Information Integrity: Research & Education Issues

Vijay V. Mandke
Research Leader,
Center for Information Integrity Research,
B-64 (FF), Gulmohar Park, New Delhi-110 049
E-mail: vmandke@unitechsys.com; vijaymandke@satyam.net.in

- A CIIR White Paper -

Pre-amble

- *With advent of steam power, successes in production automation sharply increased the volume and speed of energy conversion and material processing. This precipitated various structured and periodic control responses for 'standard' product in high volume business model. With innovations in information technology, the volume and speed of information processing and decision-making has undergone sharp increases. Accordingly, business enterprises for their competitive survival are looking for bigger business opportunities through customized products. For competitive advantage, this is requiring businesses to pass on the control baton to controlling Information Integrity of unstructured and aperiodic, i.e., continuous individual information origination and processing situation in the presence of uncertainty (The Information Integrity Imperative – Section (8), Sub-section (8.3)).*
- *Information is an organizing mechanism, which provides an ability to deal with the environment. Given the reality of ever changing environment, the assumption that data and information are perfect, once validated, and the practice that most information processing systems do not anticipate defective data and information are not acceptable (Section (10) Item (o)).*

Index

8. The Information Integrity Imperative
   8.1 Distinguishing between rational information decision and correct information decision making process
   8.2 Recognizing Originating Information As A Costly Process
   8.3 Criticality of Information Integrity for Competitive Advantage
9. I*I – Existing Perceptions: The main limitation of
10. I*I Research and Education Issues – An interdisciplinary view
    a. Secured computer research
    b. Accounting/auditing research
    c. Quality paradigm
    d. Communication Theory
    e. Decision Theory
    f. Information Model
    g. Decision situation – Defining information flow
    h. Information Economics – A case for a paradigm shift
    i. Integrity approach to error reduction
    j. Information and its Usefulness & Usability requirements – Developing basis for Cost benefit Analysis of I*I
    k. Information Value and Value of Improvement in I*I due to additional information
    l. Quantitative measures for I*I
    m. Information Integrity Risk
    n. Risk Aversion – Not a costless but a costly activity
    o. Integrity *IS*
        i. Reasonably well developed *IS*
        ii. Quality *IS*
        iii. Integrity *IS*
    p. Approach to Prevention, Detection, and Correction of error
    q. *SDILC* Model
    r. I*I Processes
    s. I*I Standards
    t. Application of System Dynamics methodology for development of I*I Technology
    u. Information Origination Resource Management
11. Information Integrity Research & Education Issues and Possibilities – A General Topography

Information Integrity: Research & Education Issues

Vijay V. Mandke
Research Leader,
Center for Information Integrity Research,
B-64 (FF), Gulmohar Park, New Delhi-110 049
E-mail: vmandke@unitechsys.com; vijaymandke@satyam.net.in

- A CIIR White Paper -

Abstract

Information Integrity (I*I) is dependability and trustworthiness of information and controlling it is a key factor determining strategic business advantage. Its attributes are accuracy, consistency and reliability of information system (*IS*) and information there from. Systems have changed a great deal but it is the processes by which systems are designed and implemented that have changed the most. The majority of systems are still comparatively simple, but internal as well as external user aspirations are becoming increasingly local and instant. In a business traditionally geared to work under exogenously generated alternatives and information, need - in the presence of the reality of ever-present external and internal system environmental factors - is now to continuously and locally (endogenously) generate alternatives and their information requirements. This dynamic decision-making process model is demanding informational work activities, hitherto not considered, namely, information processing for observation, storage, retrieval, verification, prediction of future states, precompiled responses and abstract reasoning, manipulation, communication, *use* and discard or storage for future *use*. And in the presence of system environmental factors all these activities assume complex and changing character besieged with uncertainty. Even one generation ago, large complex systems were implemented quickly and with very little documentation, but as a result of these new information processing demands today the businesses are increasingly burdened with the data and information, that too fraught with uncertainty, which in the wake of information processing methods of yesterday not only tend to become ends in themselves, but also slow down the process and increase the risks of failures. These informational failures, which can now be seen as an increasing problem as systems move from crisis to crisis call for simplification and speed. Within this framework this white paper examines I*I research & education issues and possibilities.

Information Integrity: Research & Education Issues

Vijay V. Mandke
Research Leader,
Center for Information Integrity Research,
B-64 (FF), Gulmohar Park, New Delhi-110 049
E-mail: vmandke@unitechsys.com; vijaymandke@satyam.net.in

