

INFORMATION INTEGRITY- A SYSTEMS APPROACH TO INFORMATION SYSTEM ASSURANCE

Vijay Mandke
Research Leader,
Center for Information Integrity Research
Unitech Systems (I) Pvt. Ltd.,
B-64 (FF), Gulmohar Park, New Delhi-110 049, India.
E-mail: vijaymandke@satyam.net.in

PREAMBLE

Information systems provide support for management at all levels- operational control, management control, and strategic planning. Each of these classes of management activity includes planning, control and decision making. This makes information a *critical* organizational resource in that incorrect, inconsistent and unreliable information has an adverse effect on organizational performance.

At another level, correct and complete information requirements are *key* ingredients in planning information systems, in implementing information systems applications, and in building databases. For IS applications integrated with databases, this calls for careful planning and significant communication effort between users and information system designers.

From the IS modeling point of view, this suggests process-based description at all levels of enterprise hierarchy and activities linking IS model with user, i. e. customer, internal as well as external and direct as well as indirect. Here processes constitute the means by which enterprise delivers added value to its customers. This facilitates description of enterprise structure via processes, rather than by vertical links, via functions and activities. Users (human as well as machines) are thus fundamental source of requirements, but in the face of complex and changing environments it is difficult for users to define them resulting in incorrect and incomplete information requirements and having adverse effect on IS performance.

All this makes information system assurance through dependability or trustworthiness of information, i.e., Information Integrity critical for competitiveness.

QUALITY NECESSARY BUT INSUFFICIENT

Traditionally, auditing has been used to check whether the various IS quality assurance procedures are being carried out correctly, and whether they are successful. Here no heed is paid to demands of customization. With globalization, international competition, and changing customer expectations, quality paradigm came in to prominence.

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'. 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.

Even then, when, quality control is considered for the production of information system, there is still another issue. Specifically, when the quality control is applied, a comparison is made between the output of the process and the original specification, and the performance of the production process is then assessed. On the basis of this assessment modifications are made in an attempt to fine-tune the operation. This calls for assessing value of information system and of information product therefrom. Unfortunately, this process is far from easy because there is lack of precise definition of the attributes of information system and information value as also of methods of quantifying them. This renders quality paradigm as applied to an information system a necessary but not sufficient condition for competitiveness.

INADEQUACY OF BUSINESS PROCESS REENGINEERING

The limitation of quality paradigm was sought to be done away with business process reengineering (BPR), a management technique that in 1990s swept through corporations and governments around the world. Propelled by the need to meet business objectives of mass-customization, agility focussed on customer responsiveness, and IT driven market differentiation for competitiveness, BPR emphasized customer-supplier models wherein processes transforming supplier inputs into outputs (i.e. products) for customers were seen as the means by which enterprises deliver added value to their customers – whether that added value consists of data, a service, or a tangible product (information product included). The key contribution of the process-based description is to link IS modeling exercise, at all levels in the enterprise hierarchy, with the customer, internal as well as external and direct as well as indirect. The structure of the enterprise IS is then described by these core horizontal links via processes, rather than by vertical links, via functions and activities. In actual system, these elements of the core customer-supplier based IS model are combined in a variety of ways. They may be repeated, paralleled, and interrelated. Further, output from one set of elements (information based customer product) may become input to another set (data based supplier input) and so on; thereby assigning interchangeable roles to supplier input, i.e. data, and customer product output, i.e. information.

Notwithstanding the above stated business objective, in practice, BPR reduced its goals to streamlining of processes and procedures and cost reduction – mainly through manpower reduction. In the process, though, BPR invested heavily on information system applications, it kept applications non-integrated, host-based, proprietary and opting to work mainly with structured and periodic information for command and control computing. In the process, the information systems got infested by problems of information overload, absence of standardization, lack of relationship between data in several applications, etc. leading to errors in IS. Also, they acquired complex structures which in the face of emerging business environments got further exposed to information failures that come with time delay and with complex fault mechanisms in hard and soft IS components. This led to low information system assurance in respect of BPR applications.

