

# Research in Information Integrity : A Survey and Analysis

Vijay V. Mandke<sup>1</sup>

## Abstract

*With unprecedented advances in information technologies and with shared data environments, organizations and individuals are increasingly concerned with capture, use and control of growing, complex and diversified volumes of data. However such databases are invariably affected by data errors; in turn requiring critical attention to the question of impact of data defects on the working of information processing systems. The paper submits that this essentially is the question of researching methods for data repairs and for enhancing data integrity of information systems. With this objective, the paper surveys research effort in Information Integrity starting from early studies in the area of Electronic Data Processing on data errors and Svanks' suggestions for research thrust in areas of data defects and data repair and Integrity Analysis Methodology. This is followed by survey of researches in the areas of Accounting, Data Quality attributes and Total Data Quality Management and Technology. Based on this research survey, the paper in the end presents suggestions for research initiatives in Information Integrity in terms of : Defining Information Integrity, Integrity attributes - Objective and Subjective, Designing of Information Systems with Integrity measures, Systems Technologies for Information Integrity and other research areas in terms of Integrity Management.*

## 1. Introduction : In search of Data Era

Electronic data processing (EDP) came into being in the mid 1950's for the capture, use and control of increasingly growing, complex and diversified volumes of data. Within two decades to follow, i.e. by 1970's, in the design and operation of Information Processing Systems, the hardware reliability issue appeared to be settled and software-related research was in full swing and there was considerable talk of the dawning of data era. It is 1995 today and if one is to go by the statistics on abuses of computers and by vulnerability of computerised information systems of today, this era is yet to arrive.

---

<sup>1</sup> Director, School of Engineering & Technology, Indira Gandhi National Open University, 198, Asian Village Complex, New Delhi - 110 049

## **2. Early Studies on Data Errors : Experimenting with Real Data**

In the early days of computing, numerous studies were carried out on errors in data transcription. Experiments were devised where groups of people were assigned the tasks of encoding, writing, copying, keying, and reading large volumes of data and results of these experiments were then analysed to determine the frequency of various types of errors. Many commonly used data validation techniques along with handwriting and forms design standards, evolved on the basis of these findings [1].

Another popular research area was the formulation of check digit algorithms, supported by verification studies with real numbers. These studies yielded confidence levels for the various algorithms. For example, it was established that the MOD10 check digit incorporated into the Canadian Social Insurance Number (SIN) is only 74.8% reliable in detecting keying errors, whereas the commonly used MOD 11 check digit is 98.7% reliable (Wooldridge 1974) [2].

Data encoding schemes for replacing descriptive and narrative text, handwriting and forms design standards, programmed data validation procedures, batch balancing, hash totalling, check digits, and many other control techniques developed in 50's and 60's still constitute the present repertoire of methods for enhancing data integrity.

## **3. Lack of research continuity in Data Science**

In her significant research work of 1984 entitled "Integrity Analysis : A Methodology for EDP Audit and Data Quality Assurance", Dr. Maija I. Svanks [1] points out that this type of experimentation with real data has long since ceased. Later advances in data-related EDP technology pertain to on-line access methods, data transmission accuracy, data security, data encryption, and many other techniques for manipulating and protecting existing data. Emphasis has, therefore, shifted from the research and technology for the enhancement of data integrity to preserving its status quo. The assumption prevails that data is perfect, once validated, and most Information Processing Systems (IPS) generally do not anticipate defective data.

This has certainly brought in a kind of stagnation in the state-of-the-art of data integrity. More specifically, Svanks observes "E" in EDP, i.e. Equipment or hardware design employs the principles of well-established sciences such as electronics, electrical engineering and physics.

"P" in EDP, i.e. Programmes or software have, likewise, been the subject of extensive research, and incorporate the abstractions of mathematics and logic. The recognised discipline of software engineering and the increasing formalism of programming methodologies, computer language structures and syntax, data

access algorithms and data base management models are rapidly narrowing the gap between art versus science of software design.

