# Information Integrity Technology Product Structure

by

Vijay V. Mandke
Director and Professor
School of Engineering & Technology
Indira Gandhi National Open University
(EMPC Premises), IGNOU Campus, Maidan Garhi
New Delhi-110 068 (INDIA)

and

Madhavan K. Nayar
President
Unitech Systems, Inc
1240 E. Diehl Road, Suite 300
Naperville, Illinois 60563, USA.
E-mail : mnayar@unitechsys.com

## Abstract

Due to factors drawn from system environment, external to the application system and overlapping the user environment, networked computerized information systems of today have errors that are made but not corrected. As a result their is loss of integrity in networked computerized information system in terms of intrinsic integrity attributes of accuracy, consistency and reliability of information obtained. This calls for incorporating on-line learning and error correcting mechanisms in the IS models in the form of automatic feedback control systems with error detection and correcting technologies for improved information integrity of information systems – Information Integrity Technologies. The paper addresses this research issue of developing integrity technology product. Specifically, the paper begins with a critical look at research investigations in Information Integrity and then goes on to study the choice of IS model for integrity improvement. The paper then presents some alternatives for quantification of intrinsic integrity attributes and for developing integrity profile and cumulative information integrity index, followed by presentation of information integrity development steps. Finally, the paper gives a description of Information integrity technology product thus emerging as a software product and details it.

## 1. Introduction

From centrally located, batch-oriented systems of 30 years ago, computer systems have increasingly become networked, with applications increasingly sharing data with one another and databases becoming increasingly distributed. In the process the risk of data/information error – including risk of inherited error – in the networked computerized information system has increased, amounting to the issue of data/information pollution [9].

These errors are present at each stage of an information system, namely, data origin stage, communication channel prior to processing stage, processing stage, communication channel at post-processing stage and output stage and are caused by factors not amenable to controls including application controls conceived at system design stage. Literature

reports research efforts in terms of identifying foolproof information requirements [6, 11, 19], but design experience shows this is something not easy to achieve.

These factors responsible for errors invariably have their presence mainly through the system environment which is external to computing (and hence the application) system and overlaps the user environment, though together they (the computing system and its external environment) constitute the information system. Inspite of application controls, it is these external factors that then make information systems give rise to information which is inaccurate, inconsistent and unreliable [12].

## 2. A Critical Look at Research Investigations in Information Integrity

With military requirements dominating the research in information systems, the issue of secured computer systems and of confidentiality of information has always been a high priority query. As a result, for over twenty-five years, there have been efforts to work on information security programmes. Further, security has normally been taken to mean confidentiality, integrity and availability [16, 4], where the meaning of word "integrity" is not adequately resolved.

### Integrity and Security are different, equally important and require different mechanisms

However, a critical look at research investigations in the area of information integrity points out that security meaning controlling dissemination of information (confidentiality) and integrity meaning validity of information in a computer system – requiring control over modifications made to information (also termed as correctness of information), are different [2, 3]. Further separate policies are required for confidentiality and (data) integrity, and that, with the exception of common requirements such as user authentication, much of the mechanism for supporting these two – security and (data) integrity – policies are different [13]. There is also an appreciation that there are two different notions of "data integrity": (i) one of "integrity" concerned with internal correctness of a system (consistency) and (ii) other of "appropriateness" concerned with correspondence with the real world (reliability) [18].

Coming to the networked information systems, telecommunication networks are integral to a computer based information system. From this angle, research investigations reported concede that integrity is equally important, if not more, as security (taken to mean confidentiality). Here integrity is primarily considered as a property of a system that provides assurance as to the accuracy, faithfulness, non-corruptibility and credibility of information transmitted between source and destination entities [20].

Finally, in the context of networked information system, while considering integrity issues in respect of networks, it is further realized that the unintentional but largely inevitable threat to transmitted information occurs through noise in communication systems and equipment failure. As a result, the mechanisms used in support of integrity policy requirements may need to be probabilistic [13, 20].

Of course, in all above research investigations, the concept of Trusted Computing Base (TCB) meaning thereby trusted computer base, trusted procedures, trusted processing, authenticated procedures and audit trails and segregation of duties is considered fundamental to security and integrity.

### Errors in Information Systems : Their Causes and Integrity Implications

In their study of errors in information system, Mandke and Nayar [9, 10] describe errors and thereby loss of integrity at each stage of the information system and, as a result, the loss of overall system integrity. Specifically Mandke and Nayar argue that there are errors at each stage in the information system and these errors are due to factors present mainly through the system environment. These factors represented by "5Cs" are **change** (in the content or in configuration of the system environment), **complexity** (due to introduction of new component, be it a programme,

database or network, thereby adding new interfaces), **communication** (i.e., movement of data/information within or across enterprises), **conversion** (meaning consolidation, decomposition or transformation of data) and **corruption** [refers to human behaviour – poor motivation, desire for personal gain, carelessness, actions of people; to factors leading to inherited errors polluting the information system (inherited error occurs when an error is propagated beyond the system in which it originated); to unpredictability (noise) of any kind, e.g., communication channel noise, equipment failure, etc].

Whether in addition to application controls, computerized information systems also incorporate human engineering design criteria at the system design stage itself or hardware and software vendors further incorporate error-checking filters into their products, it is these factors, external to application system, that render application controls inadequate, resulting in presence of errors in information system that are made but not corrected.