However, more significant issue is that by restricting itself to work with structured and periodic information even when IS modeling, at all levels of enterprise, is linked to the customer, the BPR has exposed (in its modeling exercise) each customer to the respective environment; in turn making the IS model an open system at all the levels of the enterprise. And it is this aspect that then makes the task of obtaining information requirements across the enterprise levels that much more difficult as these requirements are then subjected to uncertainty due to external and internal system environmental factors of 5 Cs. This understandably leads to further loss of information system assurance in respect of BPR even when, for systems assurance, information systems are equipped with advances in data-related information processing technologies of data validation techniques along with handwriting and forms design standards, formulation of check digit algorithms, verification techniques with real numbers, on-line access methods, data transmission accuracy, data security, data encryption, and many other techniques for manipulating and protecting existing data. This is because, notwithstanding the open system character of an IS, in all these technology applications for systems assurance the assumption is that data in respect of information requirements is perfect, once validated, and most information processing systems do not anticipate defective data.

In view of all these information system assurance difficulties, like quality paradigm, the business process reengineering is also found to be inadequate for meeting the demands of competitiveness in a complex and changing business environment.

TECHNOLOGICAL SHIFT TO COALITIONAL ENTERPRISE: BUSINESS ENTERPRISE REENGINEERING

What could be seen as the cause for information system assurance difficulties in BPR? Even after linking the IS model with customer at all levels of the enterprise, why has it not been possible for BPR to respond to information assurance demands of post '90s economy emphasizing business requirements of mass customization for increased markets, agility-focused on customer responsiveness? The possible answer may be found in inability of BPR to respond to demands of informational work changes that are integral

to the technological shift that has emerged as the engine for driving the post 1990s economy.

Specifically, with the technological reality of (a) availability of on-line computers; computers providing capability of moment by moment optimization of processes and decision-making; and of system integration capability, and (b) with technological reality of digital data as medium of information flow across the enterprise so as to 'put it (the supply chain) all together', it has now become possible to use information 'smarter' so as to achieve strategic and competitive advantage of larger mass-customized markets and financial optimization. For example, for a production organization, it is now possible to design a digital product. The digital product could be made to operate in the digital space, where its performance can be tested for conformance with specifications. Because product is digital, customers could be effectively brought into the network to contribute to design, so as to ensure customer requirements and any special innovational needs are incorporated and the digital product works. Further, bulk of the components and sub-systems of the product can be manufactured by suppliers who would also now constitute the part of the network. Indeed, what thus can emerge is a coalitional, networked organization or enterprise (extending beyond the parent organization) for achieving business objectives hitherto never aimed at.

Explained differently, in the event of technological shift mentioned, any business activity can be viewed as comprising informational and physical work systems. Accounting for continuous product innovation based on customer needs (product requirements) and constraints of costs and capabilities, informational work system (processing unstructured and non-periodic information) delivers flexible information decision for control implementation to process material, energy optimally, so as to deliver desired (personalized) product to the customer. Further, most importantly, such informational work systems can be applied to both hard components of production processes, machinery and equipment, and soft components like information flow and data bases.

From the analysis as above and within the framework of the technological reality of informational work system, what becomes clear is to benefit from the advantage of using information 'smarter', it is equally important, if not more, that the enterprise undertaking business activity reengineers itself by emerging as a coalitional, networked enterprise, so as to achieve business objectives (even not considered so far) so as to remain competitive. In other words, the business process should now be represented by a coalitional, networked customer-supplier model against the linear customer-supplier model as envisaged under BPR.

Thus, it can be seen that the technological shift driving the post '90s economy with emphasis on integration maximization is, as integral to it, also demanding business enterprise reengineering (BER) so as to link IS model, with a coalitional and networking structure, not only to customers at all levels of the parent enterprise but also to all customers of stake holding enterprises external to the enterprise (competitors included). In its merely cost saving attempt to work with non-integrated technology and, therefore,

with structured and periodic information, what BPR did not take into account was this enterprise reengineering requirement. In the process, it accorded no capability to the business process to generate new values (objectives) and in the process create new market space; the whole exercise of BPR remaining limited to value added business performance. When faced with the problem of changing information requirements of the emerging economy, this rendered the business process IS characterized by low information system assurance.