- A CIIR White Paper -

1. **I*I Definition**
   Information Integrity (I*I) is dependability and trustworthiness of information and controlling it is a key factor determining strategic business advantage. Its attributes are accuracy, consistency and reliability of information system (*IS*) and information there from.
2. **From fixed to flexible information – Shift in control decision for competitive advantage**
   Traditionally, with enterprise seeking to produce only 'standard' product in high volumes, control systems tuned to structured & periodic (i.e., collective or fixed) 'data/information decisions' could ensure business objectives of operational optimization and cost efficiency, thereby giving the business strategic advantage. The enterprise did have computerized information systems (*IS*) developed in isolation, there being no effort to optimize data or information for improved decision making, i.e., for environmental or local market factor based product beneficiation for strategic advantage. However, with data-driven technologies keyed to the flow of digital data throughout an enterprise and on the Net and with competitive survival pressure of finding bigger business opportunities, business enterprise has a further requirement for utilizing data/information decisions 'smarter'. This calls for automation of 'informational work' carried out by the soft components of the enterprise, which is an application of flexible automation - a structural variant from inflexible automation.
3. **Open system view of generic business process – Informational work system**
   Given the reality of data-driven technology and pressure for competitive survival, business organizations, their systems, sub-systems and their components are getting recognized to be "open systems". Unlike close systems, open systems have purpose (objective), possess porous boundaries with their environment, and, whatever else they do, they necessarily process (export and import) information. With system integration, this makes for business activity *emphasizing* information and comprising informational and physical work systems. For competitive advantage, requirement then is: (i) to recognize information processed to be function of source, process and *recipient* (i.e., *user* or customer) and (ii) to maximize informational work (IW) comprising activities of: (a) generating from business process activities raw data/information in a complex and changing real world environment, characterized by uncertainty and hence errors, and (b) processing this information on current basis for generating *endogenous* business process design alternatives and evaluating them to deliver *flexible* information decision for control implementation at the physical work system.

4. **Business Process IS View: Continuous individual information origination situation under uncertainty**
   This necessitates viewing a system as a potential source of information and describing it "as a network of information variables in causal relationship to one another and in situations even to themselves" and models the business process as an integral to a closed loop information and control system characterized by uncertainty. This Business Process *IS* View is an open system with multiple decision stages. These are: (a) Problem Information Origination Decision Stages, namely, based on long term goal set (vision and value statements), *obtaining* 'many factors' & 'multiple criteria' characterizing problem (task) complexity (D1); from multiple criteria (conflicting goals inclusive), *recognizing* (deciding) on operable goal (D2); from operable goal statement, *defining* opportunity and constraints spaces (D3); and from 'many factor' information variables characterizing problem complexity, *culling out* useful (relevant) information variables (D4), (b) Problem Environment Information Origination Decision Stages, namely, *recognizing* relationships (interdependencies) between *culled out* information variables (D5); and *developing* state transition models defining dynamic behavior of *culled out* state (information) variables (D6), and (c) Endogenous Alternatives Information Origination Decision Stages, namely, undertaking customized planning & design for *generating* robust alternatives for evaluation and selection of *flexible* information decision for control implementation (D7). Further, each of these decision stages has it's own uncertainty. In an ever-changing environment, this presents Business Process *IS* View as a *continuous individual information originating and processing situation in the presence of uncertainty* or "*information origination situation under uncertainty*" for the short.
5. **Information production (origination) errors – Environmental view**
   Integrity deviation is as a result of errors. Like Business *IS* View, its each decision stage is also an information origination process. Therefore, Business Process *IS* View and its each decision stage commit information origination resources. Specifically, each decision stage produces its respective problem information, problem environment information, and endogenous alternatives information. For a given decision situation, due to the uncertainties in the elements of the information origination process, there can be incorrect production of, say, problem environment information – implication of environmental variations. This, by the way of *distortion*, contributes to incorrect production of the said decision and results in loss of information origination integrity and its consequences for the Business Process *IS* View. On the one hand, for Business Process *IS* View, this indicates requirement for *correctness* aspect of information, where *correctness* also includes *exactness*. On the other hand, this presents information production errors as those arising from system's errors in dealings with environmental anomalies and guaranteeing system failures. This formalizes importance of the environment as a major factor in business decisions, which is critical to competitive survival in complex and changing environment.
6. **Emergent error model – Informational error, which is a decision error and unobservable**
   Discipline of reliability engineering extensively deals with the study of the consequences of incorrect operations at two levels, namely, (i) mechanistic failures, service disruptions, failure of computer hardware, etc, which are of stochastic type, and (ii) failure of system equipment that is controlled directly by the computer and, therefore, is dependent upon the correct functioning of the computer hardware and its software. Error then is defined as a

deviation from the required (standard) function/ operation of the component, sub-system, or the system. These studies, which treat error as a mechanism by which the fault in a component or a sub-system or a system becomes apparent, i.e., observable, are based on failure mode analysis methods, which use concepts of 'fault', 'error', 'accident' or 'adverse event' and 'failure'. Thus, error here is a functional, observable error and it may lead to a failure or an accident, which are also observable.

7. A paradigm shift – System failure not observed failure but informational failure in system's setting

However, as argued, there is the issue of consequences of incorrect operations at the third level, also, and it is the issue of system failure due to production (origination) of incorrect information. Informational errors take place in Business Process *IS* View at each of decision stages D1 – D7. Informational error is a decision error and when it takes place, the functional, observable error and the resulting observable system failure (or even collapse) is yet to occur. While, on the one hand, this suggests that informational errors are not observable, on the other hand, it presents informational error as one that evolves into subsequent information error and finally into the observable, functional error. This makes informational error inclusive of functional error. More significantly, this informational view permits defining system failure not as an observed failure but as informational failure in system's setting.

7.1 Informational errors far more than functional, observable errors – A rough estimate

Informational errors are far more than functional, observable errors. Precise data on the extent of informational errors is just not available and bound to vary from system to system, depending on how error is defined. Literature reports one study across various types of systems, which attributes 40% of errors to material, electrical, and mechanical failures, which are at first two levels. The remaining 60% are attributed to informational errors. Given that information, in general, has become more and more important in modern business, this is a quantitative pointer to overbearing nature of information errors and a recognition of urgent need for their rigorous study.