And it is these errors that have integrity implications at each stage in the IS and at the system integrity level. Specifically, as shown by Mandke and Nayar [9, 10], these integrity implications are in terms of requirements of integrity attributes of : accuracy, completeness, timeliness (implying accuracy inspite of time related changes in data/information), consistency (satisfying domains and constraints), reliability (accuracy with which information item represents data item in whichever way information system processed it), security (confidentiality) and privacy. Further, irrespective of the nature of use of the information obtained from the information system, attributes of accuracy, completeness, timeliness, consistency and reliability emerge as integrity attributes that an information system must satisfy, while attributes of security and privacy are optional depending on the context and nature of use. These optional attributes, therefore, can be seen as extrinsic or subjective information integrity attributes specific to areas of use [8]. Other such subjective attributes of integrity could be : interpretability, ease of understanding, traceability, cost effectiveness, flexibility, etc. [21].

## Intrinsic Integrity attributes IS should satisfy

Research investigations point out there is more to the integrity attributes mentioned above. To explain, attributes of completeness and timeliness are necessary for accuracy. That is to say, when checked for accuracy, information item also gets checked for its completeness and for its being up-to-date (timeliness), as accurate information has to be complete and timely. In that sense, it is sufficient to check for accuracy only.

Similar is the situation in respect of consistency, too, as an accurate value also has to be consistent. However, difference is that consistency check is in terms of domain values and in terms of constraints without refering to real-world objects and, therefore, a simpler and less expensive task offering first approximation of accuracy and, when checked in addition to accuracy, increasing overall reliability of integrity checking process itself.

It is within above framework then **accuracy** (includes completeness and timeliness), **consistency** (satisfying domains and constraints) and **reliability** (accuracy with which information item represents data item in whatever way information system processed it) emerge as intrinsic or basic or objective information integrity attributes and offer a precise and agreeable definition of Information Integrity [8, 9, 10].

## Inadequacy of assumption of Trusted Computing Base

There is yet another aspect and that pertains to the assumption of concept of TCB mentioned earlier. In this regard it is observed that Report of IFIP Working Group 11.5 [7] points out that this is a narrow view of what constitutes integrity and it is confined within the logical bounds of an information system as it excludes the material impact of the people and business application processing necessarily involved in any system. In fact, the report goes to state that "the concept that a system can be trusted over time without the ability to provide the evidence that the trust is well placed is incompatible with internal control principles. The concept of trust therefore is insufficient for our purposes".

## Need for Automatic Feedback Control System for On-line Error Detection and Integrity Improvement in Information Systems

It is within the framework of above research investigations that Mandke and Nayar [9, 10] have proposed need to incorporate on-line learning and error correcting mechanisms in the IS models. Specifically, to account for errors in IS that are made but not corrected, they propose incorporation of automatic feedback control systems with error detection and correcting technologies for improved information accuracy, consistency and reliability; technologies that maximize integrity of information systems – Information Integrity Technologies. They further argue that, when incorporated, it is such Information Integrity Technology that would also facilitate demonstrating improved integrity of information obtained, rather than merely trusting the computerized information systems.

There are obvious difficulties in designing and developing such automatic feedback control systems, the most important being, to study error patterns, it is not possible to track and analyse every bit of data/information for all times as it flows through the information system stages. Way out here is to consider Information Integrity Technology that takes a sample of input data at the output or at an intermediate point of an appropriately identified stage or sub-system of the IS and then follows or keeps track of the sampled records at output or intermediate points of subsequent stages (sub-system), at a given point of time or at different points of time over a required time internal [9, 10].

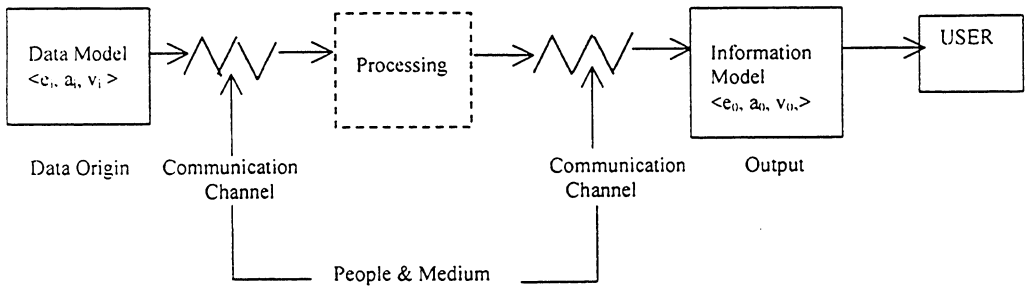
## 3. Information Integrity Attribute Quantifiers

This brings forth the central question as to what will be the structure of such an Information Integrity Technology Product. To answer this and particularly to suggest ability of such a product to demonstrate integrity improvement in information obtained, it is first necessary to consider the question of quantification of intrinsic integrity attributes of accuracy, consistency and reliability and of overall system integrity. Towards this the investigation at hand may first define the Information System (IS) model.