However, when one comes to “D” in EDP, i.e. the state-of-the-art of data technology, Svanks argues that growing preoccupation with innovations in EDP areas such as software, hardware, data base management systems, telecommunications, distributed processing, mini and micro computers, data encryption and office automation has severely detracted from evolution of a “data science” based on type of experimentation with real data mentioned earlier. In fact, Svanks submits that a formal data discipline does not yet exist and that current approaches to data preparation, validation, repair, and quality assurance are inadequate, ad hoc, and lack comprehensive methodologies. Even more serious, critical Information Processing System Logic such as data base updating and information retrieval largely assumes perfect data. In the absence of integrity-oriented techniques, result reliability cannot be guaranteed and defective data leads to further data contamination. In short, defective data used to drive “perfect” programmes may cause both programme errors and additional data defects.

#### **4. Svanks’ suggestions for research thrust**

It is within this framework, vide her 1984 research work cited above [1] Svanks identifies following research thrust for answering the data integrity problem at hand; namely,

- a) Development of conceptual models, proven mechanisms and generalised software for :
  1. The disclosure, classification and analysis of data defects,
  2. The quantification, evaluation and monitoring of data base integrity,
- b) A formal approach to data verification in lieu of ad hoc practices; that is, the use of standard repertoire of techniques where the probability of defect disclosure is known for each, and the composite effectiveness of data verification becomes measurable,
- c) Development of facilities for data repair to support both operations and quality assurance,
- d) EDP audit and quality assurance methodologies for the review and assessment of operational IPS and of diverse management control systems such as Management Information System (MIS), Decision Support System (DSS), and Information Resource Management (IRM), and,

- e) Development of standards and guidelines for integrity-oriented IPS design and operations; in particular to anticipate and deal with defects in data used to drive programme logic.

Further, she presents for the research objective at hand, a practical and powerful methodology of “Integrity Analysis”, which can work as a data inspector for information (product) control and serves as an “automated (process) auditor”.

## **5. Svanks' Integrity Analysis Methodology**

Specifically, Svanks defines “Integrity Analysis” as “the use of specifically designed software which periodically and in a systematic manner examines data bases, or relevant samples thereof, for undesirable conditions and discloses evidence of Data Defects for manual and automated analysis and for remedial action”. The Integrity Analysis Methodology presented by Svanks has following nine steps :

- Step 1 : Familiarisation with the IPS and its data,
- Step 2 : Specification and documentation of the data rules or constraints to be implemented in an integrity analysis task (Note : Violation of a constraint is a Data Defect. Further, each type of defect is assigned a unique code),
- Step 3 : Development of Integrity Analysis programme; that is, constraint programming,
- Step 4 : Development of sampling programme,
- Step 5 : Development of the reporting facility; that is, a subsystem for sorting, tallying, quantifying, summarising, classifying, reporting, retrieving, and otherwise processing data defects disclosed by Integrity Analysis,
- Step 6 : Sampling the IPS data base (optional),
- Step 7 : Integrity Analysis of the sample or of the entire data base,
- Step 8 : Printing Integrity Analysis results, and
- Step 9 : Defect diagnosis, analysis and reporting to the Integrity Analysis client or management.

## **6. Research in Accounting Literature**

Work of Svanks is the product of body of research in the accounting literature that places a specific emphasis on internal systems and audits. Specifically, as early as 1968, Feltham [3] has identified relevance, timeliness and accuracy as the three dimensions of data quality and analysed their relationship with the value-in-excess-of-cost criterion within the context of the data consumer. In 1973, Yu. & Neter proposed one of the earliest stochastic models of an internal control system that could serve as the basis for an objective, quantitative evaluation of the reliability of an internal control system [4].

