Modeling Information Flow for Integrity Analysis

Vijay V. Mandke
Research Leader
Centre for Information Integrity Research
Unitech Systems (India) Pvt. Ltd.,
B-64 (1st Floor), Gulmohar Park,
New Delhi – 110049 India

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

Abstract

This paper addresses the research issue of developing Information Flow Model for implementing error detection and correcting mechanisms in IS in the form of feedback control based, on-line integrity technologies. Towards this, the paper begins by categorizing error components in IS in terms of errors with deterministic descriptions caused by singular events and errors with stochastic descriptions caused by general, judgmental, and systems factors. In the process, the paper studies the impact of man-IS interface and of changes in business and technological environments on integrity of information. Following this, the paper suggests a basic IS model to account for these error components, particularly those caused by judgmental and systems factors as they account for material impact of business changes and people on information integrity. Based on this study, the paper suggests that to undertake meaningful integrity analysis and to develop integrity technology, the Information Flow Model should take a holistic view of IS by viewing Data Origin Stage, Conversion (Application) Stage and Output (Information Use) Stage together. Accordingly, the paper identifies data processing activities under each of these stages and develops information flow models for each of the activities alongwith their respective integrity analysis implications.

1. Introduction

Whether, in addition to application controls, hardware and software vendors further incorporate error-checking filters into their products, networked computerized information systems contain errors that are made but not corrected. These errors are due to factors drawn from system environment, external to application system and overlapping the user environment; resulting in loss of integrity in IS. Mandke and Nayar [5, 6] have identified these integrity implications in terms of intrinsic integrity attributes of **accuracy** (includes completeness and timeliness), **consistency** (satisfying domains and constraints) and **reliability** (accuracy with which information item represents data item in whatever way information system processed it). Further given that a triple <entity, attribute, value> as developed by the database research community represents data/information model and given a simplistic situation wherein value part of data/information is expressed in a numerical, Mandke and Nayar [7] present approaches for quantifying accuracy, consistency and reliability and, thereby, to build a cumulative information integrity index (CIII) providing a measurable basis for demonstrating integrity level in the IS. This then offers a basis for identifying Information Integrity Technology Development Steps [7]. However, implementation of this Integrity Technology involves constructing Information Flow Model (IFM) which is amenable for integrity analysis. In what follows this paper addresses this model building query.

2. Categorizing Error Components in IS

While discussing errors in IS that are made but not corrected, Mandke and Nayar [7] propose that (a) data/information model be represented by triple $< e, a, v + \eta >$ where η represents noise or error component and that (b) these error implications are present at each stage of an information system; namely, data origin stage, communication channel prior to processing stage, processing stage, communication channel at post-processing stage and output stage.

It is these errors in IS that provide opportunities for information pollution as information is rendered incomplete and inconsistent, and at incorrect level of detail, means different things to different users, and is based on data which is owned by multiple user areas, not current and timely, changed by multiple user areas, transformed and altered without data administrator's knowledge, re-entered into computers from incorrect media and does not represent requirements of changing business context; resulting in loss of integrity in information [4, 9].

When abstracted, these error components can be modeled to include:

i) errors with deterministic descriptions caused due to events singular in nature like software failure, denoted by η_{sing}, and

ii) errors with stochastic descriptions caused due to :

    1) general causes like mechanistic failure, service disruptions, etc., denoted by η_g,
    2) human judgmental factors operating at man-IS interface, denoted by η_j, and
    3) systems factors (external and internal to IS) like a merger, regulatory activity,legislative action, activity of a competitor, acquisition of new software or hardware, etc., denoted by η_s.

2.1 Errors due to Singular Events Deterministic in Nature

Categorization of errors by types of sources that cause them as above, is important as it affords further clearer understanding of respective error components effected. For example, in respect of error(s) caused due to a singular event, there is still the problem of locating and correcting the other instances of those errors. Having identified, through integrity analysis, presence of error(s) of this category, the responsibility of Information Integrity Technology is to correct all those errors and to locate the singular event like say the computer program which is going wrong causing those errors and then remove the cause present in the form of the singular event, i.e., correct the software, so as to improve the integrity of information obtained from the IS [4].

2.2 Stochastic Errors

2.2.1 Errors due to General Factors

Against this errors with stochastic descriptions have general, judgmental or systems causes which are embedded in processes or stages and sub-processes or inter-mediate stages which go to make the Information Flow Model (IFM) for an information system. To elaborate, errors of general type caused by mechanical and/or service disruptions are probably the one most universally accepted by researchers, designers, operators and users in any kind of system and are extensively studied [13, 10, 2, 4].

2.2.2 Errors due to Judgmental Factors

For the purpose at hand in this paper, it is therefore more necessary to understand the errors caused by factors coming under judgmental and systems types as defined here.