### 3.1 Choice of IS Model : A Basis for Integrity Quantifiers

Specifically, Networked Computerized Information Systems of today see "Data" as raw material, "Data Product or Information" as processed data used to trigger certain management action, "Processing" as the system function, and are characterized by (a) computing processes that include micro-computer and telecommunication and (b) pre- and post-processing stage communication channels at various data/information processing nodes, that are people based and include data communication and transaction processing networks with world-wide reach. Such decentralized structure of IS has certainly facilitated organizations and individuals to work with shared data environments and with capture, use and control of growing, complex and diversified volumes of data and information; in turn affording business access to bigger markets.

Such Information System can be modeled as given in Figure 1 where $< e_i, a_i, v_i >$ denotes a triple for Data Model (input to the information system) and $< e_o, a_o, v_o >$ denotes a triple for Information Model (output from the information system); < e, a, v > representing datum a triple <entity, attribute, value> as developed by the database research community. This representation which permits treating data/information as formal organized collection allows to segment integrity issue into issues concerning entities, attributes and values thereby making it feasible to study IS integrity analytically.




Where $< e_i, a_i, v_i >$ denotes a triple for Data Model and $< e_o, a_o, v_o >$ for Information Model

**Figure 1 : Conceptual Presentation of an Information System Model**

In the Information System modeled as above there are errors that are made but not corrected. When abstracted this implies, in the Information System Model in Figure 1, given that data/information is represented by a triple <e, a, v> and considering a particular example where say an output, i.e., processed data, i.e., information is represented by Entity Class, namely, employees and where specific entity (e) under consideration is an employee by name Albert and where specific attribute (a) under consideration is Albert's Salary, then, by virtue of on-line errors present in the information system, at any time, there exists a possibility of information item on value (v) of Albert's salary being inaccurate, inconsistent or unreliable, i.e. it's being affected by error or say corrupted by noise, and, therefore, a more realistic representation of value (v) is $(v + \eta)$, where $\eta$ represents noise or error component [9, 10].

It is within this framework of error implications on data/information model wherein triple $< e, a, v >$ is replaced by triple $< e, a, v + \eta >$ and, as discussed in section 1 considering that these error implications are present at each stage of an information system; namely, data origin stage, communication channel prior to processing stage, processing stage, communication channel at post-processing stage and output stage, a modified version of a conceptual schematic of an Information System Model in Figure 1 emerges, accounting for errors that are made but not corrected. The same is given in Figure 2 below.

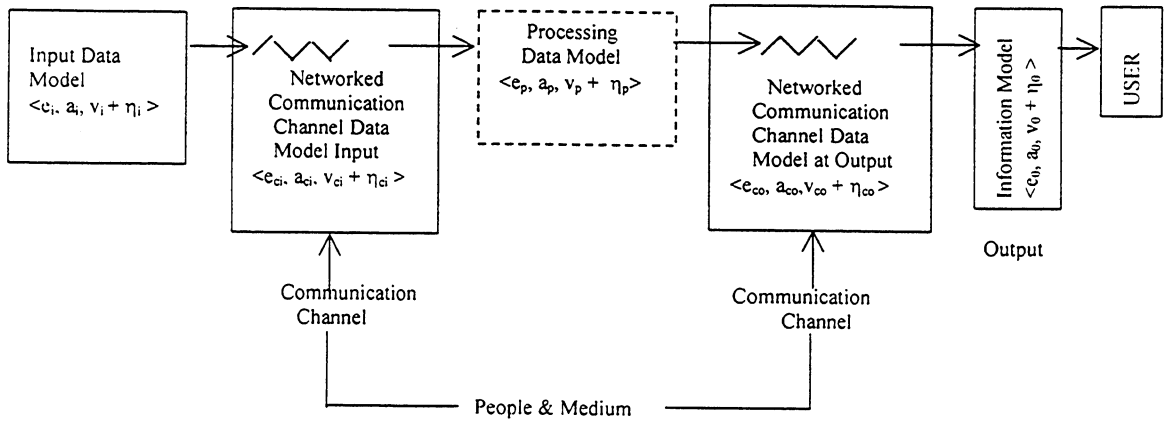



**Figure 2 : Modified Conceptual Presentation of an Information System Model in Figure 1 accounting for errors that are made but not corrected**

It is these errors that have integrity implications at each stage of the IS and at the overall system level and same are shown in Figure 3 below.

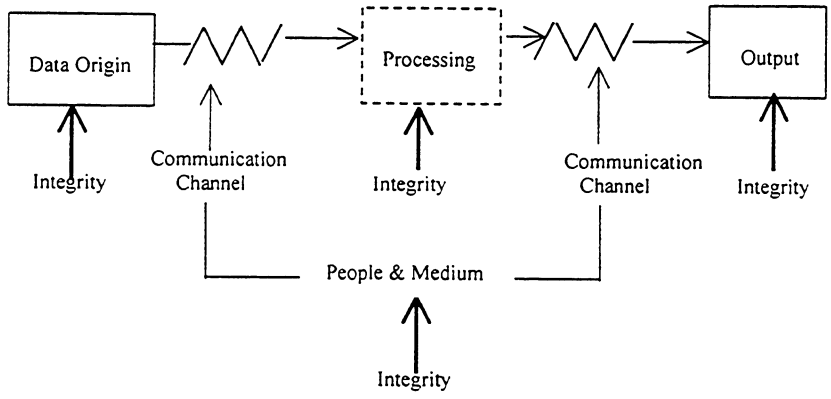



**Figure 3 : Conceptual Presentation of Integrity of an Information System**

What is important for the investigation at hand is that integrity of the overall information system is ensured if the integrity requirements of all parts of the system as in Figure 3 are ensured [14]; integrity being defined in terms of attributes of accuracy, consistency and reliability whose quantification being the query to be pursued. In what follows this section address this query [9].

