I*I attributes –
Accuracy, Consistency, Reliability
(Quantifying I*I and it's attributes)

Vijay V. Mandke

Research Leader,

Center for Information Integrity Research,

Delhi Center: B-64, Gulmohar Park, New Delhi – 110 049,

Pune Center: Flat A-2, Nikash Skies, Someshwar Wadi, Pashan, Pune-4110 08

Visit us at: centerforinformationintegrityresearch.org

Information Integrity/Integrity Information System/Management Information System
Course Lecture (s) # 48-49

2006-2007

Lecture # 48-49
I*I attributes –
Accuracy, Consistency, Reliability
(Quantifying I*I and it’s attributes)

RECAP…1

- In Lecture # 25, we saw that UUI paradigm presented information requirements of usefulness, usability and integrity as the determinants of information value. Further, we saw that information requirements of usefulness and usability must also have integrity.
- This presented integrity as all pervasive information requirement, and rendered integrity attributes of accuracy, consistency and reliability fundamental or basic to the value of information; and as a result a critical requirement of an *IS*.

RECAP…2

- As information value can be seen to define information use (IU) quantum, within above framework, the control of integrity objective then can be seen to optimize "IU" quantum for a given information processing situation, so as to offer competitive advantage.
- For an analytical understanding of the above thesis (which incidentally is the subject matter of Lecture # 27), it will be helpful to first have a clearer view of integrity attributes of Accuracy (A), Consistency (C), and Reliability (R).

I*I ATTRIBUTES OF A, C, R

- Contrary to information “exactness” requirement as pursued by traditional Integrity research (Ref.: Lectures # 7, 8, 10, 24), the search here is for “*correctness*” aspect of information requirement.
- For describing Information Integrity attributes (of A, C, R), this suggests need to define “information error”, which was the subject matter of Lecture # 14.

ACCURACY (A) -1

- Within above framework then Accuracy attribute (A) is defined as the degree of agreement between a particular value and an identified source.
- It can be assessed by identifying the relevant established source (standard) and by determining an acceptable tolerance.
- Specifically, the identified source provides the correct value – preferably the value corresponding to the optimum Integrity. 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 database.

ACCURACY (A) -2

- Given that value of data/information is expressed in a numerical, accuracy of the data/information can be quantified in a number of ways:

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

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

ACCURACY (A) -3

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

iv) Number of records examined: R

Number of records with at least one defect of loss of Accuracy: D1

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

ACCURACY (A) - 4

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

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

ACCURACY (A) -5

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

Number of records examined: R

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

ACCURACY (A) - 6

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

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.

CONSISTENCY (C) -1

- Consistency is the degree to which multiple instances of a value satisfy a set of constraints. The multiple instances may exist across space (such as databases or systems) or over time.
- Thus, consistency is with respect to a set of constraints and data/information is said to be consistent with respect to a set of constraints if it satisfies all constraints of the data/information model.

CONSISTENCY (C) - 2

- 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).

CONSISTENCY (C) - 3

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

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

RELIABILITY (R) –1
IT'S DUAL INTERPRETAION

- Reliability, which traditionally is a large concern in the system development lifecycle model, is a little complex attribute to define as it has a dual meaning in modern technical usage.
- Interpretation 1: In the broad sense, it refers to a wide range of issues relating to the design of large systems (complex computerized information system [CIS] included), which are required to work well for specified periods of time.

RELIABILITY (R) - 2
IT'S QUALITY CONNECTION

- In such a case, the term "reliability" includes descriptors such as "quality" (commonly understood from the traditional "standard' product angle) and "dependability" (Reliability Analysis), and is interpreted as a qualitative measure of how a system matches the specifications and expectations of a user.
- From this point of view for an IS the definition of reliability given as "accuracy with which information obtained represents data item in whatever respect the information system processed it" can be seen to define the reliability requirement for the IS as a whole; reliability index being amenable to quantification through techniques such as Analysis of Variance (AOV).

RELIABILITY (R) –3: IT'S COMPLETENESS CONNECTION IN PRESENCE OF NOISE (EXACTNESS REQUIREMENT)

- Interpretation 2: Then, as mentioned in Lecture # 10, reliability is also seen as 'completeness' issue. Of course, the completeness requirement itself has two different aspects. One is that of "exactness" requirement.
- This narrower requirement occurring due to the ever-present system "noise" is the main concern in communication theory and in security research as also in the "standard" product in high volume seeking business models under quality paradigm emphasizing "reduced defects" in system processing.

RELIABILITY (R) – 4: IT'S COMPLETENESS CONNECTION IN PRESENCE OF NOISE (EXACTNESS REQUIREMENT)

- Here, reliability is a measure denoting the probability of the operational success of an item under consideration. The notion of reliability, in this case, may be applied to a single component (e g., a diode or a light bulb); a complex system (e g., an aircraft, a computer or a network of computers); a computer program; a procedure (e g., conversation between a pilot and an air-traffic controller); an element of an IS; namely, data, i. e., IS input, or IS output which is "data processed", i.e., information; or even a human.

RELIABILITY (R) – 5: IT'S COMPLETENESS CONNECTION IN PRESENCE OF NOISE (EXACTNESS REQUIREMENT)

- Specifically, reliability analysis is concerned with occurrences of undesirable or unanticipated events during the course of operation of a system or an item and the impact of these events on the system's behaviour or the item's use. And the undesirable events may be failures of components (and, for information systems, failures of resulting data and information in the form of data/information errors) caused by deterioration or wearing out of components due to age and usage or even design problems and inadequacies, etc. surfacing in the course of the use of the system.