SHIFT FROM INFORMATION TECHNOLOGY TO INFORMATION

Coalitional, networked customer-supplier model represented as IS model, technologically amenable for processing information ‘smarter’, facilitates modeling the generic business process (and each of its activities) *emphasizing* ‘information’. This essentially facilitates viewing every business activity (whatever else it does) as an IS activity (a cybernetic view), wherein ‘data’ from the data origin stage is seen as raw material, ‘data transformation’ at processing stage as a system function characterized by pre- and post- processing communication channels (which include medium and/or people), and ‘data product’ or ‘information’ from the output, i.e. information use stage, as processed data used to trigger information use, so as to deliver ‘information decision’ for controlling the physical process characterizing the activity. In other words, data and information for use are different entities, and database is seen together with data acquisition and information utilization cycles.

Under BPR, informational work revolves around processing of structured and periodic information and has stages of planning of actions and decision making. Against this, the technology shift requiring BER forces a rethinking of value chain. Specifically, the IS model representing the coalitional, networked customer-supplier activity now has the capability to *generate* value so as to create new market space. This *adds* stages of long term and operable goal setting, of formulation of models and information gathering for the purpose, and of prediction and extrapolation to the informational work systems, so as to account for processing unstructured and non-periodic (flexible) information in the wake of changing information requirements.

What is interesting is each of above informational work stages by themselves are also information processing stages. This introduces a very significant dimension in the research query pursued in that the problem challenge shifts from one of dealing with information technology to that of information. As explained, this information is at different levels. At one level, information refers to the potential message in an entity or event, or in reports about it. Information is viewed, in this context, as ‘data’, i. e., as a function of the source only. At another level, it refers to transmission of message; as a function of both source and means of conveyance as in communication theory which uses probability to quantify the properties of symbols to convey message. Its primary value in studying information systems lies in the key ideas of probability and reduction of uncertainty and notions of noise, lag and error in transmission. And at a higher level, information refers to the meaning gained by the recipient customer as well supplier (human as well as machine); the extent to which uncertainty is reduced and recipient

knowledge increased. Information, in this context, is a function of source, communication channel and the specific recipient.

What this then calls for is an information development and implementation life cycle approach for unstructured and non-periodic information – a structural variant from the days of structured and periodic information implementation. The life cycle would comprise stages of information origin, storage, verification, manipulation, validation, use and discard. The challenge is to undertake this life cycle approach involving information development and implementation stages for activities of long term and operable goal setting, of formulation of models and information gathering, and of prediction and extrapolation, so as to set up IS models for planning actions and decision making for control implementation for physical work systems.

And, in presence of complex and changing system environment (external as well as internal), each of information development and implementation life cycle stages as also IS development and implementation life cycle stages for the business process IS are characterized by 5 Cs' based uncertainties of their own kind; thereby further increasing the requirement of information system assurance by researching the dependability or trustworthiness of information, i.e., Information Integrity.

TOWARDS A WORKABLE DEFINITION OF INFORMATION INTEGRITY ATTRIBUTES

Auditing, quality paradigm and BPR all are found to be inadequate to ensure information system assurance for IS for competitiveness. Information systems compete for organizational resources against alternative uses for these resources. Therefore, the value or integrity of information systems and information therefrom must be assessed to compare with estimates of cost. This tends to be difficult. As mentioned earlier, apart from the implications of the system environmental factors of 5 Cs, one difficulty in ensuring information system assurance is lack of precise definition for information system value factors.

Initial efforts to answer this query are to be seen in accounting literature that places emphasis on internal systems and audits. Indeed, as early as 1968, Feltham identified accuracy, timeliness and relevance as the three dimensions of data quality and analyzed their relationship with the value-in-excess-of-cost criterion within the context of the data consumer. The current literature on data/information quality considerably expands this list of information value or integrity attributes to include requirements of: accuracy, usability, reliability, independence, timeliness, precision, completeness, relevance, sufficiency, understandability, freedom from bias, consistency, trustworthy, brief, etc.