### 3.2 Accuracy

Accuracy refers to correctness, i.e., preventing unauthorized modification, i.e., degree of conformance between a particular value of data/information and an identified source. The identified source provides the correct value [5]. It can be an object or relationship in the real world; it can also be the same value in another database, or the result of a computational algorithm.

Given that value of data/information is expressed in a numerical, accuracy of the data/information can be quantified in a number of ways [15, 5, 17, 1]:

i) Difference between the actual value (i.e., value of the identified source) and the value processed by the information system.

ii) $\text{Error Ratio} = \dfrac{\text{Actual Error}}{\text{Acceptable Error}}$

iii) $\text{Accuracy Index} = \dfrac{\text{Number of correct values}}{\text{Number of total values}}$

iv) Number of records examined : R

Number of records with atleast
one defect of loss of Accuracy : D1

$$\text{Percent Defective} = \left[\frac{D1}{R} \times 100\right]$$

$$\text{Accuracy Index (A)} = \left[1 - \left(\frac{D1}{R}\right)\right]$$

**Note :** Percent Defective is a quantifier used extensively in statistical quality control.

v) Number of defects (cases of loss of accuracy) detected : D

Number of records examined : R

$$\text{Defects/Losses of accuracy per record} = \frac{D}{R}$$

$$\text{Accuracy Index (A)} = \left[1 - \left(\frac{D}{R}\right)\right]$$

It may be mentioned that defect denotes accuracy violation, i.e., presence of error, and hence the absence of accuracy. Ratios based on defects/errors can be converted into accuracy ratio by the transformation:

Accuracy Ratio = 1 – Defect (i.e., Error) Ratio.

Understandably notion of accuracy quantified as above has many issues not considered here. What if correct value of the identified source is undefined, or simply unknown. And of course what if data/information is say a name or has an alphanumeric value or is a video image; how is error or defect defined then ?

### 3.3 Consistency

Consistency is with respect to a set of constraints. As pointed out earlier, data/information is said to be consistent with respect to a set of constraints if it satisfies all constraints of the data/information model [5]. Constraints can apply to the same attributes in different entities (such as the salary attribute in the entities of several employees); they can also apply to different attributes in the same entity (such as the salary level and salary attributes in the entity for a particular employee).

Given the number of constraints specified (CS) and given the number of constraints for which error/defect detected in the sense constraints are not satisfied (CE), then consistency can be quantified as follows [17]:

$$\text{Consistency}\,(C) = \left[\, 1 - \left(\frac{CE}{CS}\right) \,\right]$$

### 3.4 Reliability

Finally, as mentioned in section 2, Reliability (R) may be considered as an accuracy with which the information obtained represents the data item in whatever respect the information system processed it. For this purpose, a model may be considered where any processing of data has a large error component, random in nature. As a result volume of error in the processed data will be different each time the data processing is repeated, leading to significantly different information in each case; thus reflecting a low reliability of the information. Thus 'Reliability' refers to the extent of existence of random errors in an information, or in other words, the degree of consistency with which an information can be repeated, without any intervening or additional instruction.

Coming to the quantification of reliability (R), in any data/information model, for an entity (i), the value ($v_i$) for an attribute or processed value for ith data item for the entity may be expressed as $v_i = t_i + e_i$ , where '$t_i$' is the true component of the value and '$e_i$' is the error component. It is assumed that :

(a) $v_i$ takes values on a real line,

(b) $e_i$'s are distributed independently and randomly over the whole population of data items (i's) and that $e_i = 0$, and

(c) $e_i$'s are uncorrelated with $t_i$'s.

Then reliability 'R' is given by :

$$R = 1 - \frac{V_e}{V_v}$$

where

$$V_v = \frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v}_i)^2$$

is the variance of the processed value and

$$V_e = \frac{1}{N}\sum_{i=1}^{N}(e_i - \bar{e}_i)^2$$

is the variance of the error component.

From above it follows that reliability "R", also termed as "Coefficient of Reliability" or "Reliability Index", will have a value between 0–1.

It is appreciated that it may not be possible to repeat every data processing. In such case internal consistency of a data/information set comprising (a) information from processed data, (b) information from relevant identified source, (c) information from another related database, (d) results from relevant computational algorithm, etc. could be studied to obtain the reliability.

Various methods exist for calculating the Reliability Index (R); Analysis of Variance (AOV) technique being one such. Choice of a method would depend on advantages, disadvantages and convenience of application in a given situation, while accounting for factors like nature of available data, form of data and computation aids available for processing.

### 3.5 Integrity Profile

Consider an information system designed and developed for an application area. It is appreciated that each application area, consistent with information usage requirements, will have application area specific order of significant for integrity attributes. Let $W_r$ represent significant weightages for the integrity attributes accuracy, consistency and reliability, respectively, for the application area under consideration. These weightages may take values between [0-10].