Following this, Burns & Loebhecke (1975) have studied a tolerable data error compliance level [5], while Bodnar (1975) considered application of reliability engineering techniques to internal control systems analysis [6]. Accounting literature continues to report publications after 1984, too, and they cover works of : Fields, et al (1986) who present models of internal control process that respond to guidelines and regulations calling for auditors to evaluate management effort to assure that accounting data are correct [7]; of Amer, et al (1987) reviewing information systems research as applied to accounting and auditing [8]; of Groomer and Murthy (1989) who address to the question of data quality as it relates to audit populations [9].

Researches in accounting and auditing as above, where objective is maximum reliability with minimum cost, the key data quality dimension used is accuracy defined in terms of the frequency, size and distribution of errors in data. Further, as mentioned above, researchers have also identified relevance and timeliness as desirable attributes of information.

## **7. Research in Data Quality Attributes**

Query as above in desirable attributes of information, has also given rise to body of research on evaluating information systems success from the user point of view. Thus Halloran, et al (1978) propose attributes such as usability, reliability, independence, etc. [10]; Bailey & Pearson (1983) suggest data quality attributes of : accuracy, timeliness, precision, reliability and completeness [11]; King & Epstein (1983) empirically investigate linear models for quantifying information systems value, attributes studied being reporting cycle, sufficiency, understandability and freedom from bias [12]; and, in 1987, Agmon & Ahituv apply reliability concepts from the field of quality control to information systems by developing three measures for data reliability; namely, (a) internal reliability (the “commonly accepted” characteristic of data items), (b) relative reliability (the compliance of data to user requirements), and (c) absolute reliability (the level of resemblance of data items to reality) [13].

In the Data Quality area, mention may also be made of works of Morey (1982) who estimates the “true” stored error rate [14]; of researchers at SUNY, Albany who define : (a) accuracy which occurs when the recorded value is in conformity with actual value, (b) timeliness which occurs when recorded value is not out of

date, (c) completeness which occurs when all values for a certain variable are recorded, and (d) consistency which occurs when the representation of data value is the same in all cases [15, 16, 17]; and of works of the team from AT&T which identifies more than twenty dimensions of data quality, including accuracy, completeness, consistency and cycle time [18].

Apart from various research publications as above, 1980s also report publications of books in the area of Information Integrity. In this context, mention may be made of books by Becker B.B. entitled "Information Integrity - Structure for its Definition and Management", [19] and by Lane V.P. entitled "Security of Computer Based Information Systems" [20].

Thus researchers have come forward with different perceptions of Information Integrity - from those in terms of differing attributes of Data Quality to that where information integrity is seen to serve requirements of control (data exists, data possesses certain degree of quality), privacy, security and disaster recovery. To ensure these requirements, while comparing "integrity" with "insuring" the system, Backer (1983) outlined Risk Identification Procedures, Approach to Information Integrity Surveys and Risk Analysis and Risk Management methods through physical and logical integrity measures.

In May 1990 a MSc Thesis in Management by Heber Regal Norckauer, Jr. is reported at MIT, Cambridge, MA. The Thesis deals with "Duplicate Entry Detection in Mailing and Participation Lists". The Thesis resulted in more efficient algorithm in terms of mature rule based expert system applicable for the problem at hand in United States Postal Service and number of members of the Direct Marketing Industry [21].

Then in March 1991, Hansen and Wong at Sloan School of Management published a working paper entitled "Managing Data Quality : A Critical Issue for the Decade to come". The paper presented a study of data quality while introducing analogy between manufacturing and information systems, defining the dimensions of data quality and developing the concept of data value chain. The study thus revealed data quality along four dimensions : Accuracy, Interpretability, Availability and Timeliness. Majority of field study participants also expressed difficulty tracking down the sources of their data quality problems [22].

## **8. Research in Total Data Quality and Information Management & Technology**

Subsequent to this in 1992 one comes across two important publications by AT&T entitled "Data Quality Foundations " and "Improving Data Accuracy : The Data Tracking Technique". More specifically, the publication "Data Quality Foundations" presents Data Cycle Model in terms of two basic cycles : acquisition and usage, and then gives Quality Dimensions of Data Values which,

based on AT&T research and accounting for customer requirements, are defined to be : Accuracy, Timeliness, Completeness and Consistency [23].