With military requirements dominating the research in information systems, the issue of secured computer systems and of confidentiality of information gained a high priority query. Security has normally been taken to mean confidentiality, integrity and availability *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) such as: validity of information in a computer system; prevention by customer of unauthorized program from: circumventing memory protection mechanisms, avoiding access control mechanisms and obtaining supervisor state; preventing unauthorized modification of data; correctness and protection of Trusted Computing Base and Procedures; reliability, accuracy, faithfulness, non-corruptibility and credibility of information transmitted, protection from noise in communication systems and equipment failure, consistency, accuracy, concurrency, data recovery, modification access, credibility of information ; internal consistency of a system (a correctness aspect) and external consistency (correspondence with reality – an appropriateness aspect); property of data which has not been subject to unauthorized alteration or destruction; property of internal correctness of a system - error control objective, and property of correspondence to the real world – fraud control objective; correctness and consistency of data and their classification labels; etc. Further, there is a recurring theme in the literature that integrity and security are different and that the assumption of trusted computing base and procedure is unacceptable as the trust must be demonstrated.

Similar things are to be observed with reference to commercial IS applications. For example, in case of hardware components or subsystems, the assurance factors to be assessed include not only functionality, but also more complex issues such as reliability, maintainability and safety. And for high integrity systems the required values for some of these factors may not be measurable, thereby making quality control a difficult task. With software the situation may be even more complicated. Firstly, it is generally impossible to test a program completely to determine its correspondence to the functional aspects of the specification. Secondly, the requirements of software will include features such as reliability, efficiency, portability, which may be extremely hard to quantify. Assurance control depends on ability to measure parameters within the product that is desired to be controlled. Unfortunately, in highly complex systems this is often very difficult.

Finally, many standards related to system assurance and control are concerned exclusively with software. This reflects the general view that it is within software that most problems exist. However systems assurance or integrity is a global issue and has implications for all aspects of IS.

Within the framework of above issues, of cybernetic view of an activity, and of requirement that information is for use, that then a workable definition of intrinsic Information Integrity attributes can be considered and they can be defined as: Accuracy – the degree of agreement between a particular item (value, algorithm, configuration) and an identified source (standard); Consistency – satisfying constraints, i. e., the degree to which multiple instances of the item satisfy a defined domain; and Reliability. These are intrinsic or objective or basic Information Integrity attributes in the sense all information systems must demonstrate them. As information is for use, depending on IS application area and independent of any specific user, there will be different requirements/standards of information. It is understood that the intrinsic attributes would satisfy these different application area specific industry standards. Further, it is suggested that for data/information model represented by a triple $\langle e, a, v \rangle$, this identification of intrinsic

attributes is workable in that it allows segmenting integrity issues pertaining to those concerning entities, attributes and values.

Coming to the security, it may be seen to have two aspects; namely, undamaged information and confidentiality. Security implying 'undamaged' is synonymous with the requirements of accuracy or correctness, and to that extent is synonymous with the intrinsic requirement of accuracy which is seen as integral to the definition of Information Integrity. As regards to the confidentiality aspect, it can also be seen from the point of view of 'privacy'. It is submitted that both the attributes of confidentiality and privacy, though they emerge as implications of errors in IS, i.e. low information system assurance, may not be central requirements for all information systems, as there can be information where confidentiality and privacy may not be required. In other words, attributes of security, meaning confidentiality, and privacy may be seen as optional to an information system and depend on the specific context and nature of use of information. Within this frame of reference then, security, privacy can be seen to present extrinsic or subjective attributes of Information Integrity. As indicated earlier, there can be other extrinsic Information Integrity attributes, too, such as usability, independence, sufficiency, understandability, freedom from bias, brief, faithfulness, non-corruptibility, credibility, concurrency, credibility, etc.

INFORMATION INTEGRITY – A SYSTEMS VIEW

This visualization then focuses the research issues in Information Integrity to 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.