Consider a using the above information system for the application at hand. Let the Information Integrity attribute indices as observed at the user end in this specific example be : Accuracy (A) = 0.78, Consistency (C) = 0.55 and Reliability (R) = 0.85. Then Information Integrity Profile from the user end can be represented as follows :

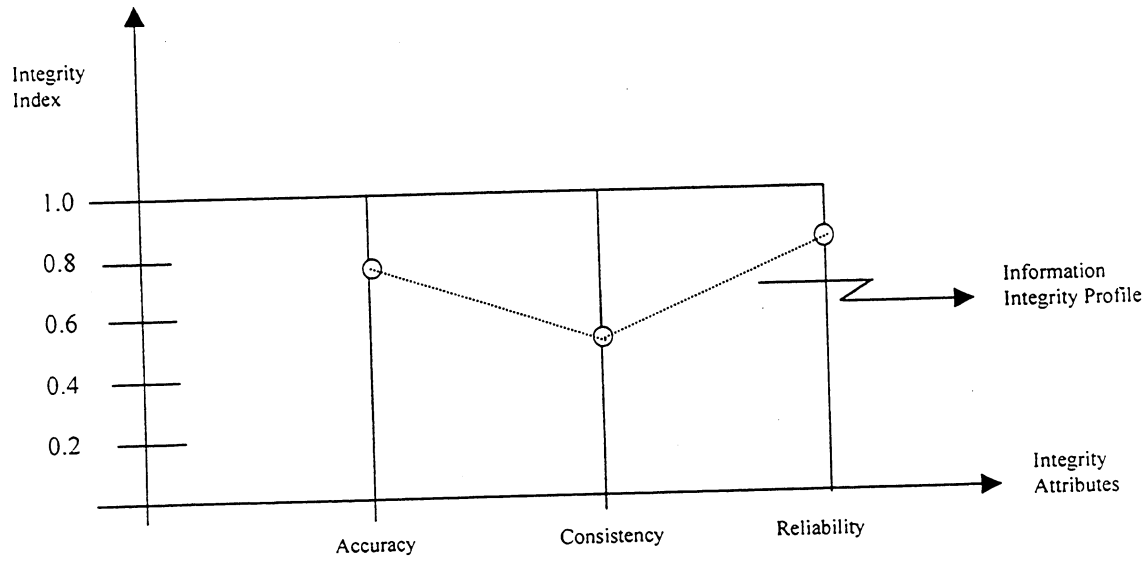



**Figure 4 : Information Integrity Profile**

### 3.6 Cumulative Information Integrity Index (CIII)

Let Information Integrity attribute, depending on the range in which the attribute index value falls, be assigned a 5-point scale as shown in Table 1 below :

**Table 1 : 5 - Point Scale for Information Integrity Attributes**

| Attribute Index Value Range | Scale | Points |
|---|---|---|
| [1–0.8] | H | 5 |
| [0.8–0.6] | G | 4 |
| [0.6–0.4] | F | 3 |
| [0.4–0.2] | E | 2 |
| [0.2–0] | D | 1 |

In the example under consideration, the Information Integrity attributes then have the scales and points as given below :

| Attribute Index Value Range | Scale | Points |
|---|---|---|
| Accuracy Index (A) | B | 4 |
| Consistency Index (C) | C | 3 |
| Reliability Index (R) | A | 5 |

Then with a view to quantify the overall Information Integrity Index for the given application by the user, a Cumulative Information Integrity Index (CIII) may be given by :

$$CIII = \frac{4W_a + 3W_c + 5W_r}{W_a + W_c + W_r}$$

For example, if $W_a = 6$, $W_c = 5$ and $W_r = 8$, then

$$CIII = \frac{(4 \times 6) + (3 \times 5) + (5 \times 8)}{6 + 5 + 8} = \frac{79}{19} = 4.158$$

CIII will thus take a value between [1–5]. It could be the situation that this value of CIII may be low from the user point of view and the user may be requiring minimum CIII value of 4. Further, user may want to improve CIII with additional requirement of Consistency Index having minimum "B" scale. It is to achieve this Integrity improvement that the user would then need to incorporate Information Integrity Technology.

Before one proceeds with further development of Information Integrity Technology Product structure, a word of caution is warranted here. The quantification of integrity attributes is not a trivial task even when it is possible [15] and quantifiers suggested above do not bring out the complexity involved. In respect of accuracy quantification, it is already mentioned that there could be a problem of correct value of the identified source (also called standard) being undefined, or being simply unknown. In situation an assumed standard itself may be incorrect as is often the case with data gathered some time in the past and with no corroborating evidence. In yet another situation there may be more than one correct value. Then there is a problem of how to quantify accuracy if the value does not lie on a real line, i.e., it is not a numerical. As regards to consistency quantifiers, though a relatively simpler concept than accuracy, it can assume complexities when all real database inconsistencies are to be measured (and which will be the need) or when Consistency is also to be studied for the conceptual view of the data or information. Finally, as already mentioned, reliability quantifier gives an index of an accuracy with which the information obtained presents data item in whatever

way the information system processed it. There can be no one way of calculating the reliability index and there will always be a need to develop one based on nature of available data, form of data and computation aids available for processing. All these areas then constitute the further research needs in the context of Integrity attribute quantifiers for Integrity improvement.