Against this the AT & T publication “Improving Data Accuracy : The Data Tracking Technique”, presents a technique to see where distortion occurs in data bases. This technique is based on following seven steps [24] :

- i) Establish process management responsibility,
- ii) Define process and identify customer requirements,
- iii) Define and establish measures,
- iv) Assess conformance to customer requirements,
- v) Investigate process to identify improvement opportunities,
- vi) Rank improvement opportunities and set objectives, and,
- vii) Improve process quality.

All this work of AT & T has then been comprehensively presented by Thomas C. Redman of AT & T in his excellent book published in 1992 titled “Data Quality : Management and Technology”. Treating “Information” as a product, Redman basically applies the well known Statistical Quality Control method for measurement and improvement of Data Quality attributes of Accuracy, Timeliness, etc [25].

Thus works of Svanks and Redman open other avenues of quantitative techniques to answer the question of measurement and improvement of data quality.

Following this in December 1992, Wang and Reddy at Sloan School of Management, MIT, USA have published paper titled “Quality Data Objects” [26], and, in Spring 1993, Madhavan Nayar [27] has paper in “Information Systems Management” titled “Achieving Information Integrity - A Strategic Imperative”. Specifically, Wang and Reddy maintain that behaviour of the quality object includes a set of methods to measure quality dimensions (such as timeliness, completeness, credibility) and claim to develop a quality data object algebra.

Following this, there are a series of publications of Sloan School of Management under its Total Data Quality Management (TDQM) Research Project and few are enumerated here :

- a) Kon, Lee, Wang, “A process view of Data Quality”, March 1993 [28]

- This paper characterizes concept of data quality from a process perspective
- b) Jang, Ishii, Wang, "A Qualitative approach to Automatic Data Quality Judgement", March 1993 [29]
- This work establishes reduction-based data quality calculus or shortly "data quality calculus"
- c) Wang, Storey, Firth, "Data Quality Research : A Framework, Survey and Analysis", July 1993 [30]
- Using the analogy between product manufacturing and data manufacturing, a framework for Total Quality Management of Data Products is developed that consists of seven elements:
    - Management Responsibilities
      - Data Quality Policy
    - Operation & Assurance costs
    - R & D
    - Production
    - Distribution
    - Personnel Management
    - Legal Function

All through this, data is seen as a "product"

- d) Lee & Wang, " On Validation Approaches in Data Production", August 1993 [31]
- e) Wang, Strong, Guarascio, " An Empirical Investigation of Data Quality Dimensions: A Data consumers perspective", October, 1993 [32]

The " Data Quality Research" paper at ( c ) above by Wang, Storey and Firth is certainly a useful reading. This paper has extensively surveyed the Data quality literature for each of the elements of the Total quality Management framework proposed for the data products. Similarly, in February 1995, International Institute for Management Development ( IMD ), Lausanne, Switzerland has published a guide to provide brief, readable reviews of recent academic, professional and industrial publications as an introduction to the concepts and topics mentioned below [33] :

- Managing knowledge assets in the learning organization,
- Information management and Process Redesign in Enterprise Transformation
- Information Integrity and Use,

- Linking Information Systems and Technology Architecture to Business Transformation, and
- IS Outsourcing and Partnering Strategies

Under the theme of "Information integrity and Use", this guide lists 19 papers published through 1988-1995. Of these, mention may be made of 4 papers; namely, a paper " The Data Pollution Problem" by Bob knight in Computerworld, September 1992 [34], the paper " Devil in your Data" by Linda Wilson in Information Week, 1992 [35] and two papers by Madhavan Nayar, one of which has already been referred to earlier and the other titled "TQM starts with information Integrity" in Technical Support Magazine, February, 1993 [36]. In this paper Nayar argues that Experts typically address the quality concerns of production, sales, service, process improvement and cycle time reduction, but fail to account for the quality information. Nayar further argues that the production of accurate, complete and timely information reduces rerun, paper and overtime costs and augments productivity with the elimination of manual verification, custom coding of controls and redoing 'yesterday's work"

Finally, the most recent research initiative in Information Integrity is the Conference at Toronto in May 1995 by Unitech Systems, Inc., USA which considered the Information Quality and Re-engineering issues in the face of demands of competition due to globalization.

