase had the lowest level of observed impairments while system operation had the highest.

Of the five system components assessed by the respondents (IT infrastructure, software, human infrastructure, procedures and data) the IT infrastructure had the lowest level of observed information integrity impairments. Understandability of procedures appears to play a key role in reducing the severity of observed impairments in the procedures component.

A study such as this one has a number of limitations which should be considered when interpreting the results. The sample of participants was small and they were self-selected. Thus, they may not be representative of the practitioner community. Hence, their ratings of attribute/enabler importance and observed severity of information integrity impairments may not be generalizeable. Also, the participants may have been biased to endorsing the concepts included in the pre-readings due to the research support provided by ISACA for this project. Also, the participants interacted with the author of the pre-reading materials during the workshop and may have been inclined towards being supportive rather than critical or challenging. On the other hand, the participants were experienced and qualified in IS assurance-related considerations. They had no incentives to support ideas that they did not agree with and were encouraged to be critical in the workshop sessions.

With these qualifications, the findings of this study suggest that the COBIT definition of information integrity be reconsidered. Also, a two-layer framework of core attributes and enablers should be considered. Moreover, measures aimed at improving information integrity should differentiate amongst the factors associated with observed impairments. For example, controls should consider the high risk associated with input, the moderate risk associated with processing and output and the low risk associated with the storage and transmission phases of transaction processing. Similarly, control strategies should consider the moderate risk of most stages and the low risk of the initiation phase. Also, control strategies should consider the moderate risks associated with most of the system components and the low risk associated with IT infrastructure.

Acknowledgements

Funding for this study was provided by the Information Systems Audit and Control Association (ISACA) and The University of Waterloo Centre for Information Systems Assurance. The views presented are solely the author's and do not necessarily reflect the views of ISACA. The research assistance provided by Malik Datardina is gratefully acknowledged.

References

AICPA/CICA SysTrust, Principles and Criteria for Systems Reliability, Version 2.0, New York/Toronto: American Institute of Certified Public Accountants and the Canadian Institute of Chartered Accountants, January 2001.

Agmon N, Ahituv N. Assessing data reliability in an information system. J Manage Inf Syst 1987;4(2):34–44.

Betts M. Dirty data: inaccurate data can ruin supply chains. Computerworld [December].

Bolour A, Anderson TL, Dekeyser LJ, Wong HKT. The role of time in information processing: a survey. SIGMOD 1982;Record 12(3):27–50.

CICA (Canadian Institute of Chartered Accountants). Information technology control guidelines. Toronto (ON): CICA; 1998.

CICA (Canadian Institute of Chartered Accountants). 20 questions directors should ask about IT. Toronto (ON): CICA; 2002.
English LP. Improving data warehouse and business information quality: methods for reducing costs and increasing profits. New York (NY): John Wiley & Sons, Inc.; 1999.
FASB (Financial Accounting Standards Board). Statement of financial accounting concepts No 2: qualitative characteristics of accounting information. Norwalk (CT): FASB; 1980.
Gelinas Jr UJ, Sutton S, Fedorowicz J. Business processes and information technology. 2004 [Thompson-South Western].
Halloran D, Manchester S, Moriarty J, Riley R, Rohrman J, Skramstad T. Systems development quality control. MIS Q 1978;2(4):1–13.
IFAC (International Federation of Accountants). Int Stand Audit (ISA) 2004;315.
ISACA (Information Systems Audit and Control Association). COBIT (Control Objectives for Information Technology). 3rd edition. Rolling Meadows (IL): ISACA; 2000.
ITGI (IT Governance Institute). IT governance executive summary; board briefing on IT governance. Rolling Meadows (IL): ITGI; 2001.
ITGI (IT Governance Institute). Managing enterprise information integrity. Rolling Meadows (IL): ITGI; 2004.
Kahn BK, Strong DM, Wang RW. Information quality benchmarks: product and service performance. Commun ACM 2002;45(4):184–92.
Nayar MK. A framework for achieving information integrity. IS Audit Control J 1996;2:30–4.
Newell R, Wilson G. A premium for good governance. McKinsey Q 2002;3;20–3.
O'Reilly III CA. Variations in decision makers' use of information sources: the impact of quality and accessibility of information. Acad Manage J 1982;25(4):756–71.
PricewaterhouseCoopers. Global data management survey. New York (NY): PricewaterhouseCoopers; 2001.
Redman T. The impact of poor data quality on the typical enterprise. Commun ACM 1998;41(2):79–82.
Richters JS, Dvorak CA. A framework for defining the quality of communications services. IEEE Commun Mag 1988;26(10):17–23.
Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. J Manage Inf Syst 1996; 12(4):5–34.
Wang RY, Reddy MP, Gupta A. An object-oriented implementation of quality data products. Workshop on information technologies and systems, p. 48–56.
Wang RY, Storey VC, Firth CP. A framework for analysis of data quality research. IEEE Trans Knowl Data Eng 1995;7(4):623–40.
Webster's third new international dictionary. Springfield (MA): G. & C. Merriam Co.; 1971.
Weill P, Ross JW. IT governance. Boston (MA): Harvard Business School Press; 2004.
Winter W, Huber L. Part 3: ensuring data integrity in electronic records. BioPharm 2000;13(3):45–9.