## 4. Information Integrity Technology Development Steps

With a suggestion for Cumulative Information Integrity Index (CIII) as above, within the framework of Information Integrity attributes of Accuracy (A), Consistency (C) and Reliability (R) argued, one can then identify Information Integrity Technology Development steps as follows:

i) Understand the user application of the computerized information system under consideration.

ii) Based on application area and based on organizational practices studied, establish organizational standard pertaining to data/information with reference to requirements of : accuracy, consistency, reliability and cumulative integrity.

iii) Study data/information flow through the Information System and define database(s).

**Note :** Apart from knowing how the Information System processes the data and apart from understanding more about the "noise" in the system, the study would also necessitate knowing wherefrom, how and data/information of what integrity flows into the system.

iv) Based on the understanding of data/information flow in the system, for the database identified, develop the Information System Model as in Figure 5.

v) Specify and document the data rules, also known as edits, to be implemented to study accuracy and consistency of the data/information.

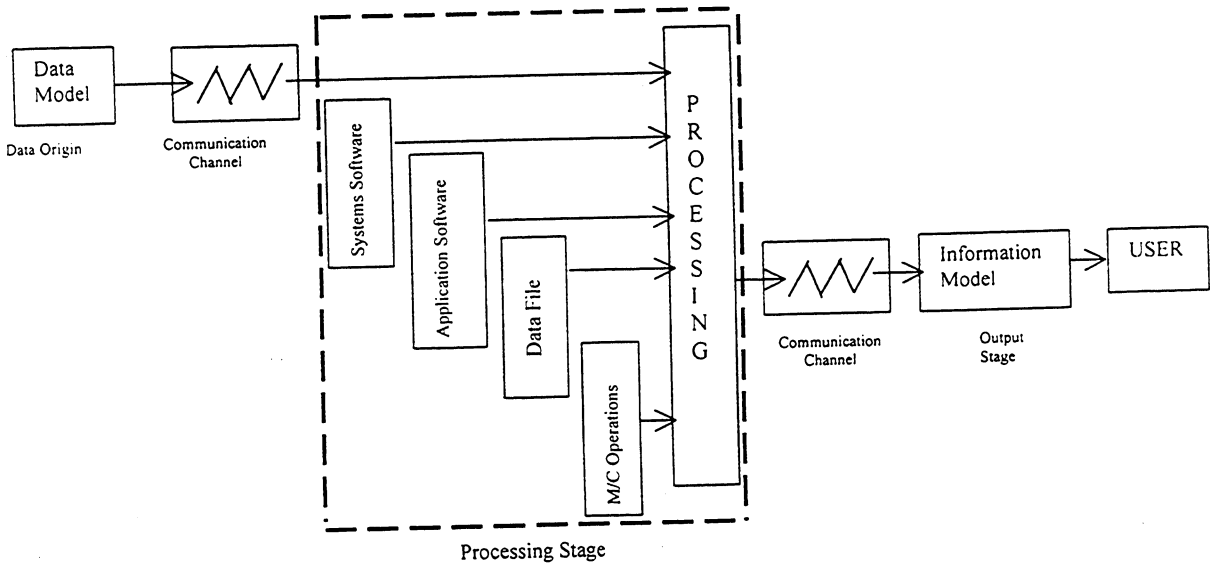



**Figure 5 : Data/Information Flow Model for an Information System for Developing Information Integrity Technology**

vi) While accounting for factors such as nature of available data, form of data and computation aids available and keeping in view advantages, disadvantages and convenience of application, choose a method for calculating Reliability Index.

vii) Develop Integrity Analysis Software for analyzing intrinsic Information Integrity attributes of accuracy, consistency and reliability.

viii) For the Information System Model in Figure 2, select a data sampling point at the output of a subsystem (or at an intermediate point within the subsystem), as close to the beginning of the Information System Model as possible.

ix) Depending on how data arrives at the sampling point (continuously or in batches), develop a continuous or batch processing sampler (a sampling programme) to randomly select a sample or records arriving at the sampling point. Along with sampling records, the sampler programme should also select some identifier of the sampling point and record of the data and time of sampling.

x) Following the selection of a sampling point and development of a sampler, select points for maintaining audit trail for sampled records.

xi) These points for maintaining audit trail may be selected at points at the output of subsystems (or at intermediate points within the subsystems) following the sampling point.

xii) Once the points for maintaining audit trail for records sampled are identified, develop a Sampled Records' Audit Trail (SRAT) programme to separate or pull out (at the points selected) the audit records.

xiii) Ensure that sampler programme and SRAT programme so developed can download sampled records and records for audit trail as in (ix) and (xii) above into a database to be set up (see Step (xiv) below).

xiv) Accordingly, based on hardware and software considerations and based on number of sampled and audit trailed records, download the sampled and audit trailed records on mainframe or minicomputer or on personal computer/workstation so as to set up an Error Detection Database.

xv) Using the Integrity Analysis Software developed in (vii), analyze the Error Detection Database to :

a) identify data rule violations in respect of accuracy and consistency attributes,

b) based on data rule violation statistics, establish degree of integrity of data/information in respect of Information Integrity attributes of accuracy and consistency,

c) obtain reliability index for the database along with analysis of factors contributing to the level of reliability,

d) based on indices for accuracy, consistency and reliability attributes, develop Integrity Profile and Cumulative Information Integrity Index, and

e) study changes in database not expected, i.e., irregular changes.