## 9. Analysis

### *Information System Model*

While discussing Data and Information, Rajaraman observes that they are not the same [37]. Specifically, he points out that "Data" is the raw material with which we start and "Information" or Data Product is processed data which is used to trigger certain action or gain better understanding of what the data implies; thereby offering following relationship between Data and Information :

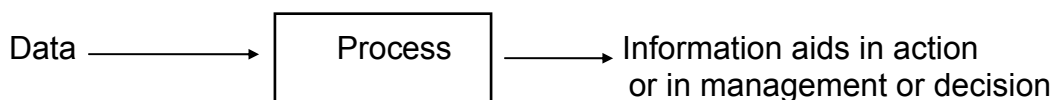


Figure 1 : Data - Information Relationship

In other words, Data-Information Relationship as in Figure 1 presents a systems description with Data as "input", process or procedure constituting the "systems

function”, and Information which aids in actions or in management or in decisions, as the “output”.

### ***Data Errors***

AT&T research publication [24] points out that decisions one makes based on outputs from databases as above, are only as good as data in the database themselves. If the data in the databases one uses contains errors, one risks making blunder after blunder, allocating resources poorly, alienating and losing customers, and doing the same work over again because of the mistakes made. And one can appreciate the order of this problem when one notes the experts estimate that error rates of 10% -50% are common in databases in a wide variety of application. And even these estimates are observed to understate the true extent of the problem because they concern only the accuracy dimension of data quality, they do not address whether the data records are current, complete, or consistent with other data.

### ***Need for Experimentation with Real Data***

It is in the context of this reality of data errors in databases, that Section (3) of the paper has stressed the need for research in “Data Science” which includes experimentation with real data to determine the frequency of various types of errors; in turn leading to development of appropriate data validation techniques.

### ***Defining Integrity***

When considering impact of data errors on Databases, the literature develops the concepts of Data Security and Data Integrity. However Date [38] distinguishes between the two as follows :

- \* Security involves ensuring that users are allowed to do the things they are trying to do.
- \* Integrity involves ensuring that the things they are trying to do are correct.

In other words, Security means protecting the database against *unauthorised* users; Integrity means protecting it against *authorised* users. Indeed, Integrity, unlike Security, is applicable even in a single-user system (it is always desirable to avoid errors); however, it is far more relevant when the system is shared. The reason is that in a shared environment one person’s mistakes can affect other people. Each user is dependent to some extent on the correctness of other people’s actions.

In the above context, literature also reports concepts of “Data Privacy” and of “Data Disaster Preparedness and Recovery” [19]. However, when seen against

the definition by Date, “Integrity” stands separate from these concepts and clearly goes beyond the concept of “Security”.

It is within this framework then the concept of “Integrity” needs to be pursued for research.

### ***Integrity Attributes***

Accordingly Date defines “Integrity” to refer to the Accuracy, Correctness, or Validity of the data in the databases [38]. Thus what one is considering are Data Quality attributes. The literature survey mentions many quality dimensions of data; namely, Accuracy, Usability, Reliability, Independence, Timeliness, Precision, Completeness, Relevance, Sufficiency, Understandability, Freedom from bias, Consistency, Conciseness, Trustworthy, Uptodate, Brief, etc. Indeed Delone & McLean [39] have conducted a survey of quality dimensions of data/information and their review reports twenty three measures of information quality.