8. The Information Integrity Imperative

For competitive advantage, requirement then is to minimize informational errors due to distortion and noise, i.e., to ensure *correctness* aspect of information, which also includes *exactness* aspect. This demands going beyond "reliability" requirement and to ensure "accuracy" (A), "consistency" (C), and "reliability" (R), i.e., Information Integrity (I*I) of information system (*IS*) and information there from. This introduces requirement to originate information effectively and economically, which in turn calls for recognizing that information "origination" is a *costly* activity.

8.1 *Distinguishing between rational information decision and correct information decision making process*

Business Process *IS* View comprises decision stages, which are information origination processes under uncertainty. Information origination is a costly activity and hence decision stages under Business Process *IS* View are costly processes. The production of information consumes resources, which have alternative uses. For this reason the concept of a rational information decision must be distinguished from the process of rational information decision-making. A rational information decision depends upon a determination of appropriate operable goals to guide choice and implies the selection of a course of action, which will make the greatest contributions to those goals. The process of rational information decision-making, on the other hand, recognizes that the costs of reaching a rational information decision may exceed the benefits. Rational information decision, which could be viewed as correct information decision, which will eliminate uncertainty about appropriate goals or the range of alternatives, is, like any other commodity, not costless. Unless its costs are exceeded by the gains expected from solving the correct (as opposed to incorrect) problem or finding a less expensive alternative, it should not be acquired. Because information decisions can and should be made in the face of uncertainty, the objective of correct information decision making process then is not a rational information decision. It is the efficient use of resources in the correct information decision making process to acquire control, the ability to direct means to ends.

8.2 *Recognizing Originating Information As A Costly process*

For efficient processing of information {I}, there *has* to be economic trade off between: (a) Costs of originating information "I", and (b) Loss due to incorrect information, i.e., due to I*I Risk (see Section (10) Item (m)). In other words, *that IS* which, for a certain kind of information *origination* is able to arrange them (costs) at the lower level *will* tend to prevail. In view of this it follows that, *to compete successfully*, "I", must have *integrity*, i.e., it must have accuracy, consistency and reliability.

8.3 Criticality of Information Integrity for Competitive Advantage

For simplicity of presentation, only Accuracy attribute of I*I is considered. It is denoted by A(I). Let, I(t)=Information, IUUB=Information Use (IU) Upper Bound, ΔIU(I)= Net information Use Benefit, α(I)=Usefulness Factor, β(I)=Usability Factor, A(I)=Accuracy of I(t). Then, Gross "IU" Benefit=GIUB(I) is given by GIUB(I)=[α(I)×β(I)×IUUB(I)]×[A(I)] and ΔIU(I) by:

$$\Delta IU(I)=GIUB(I)-[COST_{ORIG}(I)+ COST_{ANALY}(A(I))+COST_{OPPOR}(A(I))] \quad \ldots\ldots \text{Equation (1)}$$

where, [$COST_{ORIG}(I)$=Cost of originating information (I)], [$COST_{ANALY}(A(I))$=Cost of analyzing I*I, i.e., A(I) in this case], and [$COST_{OPPOR}(A(I))$=Opportunity cost of analysing I*I, i.e., A(I)].

From Equation (1) what emerges is *IS* under consideration *will* have a maximum value at a given time, and, among other things, *for* a given (what can be seen as an optimum, i.e., desired or, say, intended) value of integrity quotient "A". In other words there *is* an optimum I*I at which net increase in information *use* benefit is maximum; achieving that I*I (implying accuracy, consistency, and reliability) is a costly process; and, to meet the demands of competitive advantage, resource commitment for achieving improved I*I, preferably optimum I*I, is critical. Equation (1) (with I*I (I) substituted for A(I)) gives the generalized *IS* model presenting I*I (I) as a resource for improved information *use* and establishing requirement for I*I analytically.

9. I*I – Existing Perceptions: The main limitation of

Literature reports integrity studies from different angles: security based definitional approach to integrity, auditing research, process centered quality approach, noise reduction based technology under communication theory, and the Savage (Subjective) Expected Utility (SEU) theory under decision-making. The main limitation of these approaches is that the *IS* models considered do not account for the requirement of continuous origination of information endogenous to the specific decision situation. Their main concern is only that information technology accesses, communicates, processes and distributes the already generated information. With information technology costs ever decreasing, the information processing for decision-making is, therefore, taken as a *costless* activity. This results in *IS* models having no explicit reference to cost-benefit of information processed, there being no analytical basis for comparing two information processing with reference to net information *use* quantum delivered. These *IS* models are characterized by uncertainty resulting in information errors leading to loss of Information Integrity, which is seen as data integrity problem. Accordingly, it is appreciated that it is good to have data integrity, *except* that how much remains the issue. This certainly is not amenable to any analytical comparison and selection for competitive advantage; thereby integrity not receiving its deserving attention in terms of resource allocation for systems emphasizing it.

10. I*I Research and Education Issues – An interdisciplinary view

This gives rise to interdisciplinary I*I research and education issues and the same are briefly mentioned below.

(a) **Secured computer system research**: Integrity research in computer science has its origin in study of secured computer systems and of confidentiality of information. Here, security has normally been taken to mean confidentiality, integrity and availability. Researchers involved with information security issue are at ease with this terminology *except* that the meaning of the word "integrity" is not adequately resolved, the word being frequently used to describe a range of attributes (or requirements). Further, database integrity models and methods, while context specific, do not lend themselves to any comparative, analytical studies. In the computer security field, despite the Clark-Wilson model and the considerable integrity discussion it prompted, there is still nothing like coherent framework. The integrity research effort has been either very pragmatic, and/or technological, or almost semantic in nature, and in any case there is no reference to the cost benefit framework for Information Integrity – an aspect so crucial to business decisions.