xvi) Compare the Integrity profile and indices obtained as in [(xv(b)) – (xv(d))] with standards in (ii) – local, regional, national, international as the case may be – and with the user specifications on Integrity, so as to know what is expected of Information Integrity Technology. This would also facilitate ordering or ranking the Integrity attributes from the point of view of which attribute needs maximum improvement effort.

xvii) Then, for each of the Integrity attributes of accuracy and consistency, by further analyzing the irregular changes either by subsystem or by field (in that order of priority of choice) locate separate Integrity improvement opportunities at each of appropriately identified pairs of a given field at a given subsystem.

xviii)Similarly based on reliability factor analysis in (vii), locate reliability improvement opportunities at each of subsystems.

xix) Having located pairs of a given field at a given subsystem each for improvements of accuracy and consistency and having located given subsystems for reliability improvement opportunities, further analyze the Error Detection Database and study irregular changes at each of pairs corresponding to each of accuracy and consistency attributes and study reliability factors at each of the subsystems, so as to understand over the time error patterns and causes contributing to loss of accuracy, consistency and Reliability.

This would then facilitate detecting error or cause that occurred sometime in the past $(t-\tau)$, or estimating error or cause at time (t), or predict error or cause that may occur at a future time $(t+\tau)$.

xx) Based on assessment as in (xvi) of integrity improvement target and based on the understanding of error patterns and factors for loss of intrinsic Information Integrity attributes as in (xix), now develop Information Integrity Improvement Action Plan for locations identified in respect of Integrity Improvement opportunities. This Integrity Improvement Action Plan may comprise restructuring subsystem(s) previous to the point of occurrence of error, improving integrity of data origin stage, improving communication channels, etc.

xxi) Finally, study performance of the Information System on incorporation of the Information Integrity Technology as outlined above. Accordingly obtain the intrinsic Information Integrity attribute indices, Integrity profile and Cumulative Information Integrity Index and compare them with appropriate reports before implementation of Information Integrity Technology available vide [xv(b)], [xv(c)] and [xv(d)], so as to quantify integrity improvement achieved and to check if it is as per customer expectation.

## 5. The Information Integrity Product

The Information Integrity Technology product thus developed would then be a SOFTWARE PRODUCT consisting of :

- the user data rules list for error detection

  **Note :** Data rule is that which must hold true in an Information System

- the Integrity Analysis Software for :
  - Accuracy
  - Consistency
  - Reliability
  - Integrity profile for the Information System
  - Integrity Indices
- the sampling programme
- the Sampled Records' Audit Trail (SRAT) programme
- the programme for
  - Statistical analysis of errors/causes for loss of integrity
  - factor analysis for reliability
  - time series analysis

- the programme for :
  - detecting errors/causes (filter programme)
  - estimating errors/causes (estimation programme)
  - predicting errors/causes (predictor programme)
- the generation of Error Detection Data Base
- the reporting based on analysis of Error Detection Data Base in terms of :
  - errors and causes detected; their locations in the Information System and their significance
  - error and cause patterns and trends obtained through statistical techniques such as time-series analysis
  - detection, estimation and prediction of errors and causes
  - identification of Integrity improvement opportunities
  - deciding and implementing Information Integrity Improvement Action Plan for Integrity Improvement opportunities identified (probabilistic action plan as also manual action plan included)
- obtaining improved Integrity Profile and Index
- documentation :
  - data rules list encoding the specifications for the Integrity Analysis
  - software and the reporting facility. This calls for user interaction.
  - the individual programme in accordance with the systems and programme documentation within the user organization
  - operating instructions for each programme
  - programme maintenance and test procedures
  - training material for users

## 6. Conclusion

Computerized information systems contain errors that are made but not corrected by controls built in at system analysis and design stage of the Information System. An Information System could be viewed as a production line in a manufacturing environment. Processing stage represents logic steps which utilize input transactions as raw material or parts to yield processed database records, i.e. information, as the end product. A typical production line incorporates process control, but more importantly, also employs product control. The identification of faulty processes alerts product quality control to invoke special procedures such as tightened inspection, repair or discarding of finished goods. Conversely, the disclosure of sub-standard products suggests remedial action for specific processes.

Information System testing becomes the equivalent of process quality control in that the errors revealed call for software revisions and maintenance. If the Information System is already operative, a test result indicating error would only suggest that such error may have occurred in the past. Test procedures do not access the production database, therefore, no statement can be made as to whether or not this Information System error has occurred in real environment, nor which records have been affected by the erroneous process. Therefore the confirmation of potential or suspected anomalies on a live database and subsequent integrity improvement becomes an essential facility (beyond application controls) within an Information System.

And the users of computerized information systems have to undertake this computer housekeeping to incorporate this facility in their information systems, so as to avoid potential serious losses occasioned by errors that were made (due to factors external to application control) but not corrected. In concrete terms this facility, constituting the Information Integrity Technology, will be an application and user specific software which, for an Information System Model as in Figure 5, on-line periodically and systematically samples records arriving at an appropriately chosen point (in the Information System Model) and then follows or keeps track of sampled records at subsequently identified points through the information system and stores the records so sampled and obtained through follow up (audit trail), to set up error detection database which is then analyzed to identify errors, i.e. changes not expected (irregular changes) and to quantify resulting loss of integrity therefore, followed by integrity improvement action wherein Information Integrity opportunity is identified and implemented.