Keeping in mind Data-Information relationship in Figure 1, and as it is the Information, which is a processed data, which is what essentially the user is looking for, it becomes natural that one may basically research attributes mentioned above as *integrity* dimensions, going beyond data quality and extending to also include process quality and data product quality, to provide a scientific basis for defining Information Integrity.

As can be seen from the survey of the research presented in this paper, the main thrust of the research effort has been in studying the attribute of “accuracy”. Further, here too one is not clear of acceptable definition as to what one means by accuracy. For example, while describing “Qualities of Information”, Rajaraman defines “accurate” information as one with “correct input and processing rules” [37] and most of the research efforts in literature assume that a record is accurate if it conforms to the actual value. But Wang, Storey and Firth [30] ask : What is the “actual value”? Do values in all fields of a record need to conform to the “actual value” in order to be considered as “accurate”? Should a record with one inaccurate field value be defined as more accurate than a record that has two inaccurate field values?

### ***Objective Integrity attributes basic to an information system***

Without going into further details, it thus becomes clear that, to begin with, there is first a need for research effort to study which integrity attributes are basic to the Data-Information relationship model of Figure 1. That is to say, the attribute of “Accuracy” certainly is basic to an Information System model of Figure 1. Similarly, which other information Integrity attributes are basic and which are secondary in the sense being contextual and hence may vary from application to application.

### ***Subjective Information Integrity attributes specific to areas of use***

Information Integrity attributes of accuracy; trustworthy or reliability; keeping in mind that it is process or procedure that transforms Data into Information - of consistency; and completeness could be first candidates to identify “basic” or “objective” Integrity attributes characterising an information system. Other attributes like timeliness, usability, sufficiency, etc. may be considered as “subjective” to the application for which the Information system is put to use and hence may be treated as add-on to the “objective” attributes of Integrity.

### ***Need for research in Objective and Subjective attributes of Information Integrity***

It is such categorization of Information Integrity attributes into “basic” or “objective” attributes, intrinsic or integral to an information system, and “subjective” attributes, specific to the application areas, that would then need to be followed up by research investigations in terms of defining these Information Integrity attributes with the objective of measuring them. Once again as mentioned earlier, there is some work reported in literature in terms of quantification of accuracy attribute [1], but there is not much work reported in terms of other attributes mentioned above [30].

It is such quantification or measure building for “objective” Information Integrity attributes, that can then lead to developing an objective, holistic measure of Information Integrity for the information system designed; while, based on “subjective” Information Integrity attributes, subjective or application or use specific Information Integrity measure can be considered for development.

### ***Need for research in Integrity Analysis Methodology***

Another area of research that emerges is in terms of “Analysis and Design of Information Systems for Information Integrity”. Integrity Analysis Methodology discussed earlier presents research effort in this area with reference to the Information Integrity attribute of “accuracy”. The literature has not covered the Integrity Analysis Methodology for other Integrity attributes - objective as well as subjective.

The need is thus to initiate research to develop such Information Integrity Analysis Methodologies for Information Integrity attributes other than that of “accuracy”.

In the above context it is submitted that with appropriate modification as per the area of application, Svanks suggestions for the research thrust as given in Section (4) of this paper delineate the research initiatives thus needed.

### ***Need for research in Integrity Technologies***

The literature also reports research efforts in the area of systems technologies as reflected through AT&T work on Data Tracking Technique. As explained in Section (8) of the paper, this technique of AT&T using combination of statistical control and identification of errors and their sources to systematically locate cause of errors so as to improve data quality is a useful Integrity technology (a) in designing an information system with given degree of integrity as also (b) in enabling the information system to produce Information with specified degree of integrity.

However once again, such techniques have been studied in the literature for the Information Integrity attribute of accuracy and it will be very useful to develop such systems technologies for other Information Integrity attributes.