(b) **Accounting/auditing research**: In accounting/auditing research there seems to have been no corresponding debate concerning the exact meaning of "integrity". Specifically, the auditor assesses control risk, according to Statement of Auditing Standard: SAS 55, as determined by the relevant parts of the entity's (Audi tee's) internal control structure. With respect to accounting information, relevant part of the internal control structure is thus made up of three parts (categorizations): the control environment, the accounting system, and the control procedures. This certainly offers a way of structuring the analysis of different possible control mechanisms. However, there is a problem in that there is no explicit coupling to cost and benefits in the sense items in different categories can be compared. The categorization in three parts is essentially ad-hoc. Then there is the COSO report that provides an extended framework, but it is qualitative in nature. It sees internal control, from the management point of view, as consisting of five interlocking factors: monitoring, information and communication, control activities, risk assessment, and control environment. However, the same line of inadequacy that is leveled at SAS 55 above, that is, lack of explicit cost-benefit links between the components of model, applies here.

(c) **Quality paradigm**: Coming to the quality paradigm, it is an extensively addressed research area. Traditionally, auditing has been used to check whether the various *IS* quality assurance procedures are being carried out correctly, and whether they are successful. Quality paradigm came in to prominence in the wake of globalization, international competition, and changing customer expectations. Specifically, it has its origin in the high-volume mechanical manufacture. Accordingly, it has two aspects; namely, (a) quality assurance concentrating on the process and attempts to ensure that it is done correctly, and (b) quality control aiming to ensure that the product delivered to customer is correct, where the term 'product' represents a system or component or service. In practice, however, the quality paradigm operates in the 'standard' product mould, emphasizing incremental changes, and sees its operable goal as 'reduced defects'; thereby emphasizing cost reduction aspect but not the cost-benefit angle. This leaves the quality emphasis weighing more on the side of 'process'-centered issues rather than 'product'-centered issues. Unlike in the case of standardized industrial products, information product is so because it is 'unique'. As a result, for an information system, although quality assurance is vital, it alone is not sufficient to achieve an adequate level of quality.

(d) **Communication Theory**: Then there is information system (*IS*) model as in communication theory involving source, channel, and destination (receiver) as its fixed parts and the coder and decoder as the variable parts. In Shannon's words, "the fundamental problem of communication is of reproducing at one point (destination) either *exactly* or approximately a message selected at another point (source)". The cause of this problem is the ever-present channel noise, which tends to disfigure the message, and the noise reduction technology envisaged is optimization of variable parts so as to improve reliability, increase the data rate, or decrease the cost. Indeed this model is so widely used that, as a result of Shannon's work, in less than 50 years information theory has scaled the heights of new mathematical discipline. Even then from the point of view of the research query under consideration, it is important to note that this *IS* model does not take a decision process based view of the message through the channel. In fact, although the measurements in information theory are significant to communications engineer, they are not related to decision issues, except by chance. Accordingly, then there is no reference to the cost-benefit framework for the degree of "exactness" of message achieved.

(e) **Decision Theory**: The Savage (Subjective) Expected Utility (SEU) Theory presents a mathematical model based on the concept of information value to analytically study a decision process model in the presence of uncertainty. In view of this, when an *IS* is modeled as a decision process characterized by uncertainty, to study its integrity issues, SEU model emerges as an obvious candidate. The elegance of the SEU framework is almost beyond compare in the field of economics. However, to this date there is confusion as to the descriptive validity of SEU maximization. The sources of confusion partly, at least, seem to lie in the fact that SEU maximization is descriptively *invalid* – falsified – as a model of how individual decision makers behave. Nevertheless it is descriptively *valid*, or at least constitutes the best alternative currently available, as a model of individual decision making when building theories of collective decision making at the market level. This apparent paradox has caused much of confusion. Further, it defines the monetary value of perfect information as amount of money which renders the decision maker indifferent between using and not using information; and thus does not consider in its treatment any explicit coupling to cost-benefit analysis for the information value it measures.

(f) **Information Model**: Given the reality of complex and changing environment, business process is getting modeled as an integral to a closed loop information and control system characterized by uncertainty. The emerging Business Process *IS* View is an information origination situation in the presence of uncertainty due to implications of distortion and noise. This presents information modeled as "information envelope" so as to include requirements of problem information, problem environment information and endogenous alternatives information. This is pointer to a shift in database modeling, which should include information acquisition cycle and information *use* cycle.

(g) **Decision situation – Defining information flow**: Business Process *IS* View delivers a *flexible* information decision (D7). Also, Business Process *IS* View comprises decision stages delivering intermediate information decisions (D1-D6). It is useful to model all these information decisions as events when the information flow takes place, which formalizes environment as major factor in decision-making. If at each decision stage the information processing for decision-making and information there from (in the manner of the delivered information decision) have integrity, there is competitive advantage and if not then loss of it.

Within above framework, any given decision situation, which is information origination situation, is described as follows:

a. Given the informational view of the system, a decision stage has a set of decision rules that differ over a range of outcomes that the decision-making, *ex ante*, estimates as possible.
b. Assumption is that the decision rules will to some extent influence the realization of the outcomes over which it has preferences.
c. The decision-making process has some initial information, which makes it possible for it to see the problem in terms of a choice between different possible decision rules.
d. Decision-making, which requires information origination, is a costly process.
e. The decision-making process has limited information origination resources.