Understandably, this Information Integrity Technology will have to be developed in a computer language compatible with the information processing environment of the user organization. This calls for organizational IS planning, devising policies, standards, and guidelines pertaining to data. If this is not ensured, net result is non-compatible, and hence unshareable data/information. An important step (in the development of Information Integrity Technology) in the context is data rule specification, in turn requiring defining data rule standard also needed for undertaking Information Integrity Analysis.

Yet another area that calls for standards pertains to degree of integrity. As mentioned, the application area would influence the requirement of how much accuracy or consistency or reliability. The application area would also influence values of $W_A$, $W_C$, $W_R$. Further, quantification of Integrity attribute such as accuracy calls for identification of data/information sources and their standards, i.e. correct values. In development of Information Integrity Technologies, it would therefore be necessary to establish these application area specific standards representing requirements of degrees of integrity as also of values of Integrity attribute significance factors.

Finally, it is important to appreciate that the development of standards as above would facilitate development of Information Integrity Technology products for different subsystems of the information system as also for the total system. This would call for support of reputable software developers and vendors for the purpose. Further, these Information Integrity Technology products would cover data/information in various forms – numerical or alphabetic or alphanumeric or video-images or any other – and that too for different application areas. This would open a new vista in terms of design, development, commissioning, operation and maintenance of data technologies, hitherto not attended, for ensuring on-line integrity of computerized information system.

## References


1. Ameen D.A., "Systems Performance Evaluation", Journal of Systems Management, (March 1989), pp. 33–36.
2. Biba K.J., "Integrity Considerations for Secure Computer Systems", USAF Electronic Systems Division, Bedford, MA (1977), ESO-TR-76–372.
3. Clark D.D., and Wilson D.R., "A Comparison of Commercial and Military Computer Security Policies", Proc. (1987), IEEE Symp. on Security and Privacy, IEEE, New York, (1987), pp. 184–194.
4. Courtney R.H., and Ware W.H., "What Do We Mean by Integrity", Computers & Security, 13, (1994), pp. 206–208.
5. "Data Quality Foundations", Published by AT&T Quality Steering Committee, U.S.A., (1992).
6. Kliem R.L., "Back to Basics : Developing A Good Requirements Document", Journal of Systems Management, (October 1992), pp. 16–19.
7. List W., and Melvilse R., "Integrity in Information Systems – Executive Summary : IFIP Working Group 11.5", Computers & Security, 13 (1994), pp. 295–301.

8. Mandke Vijay V., "Research in Information Integrity : A Survey and Analysis", Proceedings of the JNCASR and SERC Discussion Meeting at IISc Campus, Bangalore on Information Integrity – Issues and Approaches, Edited by Rajaraman V. and Mandke Vijay V., published by Information Integrity Foundation, New Delhi, India (1996).

9. Mandke Vijay V., and Nayar M.K., "Information Integrity – A Structure for its Definition", Proceedings of the 1997 Conference on Information Quality, Edited by Diane M. Strong and Beverly K. Kahn, MIT, Cambridge, Massachusetts, USA (1997).

10. Mandke Vijay V., and Nayar M.K., "Design Basis for Achieving Information Integrity – A Feedback Control System Approach", Accepted for presentation at the 1998 IFIP WG 11.5 Working Conference on Integrity and Control in Information Systems, (19–20 November 1998), Warrenton, VA, USA.

11. Mostert D.N.J., and Solms S.H. Von, "A Methodology to include Computer Security, Safety and Resilience Requirements as part of the user requirements", Computers & Security, 13 (1994), pp. 349–364.

12. Nayar M.K., "A Framework for Achieving Information Integrity", Proceedings of the JNCASR and SERC Discussion Meeting at IISc Campus, Bangalore on Information Integrity – Issues and Approaches, Edited by Rajaraman V. and Mandke Vijay V., published by Information Integrity Foundation, New Delhi, India (1996).

13. O'Shea G.F.G., "Operating System Integrity", Computers & Security, 10 (1991), pp. 443–465.

14. Rajaraman V., "Information Integrity – An Overview", Proceedings of the JNCASR and SERC Discussion Meeting at IISc Campus, Bangalore on Information Integrity – Issues and Approaches, Edited by Rajaraman V., and Mandke Vijay V., published by Information Integrity Foundation, New Delhi, India (1996).

15. Redman T.C., "Data Quality : Management and Technology", Bantam Books, NY, (1992).

16. "Security Functionality Manual", DTI Commercial Computer Security Centre, V.21-Version 3.0 (unpublished).

17. Svanks Maija I., "Integrity Analysis : A Methodology for EDP Audit and Data Quality Assurance", EDP Auditors Foundation, Inc. (1984).

18. Terry P., and Wiseman S., A "New Security Policy Model", Proc. (1989), IEEE Symp. on Security and Privacy, IEEE, NY (1989) pp. 215–228.

19. Tompkins F.G., and Rice R. "Integrating Security Activities into the Software Development Life Cycle and the Software Quality Assurance Process", Computers & Security, 5 (1996), pp. 218–242.

20. "Trusted Network Interpretation of The TCSEC", The National Computer Security Centre, Fort George G. Meade, MD (1985), DoD 5200 28–STD.

21. Wang R., and Strong D., "An empirical Investigation of Data Quality Dimentions : A Data Consumer's Perspective", TDQM-93-12, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA (1993).

0-0-0-0