### ***Other research Areas in Integrity Management***

Finally, literature reports a research effort by viewing an information system as a product manufacturing system where raw data is the raw material and data product or information is the output material or physical products of the manufacturing system. This is a very interesting framework to pursue Total Data Quality Management Research for Information, i.e., Data Products, and here the areas of Information Integrity Policy for the Management; Economics of Failures external to information system; Distribution of Data Products or Information; Motivational issues involved in high Information Integrity; Legal issues surrounding data product to enhance Information safety and minimising Information liability; etc. have not received much attention and therefore constitute the possible research directions to investigate the area of Information Integrity.

## **10. Conclusion**

The question of vulnerability of computerised information systems has been attracting the attention of designers and practitioners of the information systems for quite sometime now. Particularly, last decade has been full of anecdotes on data errors and their costs to business due to consequent information failures. However, if decision makers in all walks of life, including business, are to take a serious view of information and its failures, the researchers need to demonstrate for the purpose the criticality of Information Integrity. This calls for scientific efforts to generate body of knowledge in Information Integrity : (a) to develop area of Data Science and to define concepts of Information Integrity and to measure Integrity attributes, (b) to develop Integrity Analysis Methodology for Integrity attributes and a “standard” Information Integrity metric or acceptance test that could be used at the information system design stage, (c) to establish criteria for selection of qualified data/information supplier, (d) to develop technologies to manage Information Integrity, (e) to study human or automated information processing procedures, (f) to investigate the link between problems

of poor Information Integrity and revision of the procedures used to detect and eliminate problems, etc. Towards this objective the paper has surveyed the work done in Information Integrity in the literature and presented research investigation initiatives for further pursuit.

## 11. References

1. Svanks, Maija I., **Integrity Analysis : A Methodology for EDP Audit and Data Quality Assurance**, EDP Auditors Foundation, Inc. 1984
2. Wooldridge, Susan, **Computer Input Design**, Petrocelli Books, New York, 1974
3. Feltham, G, **The Value of Information**, The Accounting Review, 43(4), 684-696, 1968
4. Yu, S. and Neter, J., **A Stochastic Model of the Internal Control System**, Journal of Accounting Research, 1(3), 273-295, 1973
5. Burns, D., and Loebbecke, J., **Internal Control Evaluation : How the Computer Can Help**, Journal of Accountancy, 140(2), 60-70, 1975
6. Bodnar, G., **Reliability Modelling of Internal Control Systems**, The Accounting Review, 50 (4), 747-757, 1975
7. Fields, K.T., et al., **Quantification of the Auditor's Evaluation of Internal Control in Database Systems**, The Journal of Information System, 1(1), 24-27, 1986
8. Amer, T., et al., **A Review of the Computer Information Systems Research Related to Accounting and Auditing**, The Journal of Information Systems, 2(1), 3-28, 1987
9. Groomer, S.M. & Murthy, U.S., **Continuous Auditing of Database Applications : An Embedded Audit Module Approach**, The Journal of Information System, 3(2), 53-69, 1989
10. Halloran, D., et al., **Systems Development Quality Control**, MIS Quarterly, 2(4), 1-12, 1978
11. Bailey, J.E., and Pearson, S.W., **Development of a Tool for Measuring and Analyzing Computer User Satisfaction**, Management Science, 29(5), 530-545, 1983