(h) **Information Economics – A case for a paradigm shift**: Traditionally economists suggest that decision makers are constrained by the opportunities available to them. Here decision alternatives are assumed generated exogenous to the decision situation. Accordingly, the study of economics is primarily the study of choice among a restricted set of opportunities. Informational view of business process, however, sees opportunities as alternatives generated *endogenous* to the decision situation. This is entirely different perspective, which for comparative advantage introduces further need for continuous origination of information, which in itself is a limited resource. This has as its consequence the requirement for information *recognition*. The governing principle is that business decision makers face unlimited opportunities coming form every direction. Unexpected opportunities surface every day. Only that some are recognized, others are missed; some are originated, others are forgotten; some are acted upon, others are passed over.

(i) **Integrity approach to error reduction**: Search for integrity is search for error reduction. The problem of error reduction can be approached in two different ways. One approach can assume that if people are more careful, pay more attention, and in general take more trouble over what they are doing, then errors can be reduced and their effects mitigated. That is to say, the error problem is seen as of that moment having no significance beyond itself. This approach, ad-hoc in nature, puts whole attention after a particular error. In real world, the error, however, does not occur again in the same form and in a same situation in a linearly predicted manner. As a result, this approach, easier to pursue, gives a false sense of having taken steps for error removal. It never minimizes the error occurrence, though, and is invariably found less effective in the long run.

The other approach – integrity approach, which is also a systems approach - may see design of objects, activities, rules and procedures, norms, commands, and patterns of behavior as being the source of errors. Clearly, systems approach is holistic, more in tune with the setting in which the real world operates. It does not see the error as, say, a "mechanical engineering" problem, but as that of (or more correctly as that of loss of) integrity, that is trustworthiness and dependability (here say in a "mechanical engineering" setting), of: each of the system components as also the complete system; of each of system development & implementation phases as also the total lifecycle model. This emphasis on integrity of component (or phase) as also of complete system (or total lifecycle) is important in that it also suggests requirement of integrity in respect of relations and interactions between the components and between the phases. Only when this entirety of integrity requirement is ensured will the error be minimized.

(j) Information and its Usefulness & Usability Requirements – Developing basis for Cost Benefit analysis of I*I: Section (8.1) shows that the objective of rational information decision-making process is not a rational information decision. It is the efficient use of resources in the correct information decision-making process to acquire control, the ability to direct means to ends. Therefore, for competitive business advantage, the Business Process *IS* View must manipulate information "analytically" so as to achieve improved decision; that is, it must process information effectively and <u>economically</u>. What information are we talking about here?

I*I research shows this information {I}, so to be *originated* and processed, in fact, has three components. These are: (a) Information (I_1), which is in a form of an aggregate or a measure so as to compare two or more alternatives and to select one flexible (customized) information decision for control implementation, (b) Information "I_2", on market imbalances indicating business opportunities, and (c) Information "I_3", constituting knowledge of working mechanisms for resource allocation. Further, *recipient* or customer (local market factors) requires that a business *IS*, as above, *originate* and process <u>that</u> $\{I=I_1+I_2+I_3\}$, which are ***useful*** (relevant) and make it easy to function in the market, i.e. which are ***usable*** to rank the *originated* alternatives for comparison and to make a customized information decision selection. What is needed is an analytical pointer to the product nature of the information "I_1".

Equation (1) in Section (8.3) provides this analytical pointer by giving *IS* model for cost benefit analysis of I*I.

(k) Information Value and Value of Improvement in I*I due to Additional Information: Equation (1) then provides a basis for developing mathematical equations for (i) value of information, which accounts for cost benefit analysis framework (unlike that (equation for information value) in case of SEU Theory (Refer Section (10) Item (e)) and (ii) for improvement in value of I*I due to additional information.

(l) Quantitative measures for I*I: Obtaining mathematical equations for information value and for improvement in I*I due to additional information would call for developing quantitative measures for accuracy, consistency, reliability attributes and for I*I.

(m) **Information Integrity Risk**: In the form of I*I risk, Section (8.2) mentions the informational view of risk. Specifically, concern *here* is the decision process in a complex and changing environment. Decision process has preferences over decision rules and over outcomes (Refer

Section (10) Item (g)). Further, decision process is characterized by (i) *Ex ante* expectation of the decision outcome, and (ii) *Ex post* decision outcome. What kind of risk will such decision process experience?

The risk it can experience is that of making a choice *ex ante*, which will, according to it's *ex ante* estimate, perhaps turn out to have been sub-optimal *ex post*. This kind of risk has to be caused by information, which is inexact (noisy) and incorrect (distorted).

Effort here is to maintain an *ex ante*, subjective view from the decision process angle. Even though it chose according to what from its point of view was certainty it (decision process) may have surprise waiting for it *ex post*. But before hand, from its point of view, this risk is captured by the values it estimates (understandably through the I*I Analyzer) for the information integrity attributes of accuracy (A(I)), consistency (C(I)), and reliability (R1(I)) for correctness and of Reliability (R2(I)) for exactness.

This is the statement of Information Integrity Risk, which consists of Information Correctness Risk and Information Exactness Risk.