As observed by Rajaraman, the issue of *Information Integrity* 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* goes beyond the subject matter of traditional view of data integrity to include integrity of stages of long term and operable goal setting, of formulation of models and information gathering, and of prediction and extrapolation, and further covers the requirements of process integrity, storage 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 life cycle, the above then also presents the requirements of ensuring design integrity, development integrity, implementation integrity and maintenance integrity.

There can be yet another detailed angle to look at Information Integrity from a more holistic point of view. Specifically, a computerized information system (*CIS*) can be viewed (modeled) from different angles; namely, (a) operating elements based *IS* model (view), (b) decision support activity based *IS* model (view), (c) management activity

based IS model (view), and (d) organizational function based IS model (view). Each of these views are described by different components. Specifically, the operating elements based view of an computerized IS (CIS) comprises of:

- Components of Operating Elements:
 - Physical components of an IS:
 - *Hardware for:* Input or entry, Output, Secondary storage for data and programs, Central processor (computation, control, and primary storage), Communications
 - *Software:*
 - System software,
 - Application software
 - *Databases:*
 - File
 - Physical storage of media (computer tapes, disk packs, diskettes, etc.)
 - *Procedures*
 - User instructions (for users of the application to record data, employ a terminal to enter or retrieve data, or use the result)
 - Instructions for preparation of input data by data preparation personnel
 - Operating instructions for computer operations personnel
 - *Operations personnel*
 - Computer operators, system analysts, programmers, data preparation personnel, information systems management, data administrators, etc.
 - Tasks these physical components do:
 - *Processing functions*
 - *Process transactions*
 - *Maintain master files*
 - *Produce reports*
 - *Process inquiries*
 - *Process interactive support allocation*
 - Information input that an information system made up of these physical components processes:
 - *Input Data which could be:*
 - *Observations*
 - *Transaction documents or screens*
 - *Inquiries/questions*
 - Form in which Information input received:
 - *Text*
 - *Voice*
 - *Video*
 - *User-machine dialog results*
 - Information output for users:

- *Transaction documents or screens*
- *Preplanned reports*
- *Preplanned inquiry responses*
- *Ad hoc reports and inquiry responses*
- *User-machine dialog results*

Against this decision support activity based view of *CIS* would consist of:

- Components of Decision Support Activities:
 - Structured, programmable decision making, and
 - Unstructured, non-programmable decision making.

As regards to management activity based view of a *CIS* , it can be seen to consist of:

- Management Activities of:
 - Strategic planning,
 - Management control and tactical planning, and
 - Operational planning and control.

Finally, the organizational function based *IS* model (view) of a *CIS* would consist of:

- Organizational functions:
 - Sales and Marketing,
 - Design & Development,
 - Production,
 - Logistics,
 - Personnel,
 - Finance and accounting, and
 - Top management.

The point is, each of the components of the different *IS* views as above are prone to failures due to system environmental factors of 5 Cs; and, hence, need to ensure integrity.

From above it follows that low information system assurance has its origin in a wide variety of errors associated with enterprise-wide generic functions: designing, developing, implementing, operating, maintaining, communicating, managing, and the like. As mentioned elsewhere, one important point is that these errors may be latent (that come with delay) or active; that is they may be complex errors. Therefore the etiology of enterprise-wide errors divides into five phases: (a) enterprise function, process, activity, procedure, or task giving rise to latent errors, (b) the consequent creation of error- and violation-producing conditions within specific work-places (design board, production rooms, test centers, processing units, maintenance activities, etc.), (c) the commission of errors and violations of various types including human errors, in respect of sharp end tasks (and in the days of data driven technology keyed to the flow of information across the supply chain, every information processing across the *IS* development and implementation life cycle becomes an operation at sharp end), (d) events in which one or more of the various defenses and safeguards are breached or by passed, and (e) outcomes

that can vary from a “free lesson” to loss of customer to a catastrophe – all forms of varying consequences of effects of low system assurance due to lack of I*I .

This totality then gives the systems view of Information Integrity.

0-0-0-0-0