12. King, W., and Epstein, B.J., **Assessing Information System Value : An Experiment Study**, Decision Sciences, 14(1), 34-35, 1983
13. Agmon, N., and Ahituv, N., **Assessing Data Reliability in an Information System**, Journal of Management Information Systems, 4(2), 34-44, 1987
14. Morey, R.C., **Estimating and Improving the Quality of Information in the MIS**, Communications of the ACM, 25(5), 337-342, 1982
15. Ballou, D.P., and Pazer, H.L., **Modelling Data and Process Quality in Multi-input, Multi-output Information Systems**, Management Science, 31(2), 150-162, 1985
16. Ballou, D.P., and Pazer, H.L., **Process improvement versus enhanced inspection in optimized systems**, International Journal of Production Research, 23(6), 1233-1245, 1985
17. Ballou, D.P., and Tayi, K.G., **Methodology for Allocating Resources for Data Quality Enhancement**, Communications of the ACM, 32(3), 320-329, 1989
18. Pautke, R.W., and Redman, T.C., **Technique to control and improve quality of data in large databases**, In the Proceedings of statistics Canada Symposium 90 (pp. 319-333), Canada, 1990
19. Baker, H.B., **Information Integrity-Structure for its Definition and Management**, McGraw Hill Book Company, NY - 1983
20. Lane, V.P., **Security of Computer Based Information Systems**, Macmillan Education Ltd., U.K. 1985
21. Norckaur, H.R. Jr., **Duplicate Entry Detection in Mailing and Participation Lists**, Master of Science Thesis in Management, MIT, Cambridge, USA, May 1990
22. Hansen, M.D., and Wang, Y.R., **Managing Data Quality : A Critical Issue for the Decade to Come**, Working Paper, Alfred P. Sloan School of Management, MIT, USA, March, 1991
23. **Data Quality Foundation**, Published by AT&T Quality Steering Committee, USA, 1992

24. **Improving Data Accuracy : The Data Tracking Technique**, Published by AT&T Quality Steering Committee, USA
25. Redman, T.C., **Data Quality : Management and Technology**, Bantam Books, NY, 1992
26. Wang, R.Y., and Reddy M.P., **Quality Data Objects**, TDQM Research Program, Sloan School of Management, MIT, Cambridge, USA, December 1992
27. Nayar, M.K., **Achieving Information Integrity**, Information Systems Management, Spring 1993
28. Kon, H., Lee, J., and Wang, R., **A Process View of Data Quality**, TDQM - 93-01, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA, March 1993
29. Jang, Y., Ishii, A.T., and Wang R.Y., **A Qualitative Approach to Automatic Data Quality Judgment**, TDQM - 93-02, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA, March 1993
30. Wang, R., Storey, V., and Firth, C., **Data Quality Research : A Framework, Survey and Analysis**, TDQM - 93-11, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA, July 1993
31. Lee, J., and Wang, R., **On Validation Approaches in Data Production**, TDQM - 93-15, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA, August 1993
32. Wang R.Y., Strong, D.M., and Guarascio, L.M., **An Empirical Investigation of Data Quality Dimensions : A Data Consumer's Perspective**, TDQM -93-12, The TDQM Research Program, MIT, Sloan School of Management, Cambridge, USA, October 1993
33. Bechler, K.A., and Marchand, D.A., **Improving Business Performance Through I/T Leadership - A Guide to the Literature for Executives**, International Forum on Managing Information and Technology, The International Institute for Management Development (IMD), Lausanne, Switzerland, February 14-15, 1995
34. Knight, B., **The Data Pollution Problem**, Vol.26, No.39, Computerworld, September 28, 1992, p: 81-83

35. Wilson, L., **Devil in Your Data**, Information Week, August 31, 1992
36. Nayar, M.K., **TQM Starts with Information Integrity**, Technical Support Magazine, February 1993
37. Rajaraman, V., **Analysis and Design of Information Systems**, PHI, New Delhi, 1991
38. Date, C.J., **An Introduction to Database Systems**, Vol. 1, 4th Edn., Reading, Massachusetts, Addison - Wesley, 1986
39. Delone, W.H. & McLean, E.R., **Information Systems Success : The Quest for the Dependent Variable**, Information Systems Research 3(1), 1992

0-0-0-0-0

**Below is given summary of discussions that followed the paper presentation :**

- \* Problem of Data Error is a real one.
  - Today the problem is that systems are generating information at a phenomenal rate. We have far more data to enter.
  - There is also a problem of non-numerical data.
  - Concept of data warehouse is already recognised.
- \* Can there be a course in the universities in data repair, data science and/or data technology?
  - One has to really look for the available literature.

---

**Legend :** \* **Question or Comment**  
- **Answer or Response**