(n) **Risk Aversion – Not a costless but a costly activity**: Organizations live in uncertain environments. The behavior of the market, suppliers, shareholders, and government is uncertain. The behavioral theory of organizational decision-making assumes that the organization will seek to avoid risk and uncertainty at the expense of expected value. The governing principle is, in general, the decision process will be willing to accept a reduction in the expected value of an outcome in exchange for an increase in the certainty of outcome and that the decision process has, *ex ante*, exact and correct information on which to base its choice of a prospect. In other words, the decision making process is certain that there is no way it can obtain additional information and that it has all the information it needs to choose. In this sense, the "classical" risk aversion of mainstream economics is costless to carry for the decision process, and causes no information demand.

However, given the reality of ever changing environment, risk aversion has a requirement for additional information and it is a costly activity. The concept of risk is function of information (Section (8.2) and Section (10) Item (m)). This lends risk to have a model- and context- specific interpretation. In the manner of local market (environmental) factor beneficiation, it requires generating alternatives <u>*endogenous*</u> to the specific decision situation. As a result, there is a paradigm shift in the view of risk from that under expected utility framework. Given the informational view of risk, the decision process can not be *ex ante* confident or certain that the information it is basing its view of the of the decision situation on is as correct and exact as it can be. Therefore, there is, clearly, a demand for additional information, i. e, individual information origination situation – for increasing I*I(I).

Unless and until the decision process becomes sure that there is no information that it could obtain which would enable it to choose better it will experience Information Integrity risk: the *ex-ante* perceived possibility that devoting additional decision capacity or resources in order to acquire and apply more information to this decision might lead to an *ex post* more optimal result with regard to its preferences. It is the presence of this Information Integrity risk (comprising information correctness risk and information exactness risk) that then gives rise to an opportunity cost of decision making in other decision situation(s) (i.e., prospect(s)) which it (decision process) faces, and which compete for its limited information processing capacity – making risk aversion a costly activity.

(o) **Integrity *IS***: In economic terms a larger and larger percentage of business resources is expended on information and information technology. Systems theory and economics point out that the complexity of an organization is limited by the amount of information that it can (economically) transform (process and transfer), i.e., by the costly bandwidth of its internal communication channels. This holds for organization as a whole, and its systems, sub-systems and their components. The degree to which higher efficiency through increased specialization (*this specialization is I*I*) is feasible is governed by the means that are available to control the ensuing complexity, that is, by the cost and efficiency of the available information systems. The question is what kind of information system (*IS*) will, in different cases, turn out to be the most cost-effective?

a. **Reasonably well developed *IS***: In traditional business, with controls tuned to "fixed" information decisions, for competitive advantage, strategy is to either have "good enough reliability" or, at a higher level as in Quality *IS*, aim at "zero error".

   Under good enough reliability requirement, the information system focuses on error discovery more in the form of observed failure. The approach is to develop information system that works reasonably well, and fix the bugs in ad hoc, repair and service manner, as and when they (bugs) appear after the system has become operational. As a result, in this *IS* design, little attention is paid to the type of information system required to manage integrity.

b. **Quality *IS***: Against this, the "zero error" aim, which is pursued by quality information systems, is based on the idea of "doing it right the first time". In pursuance of this aim, all the trial and error processes should be carried out at the prototyping stage of design, in which all the defects are identified and removed and (once again, in the manner of errors in "fixed" design decision under the assumption of static environment) no error is anticipated in subsequent phases. In other words, in comparison to "reasonably well developed *IS*", the design phase of quality information system includes two additional aspects, namely, reliability and robustness.

   Within above framework, once the system becomes operational, it is monitored, by regular collection of data for early detection of service problems. Monitoring data then becomes the basis for process improvement before any major problem arises in the system, thereby improving the system on a continuous basis. Finally, in Quality *IS*, the cost of system ownership includes future maintenance and loss due to lack of reliability, which comprises: (a) Cost of non-performance, and (b) Cost of lost opportunities. The cost of non-performance is when the system is under-designed. The cost of lost opportunities is the lost profits or benefit due to failure of the *IS* to provide its customers with needed information.

c. **Integrity *IS***: Quality *IS* is certainly analytically superior to "reasonably well developed *IS*". However, with its concern limited to "reliability" attribute of information, Quality *IS* views environment as static or incrementally linear and does not formalize it (environment) as major decision factor in decision-making. Put differently it does not have requirement to originate information; it treats information as a by-product and not as product; it does not anticipate error in information, once validated, does not process maximal information requirements, and does not account for their information integrity. This leaves Quality *IS* emphasizing competitive advantage through cost-reduction, which is necessary but not sufficient in complex and changing environment.

Concern of the Integrity *IS*, however, is *correctness* aspect, i.e., accuracy, consistency, and reliability of information. Achieving this goal requires continuously originating all through *SDILC* model phases problem information, problem environment information and endogenous alternatives information, measuring I*I, and using system engineering techniques to discover potential problem areas. As a result, monitoring I*I of all the information originated at all the decision stages D1 through D7 becomes the basis for (information) process and product improvement even before any informational error problem arises in the system, thereby adapting and improving the system on a continuous basis. Finally, as explained in Sections (8.2) and (8.3), Integrity *IS* cost includes (a) cost of origination of information (I), (b) cost of analyzing I*I(I), and opportunity cost of analyzing I*I(I).