RELIABILITY (R) – 6: IT'S COMPLETENESS CONNECTION IN PRESENCE OF DISTORTION (CORRECTNESS REQUIREMENT)

- There is another aspect of "completeness" requirement, though. In the form of "observability", it is to be found in system theory.
- Specifically, the problem considered is that of state variables derived based on measured system outputs at several times and the knowledge of the system-forcing function (control) effort. It is conceivable that the structure of the system and/or measurements taken is such that the measurements do not contain *all* the information about the system states.

RELIABILITY (R) – 7: IT'S COMPLETENESS CONNECTION IN PRESENCE OF DISTORTION

(CORRECTNESS REQUIREMENT)

- The usual technique in systems engineering is to generate control efforts (strategies) based on measurements of system outputs. If the measurements are *missing* basic information on actual system response (that is, if there is information distortion), erroneous control efforts could be generated, which is not desirable; just as, if, in the *IS*, value of information element is missing from the information record, the desired information use (IU) value is not achievable, however high may be the information usability factor.

RELIABILITY (R) – 8: IT'S COMPLETENESS CONNECTION IN PRESENCE OF DISTORTION

(CORRECTNESS REQUIREMENT)

- In In other words, when concerned with reliability factor under *correctness* requirement of information, there are incompleteness issues due to "noise" and "distortion". For the purpose of the investigation at hand, whether "inexactness" due to the 'noise' factor or "incorrectness" due to 'distortion' factor, both result in information item exhibiting error and therefore loss of integrity. As a result, reliability attribute of "*correctness*" aspect of information requirement in considering 'completeness' must account for both these possibilities.

RELIABILITY (R) – 9: IT'S COMPLETENESS CONNECTION IN PRESENCE OF DISTORTION

(CORRECTNESS REQUIREMENT)

- It is within this framework then the reliability (R) can be heuristically defined as follows: Reliability (R) refers to completeness, currency and auditability of data/information. Specifically, data/information is complete when all component elements are present (effects both of distortion and noise are counted). Information is current when it represents the most recent value. And, information is auditable if there is a record of how it was derived and that record allows one to trace information back to its source.

DEFINING RELIABILITY (R) –1: R1 & R2 (CORRECTNESS AND EXACTNESS ASPECTS OF INFORMATION REQUIREMENTS)

- For For the analytical convenience, let us denote the reliability attribute defined based on "completeness" perception, which accounts for 'distortion' as mentioned above, by "R1".
- Further, let us denote the reliability attribute, which is with reference to "noise" factor and is based on "exactness" perception, by "R2".

DEFINING RELIABILITY (R) –2 R1 (CORRECTNESS REQUIREMENT)

- Within above framework then, using the earlier described concept of percent defective, given the number of records examined (R) and the number of records with at least one defect of incompleteness, lack of currency or inability of auditability (D2), the reliability index (R1) can be quantified as follows.

$$\text{Reliability Index (R1)} = \left[\ 1 - \left(\frac{D2}{R}\right)\right]$$

DEFINING RELIABILITY (R) –3 R2 (EXACTNESS REQUIREMENT)

- As regards to reliability defined by “R2”, it can be calculated by using reliability analysis methods.

DEFINING RELIABILITY (R) –4

- Then, the reliability attribute “R” should cover both “R1” and “R2” and is given by Equation (1).

Reliability attribute index = R = R1 x R2

.. Equation (1)

INFORMATION INTEGRITY (I*I) BASIS FOR IT'S QUANTIFICATION

- What are defined are attributes of Information Integrity (I*I) for information value, i.e., for the content of information processed by *IS*. Content of information would, therefore, have I*I value as in Equation (2).

- Information Integrity = I*I = A x C x R1 x R2

.. Equation (2)

Refer to following papers, which are uploaded onto the HSS Website for the “Reading material Link”:

- 1. Refer to **Section (3)** of Paper by *Mandke Vijay V. and Nayar Madhavan K.; 'Implementing Information Integrity Technology' - Updated, Special Issue of International Journal of Informatica, June 2002.*
- 2. Refer to **Section (7)** of Paper by *Vijay V. Mandke and Madhavan K. Nayar, “Cost Benefit Analysis of Information Integrity”, Proceedings of the 2002 International Conference on Information Quality, MIT, Cambridge, Massachusetts, USA, 8-10 November, 2002.*

EPILOGUE-1

- The current Lecture # 26 provides the basis for quantifying I*I attributes and I*I.
- With determination of I*I attributes and I*I, one can now proceed with cost-benefit analysis of I*I, which is the subject of Lecture # 27.
- The need for *IS* model in the form of cost-benefit analysis equation of I*I comes because it is what will provide an analytical means to compare and select between two or more values of I*I, i.e., optimum, i.e., correct, i.e., standard value of I*I.

EPILOGUE-2

- That will analytically and scientifically establish the ever increasing importance of I*I (in Lecture # 7 we first argued it qualitatively) to control systems (comprising informational and physical work systems) to achieve competitive advantage as organizations expend increasingly more resources on information technology and information in a complex and ever changing environment.
- This is the task we set before us in Lecture # 12, Slide # 52 (Also see Lecture # 6, Slide # 34).

THANK YOU