(p) **Approach to Prevention, Detection and Correction of error**

a. **Reasonably well developed *IS***: Unless an error or defect is discovered that would require maintenance, an operational system is used as it is. Direct cost forms the basis of resource allocation to various components of the system and reliability of the system is considered an additional cost, requiring time and money.
b. **Quality *IS***: The goal is to produce zero-error operational systems. Achieving this goal requires identifying reliability of measures, collecting data, analytically discovering potential problem areas, and taking preventive actions. Quality approach accepts three phases for finding errors, namely, (i) **Prevention**: prevent errors from occurring by designing reliability into system, (ii) **Detection**: search for errors through constant collection and monitoring via control charts and other statistical techniques while the system is working, (iii) **Failure**: wait until the system fails.
c. **Integrity *IS***: The goal is to minimize occurrence of errors by using systems approach to error reduction (systems approach considers design of objects, activities, rules and procedures, norms, commands, financial mechanisms, policies, and patterns of behavior as being the source of errors). Integrity *IS* has following phases for error reduction: (i) **Anticipation**: anticipate errors at decision stage itself, (ii) **Correction**: correct decision errors by using system engineering techniques to discover potential problem areas and take preventive actions at informational as well as functional levels, (iii) **Failure**: not acting to correct decision error as soon as it occurs and saving system from occurrence of (even delayed) functional failure.

(q) ***SDILC* Model**

a. **Reasonably well developed *IS***: Design phase is just one phase in the sequential *SDLC* model.
b. **Quality *IS***: In quality *IS*, major participants in the subsequent phases of *SDLC* model take part in the design phase, and plan their functions as the design progresses.
c. **Integrity *IS***: In Integrity *IS*, all participants in each of *SDILC* model phases are concerned with the remaining phases, *also*. As a phase progresses, requirement is to

plan for decisions and actions with respect to *that* phase as well as other (phases). This facilitates each *SDILC* model phase of Integrity *IS* recognizing continuous interaction among complex systems and including additional aspects that are missing in the design and implementation of quality information system: **accuracy**, **consistency** and **reliability** of information system and of information there from.

d. **I*I Processes**: For competitive advantage in complex and changing environment, business process is modeled as Business Process *IS* View, which is an information origination process in the presence of uncertainty. For this *IS*, Section (4) mentions error implications in respect of problem environment information due to incorrect origination of environmental variations knowledge factor relationships, which impact critical *IS* variables and have consequences of loss of I*I and thereby of competitive advantage. We will term these environmental variation relationships' errors as Level 1 Errors.

There are informational errors at other levels, too. These are: information recognition errors (Level 2 Errors), information origination errors in respect of opportunity and constraining spaces (Level 3 Errors), and information origination errors in respect of strategic factors characterized by conflicting goals (Level 4 Errors). Specifically, when something is found in the environment that is anomalous (environmental integrity deviation), information origination is required and a process of origination has to begin, which requires other shifts of resource. This introduces the question of recognition of information in respect of knowledge factors relating to environmental variations. It is only if these knowledge factors are correctly recognized that the correct information can be originated, otherwise not. This draws attention to implications of errors in information recognition (Level 2 Errors), which lead to Level 1 Errors and, therefore, to loss of I*I resulting in loss of competitive advantage.

Section (4) mentions that the Business Process *IS* View has requirement for endogenous alternatives' information origination. Further, Section (10) - Item (h) points out that, given the ever changing environment, informational economics suggests that managers face unlimited opportunities coming form every direction. Unexpected opportunities surface every day, in fact every moment. The need to stimulate and control opportunities, therefore, is central to business strategy and control, and it must be met with integrity. If there are errors in origination of 'alternatives (opportunities) & constraints' information (Level 3 Errors), there will be errors at Levels 2 and 1, leading to loss of I*I in Business Process *IS* View and, therefore, loss of competitive advantage.

Finally, Section (4) mentions need to originate problem information requirements, which comprise operable goal information inclusive of conflicting goals. *IS* operates in real world and real world strategic factors necessarily involve information requirements in respect of: trade-offs between goals that conflict, factors that must not change, malfunctions that emerge with delay, and consequences of uncertainties, time factor, inadequate information origination resources, etc. Development of these information requirements must be done with integrity. If there are errors in this information origination (Level 4 Errors), there will be errors at Levels 3, 2 and 1, leading to loss of I*I in business process *IS* view and, therefore, loss of competitive advantage.

The requirement then is to ensure I*I at above levels. This calls for development of I*I Processes at Levels 1-4.

(s) **I*I Standards**: Search for I*I Processes would lead to development of I*I standards as follows:

- **(i)** With reference to I*I Process Level 1: Dictionary of environmental variations knowledge factors with models of relationships between them illustrating problem solving in different contexts.
- (ii) With reference to I*I Process Level 2: Dictionary of standard contexts and situations describing:
  - (a) Standard contexts in which standard situations occur, and
  - (b) Standards patterns in which situations evolve (e.g., implications of ergonomic anomalies), which activates certain kinds of standard environmental variations knowledge factors as relevant to evolving situation.
- (iii) With reference to I*I Process Level 3: For different combinations of contexts and situations,
  - (a) Dictionary of Standard Rules and Standard Statements of Limits for Opportunity Space, and
  - (b) Standard Defined Sanctions and Standard Credible Threats for Constraining Space.
- (iv) With reference to I*I Process Level 4: Dictionary of standard contexts and situations describing standard information decisions delivered through decision stages (D1) – (D2) in Section (4).

(t) **Application of System Dynamics methodology for development of I*I Technology**: System Dynamics is a methodology developed for modeling systems, where very little is known about how variables interact. In particular, it is an inductive-deductive modeling methodology to look at complex systems. It can be applied to I*I Processes to develop I*I Technology.

Briefly, every material object contains no less than an infinity of variables (i.e., facts, which are modeled as data, and, when processed, as information), and therefore, possible (unlimited) systems. Further, each system, its sub-systems, and their components have their own (i.e., local) environment, a situation and problem-solving goal. Given unlimited systems (alternatives), what is, therefore, required is to *cull out*, i.e., decide and organize – not necessarily physically, but mathematically – and study data (facts) and information that are relevant to the local environment, local situation, and local goal (usefulness aspect), so as to design and deliver a flexible, customized information decision. This modeling principle, which is behind the open system view of the enterprise (Section (3)), leads to an informational view presenting system as a potential source of information and describing it "as a network of information variables in causal relationship to one another and in situations even to themselves" (Refer: Section (4)).

This informational view of the system, which is also an environmental view, models an enterprise as a Business Process *IS* View, which comprises decision stages (D1 – D7) originating appropriate information decisions. Once again, each decision stage has a requirement to originate problem information, problem environment information, and endogenous alternatives information, thereby implying having its own environment. This suggests informational view for each decision stage. Given the ever changing informational environment, then each information decision is to be chosen from possible options (Section (10) Item (h)). To simplify, let us consider any decision stage (Di) from decision stages ([D1 – D7]).

The issue is the decision stage (Di) has its own environment variations knowledge factors and relationships between these factors affect the information decision from (Di), which results in its loss of I*I. The main concern, therefore, is that decision model is able to affect a control law (i.e., decision rule or integrity rule), which is based on some aspect of the decision system state(s), so as to control itself (decision stage) in the face of external variations (e.g., implications of non-critical, interdependent environmental factors). This requires that the decision stage modeled as *IS* is closed, a feedback loop is created, and a dynamic loop is ensued.

System Dynamics methodology works within this framework of feedback loop at decision stage so as to develop I*I technology.

(u) **Information origination resource management**: While distinguishing between rational information decision and correct information decision-making process, Section (8.1) emphasizes centrality of efficient use of resources in the correct information decision-making process to acquire control, i.e., the ability to direct means to ends.

Correct information decision-making process is an information origination process and, as explained vide Section (8.3) Equation (1), the control is exercised by controlling I*I. To achieve this, it is required to commit resources for generating I*I standards identified at Section (10) Item (s). It is clear even these standards would have their environments, which will have variations. To mitigate implications of these environmental variations, there would be further need for information origination, which will call for effective and efficient management of information origination resources by the way of data streams to be attended, problem and activity prioritization, and action scheduling. This resource management, which will have to be undertaken in the midst of reality of time pressure (real world always creates time pressure), constraints, side-effects, etc., is key to implementing Integrity *IS* for competitive advantage.

11. Information Integrity Research & Education Issues and Possibilities – A General Topography

This interdisciplinary view of I*I research and education issues then focuses on the study of accuracy, consistency and reliability of information system and of information processed by it. In other words this brings in the research queries of information content integrity, information process integrity and information system integrity: their definitions, measurements, and methods and technologies for their enhancements.

The issue of *Information Integrity* is valid at the problem information origination stage (includes information origination in respect of operable goal, conflicting goal(s) inclusive), at the problem environment information origination stage, at the endogenous alternatives information origination stage, at the flexible information decision origination stage, and at the flexible information decision processing stage for control implementation. It is valid at the data origin stage, at the communication channel stage (comprising medium and/or people), at the processing stage and at the output, i.e., at the information *use* stage. Thus *Information Integrity* formalizes environment as major factor in decision-making and goes beyond the subject matter of data integrity and further covers the requirements of process integrity, medium integrity, people integrity and the output integrity; all these requirements together ensuring the system integrity. As a result, *Information Integrity* emerges as a holistic and fundamental or basic requirement of an *IS* and the information processed by it.

Seen from another angle, with reference to the information system development and implementation life cycle, the above also presents the requirements of ensuring design integrity, development integrity, implementation integrity and maintenance integrity. In fact ensuring and more importantly controlling this integrity is what delivers the competitive advantage in complex and ever changing environment. Entirety of this then offers an opportunity to understand the global nature of *Information Integrity* by studying it in new ways and finding radically new approaches for achieving it. Where it may take information science is to defining what could be termed as *Information Integrity Space* (I*I S).

When visualized, I*I Space could be described as a conceptual aggregation of different dimensions of I*I with many perspectives. All dimensions have one thing in common: they all are related to Accuracy, Consistency, and Reliability of information. The other dimensions of I*I Space are: Data forms: Numeric, Text, Graphics, Voice, Video, Digital, Analogue; Content, Process, System; Security, Audit, Control; Prevention, Monitoring, Verification, Detection, Correction; Design, Development, Operation, Use, Maintenance; Manual, Batch, Online, Interactive; Risk Factors; Economic impact; Application Focus; User Domain; Knowledge continuum: `Technical, Economic, Social, Ethical, Philosophical; etc.

Information Integrity is a pervasive universal issue, which impacts business, government, and society in many ways in this electronic information age. Today, the knowledge and understanding about Information Integrity is rudimentary, fragmented, and insufficient. On the one hand it has philosophical, sociological and economic aspects, which are yet to be conceptualized, and on the other hand it has technical, technological and industry related requirements, which only now have begun surfacing. Seen from this point, Information Integrity has potential for becoming a new discipline, a new science, even a new industry, very much like the environmental science and industry, which emerged as a result of society's concerns about the quality of air, water, and the earth.

0-0-0-0-0