Specifically, as mentioned above, judgmental factors are particular to man-IS interface environments which could be at component level or sub-process or process level or at total system level. To explain, bulk of IS applications in business so far have been in respect of transaction processing systems (TPS) requiring batch processing, having straight forward and elementary procedures and based on centralized computer systems capable of handling simple documents but in huge quantities [8, 2]. Success of these application – oriented data processing systems aided by requirement to perform in relatively static technological and business environments brought in perception of automated computerized information systems, not withstanding incidents of integrity loss due to incorrect data generation at data origin stage itself or due to data entry errors at operator – machine interface or due to processing the same data more than once and in situations due to disruption of services.

However things have changed since 1980s. Real – time systems are now as common as batch systems. Communications are now an integral part of most systems. A wide range of hardware is in the market and software is available as packages. No doubt vendors and software houses attempt to deliver the best possible software. However they are torn between stability and the necessity to enhance their products. In their continuing efforts to improve the functionality of their products and remain competitive, they have to make changes many of which not anticipated. Regardless of the amount of effort and expenses to assure quality of their product, it is inevitable that errors are found when these products are used in the field, thereby generating uncertainty in the user – IS interface environment and requiring user to undertake house – keeping initiatives in the context [5, 14].

2.2.3 Errors due to Systems Factors

Further, with large number of applications thus available and with each application having its own sources of information, terminology, and classification methods, apart from the issues arising from conflicts between the results from two applications, there is the issue of more information available than what user could digest or evaluate [8]. Even then in shared environments of today, user aspires for continual updating of hardware and software products hoping improved functional efficiencies for speedy on-line access to databases and for benefits of working with data in ever increasing quantities and levels of details [9, 4].

All this has further added to efforts to extend use of computerized information systems from TPS to more complex and judgmental business applications covering office automation systems (OAS), management information systems (MIS) and decision support systems (DSS). And the challenge of these developments is that while on the one hand IS is in itself becoming more complex and susceptible to operate under uncertainty, on the other hand, due to systems factors like government actions, competitor's actions, long term prospects of technology innovation, new markets, resource availability, etc., the business environment, in which IS operates, is also becoming at all times prone to perceptible changes in business functions, processes, activities and procedures, thereby subjecting information requirements emanating therefrom to uncertainty.

In other words, it is the sources of judgmental and systems factors that now require IS to deal with data/information characterized by uncertainty. It may be mentioned that information due to systems factors may constitute controllable variables like particular production procedure information which is internal to the business process or uncontrollable variables like governmental regulation which is external to the business environment. Against this, the information requirements due to judgmental factors come into play when IS has to deal with uncertain, imprecise, incomplete, open-ended information environment. Specifically, any such information environment calls for "what if" type problem solving to process the incomplete information (what if unit price is increased ? what if procurement is delayed by another 10 days ?), then to evaluate the alternatives and finally to select the best [8]; thereby making this information processing procedure judgmental in character. As can be appreciated, there is always a chance of error or variance from standard in this process.

3. Developing Basic IS Model

3.1 An IS View of a Business Process

A business system is an orderly, harmonious group of interacting, interrelated, and interdependent procedural components. All business procedures are data processing procedures, that is, they all process data in some manner to deliver information for use in decision making [9]. In such interconnected system, data and information variables are interchangeable in that what is data for a given procedure is information from the previous connecting procedure and the information from the procedure can be data for the subsequent connecting procedure.

3.2 Significance of Judgmental Data processing in IS

An information system so modeled has the requirement to process uncertain data/information due to environmental factors – mainly external. This requirement introduces "judgmental" processing requirements in the IS. Judgmental processing requirements are also introduced when IS is designed to undertake higher level information processing functions like OAS or MIS or DSS in that order. Judgmental processing requirement is also present in transaction processing systems whenever there is operator – IS interface (through machine) which is invariably at each data entry or data conversion or machine operation point where the operator has to think, decide and act. Even in the most automated IS there exist judgmental processing requirements at total system level like say at the level of strategic information processing and through user - IS interface environments which, in shared, distributed environments of today, are exponentially on rise.

It is the judgmental data processing requirement – which uses individual and group forecasting techniques and is extensively used in a wide range of business application areas of production, material requirement, personnel, sales, inventory, advertising, budgeting, new products, pricing, competitive analysis, strategy etc. – that is then fundamental to every situation where uncertain, imprecise, incomplete, inaccurate or inconsistent data or information is to be processed. As mentioned earlier, this data processing is essentially cognitive processing and for its inclusion in IS calls for man – IS (machine) interface though at different levels depending on the nature of IS application area and degree of automation, networking and decentralization in IS architecture. Specifically man-machine interface environment within an IS draws on complementary capabilities that both sides of the interface present – the machine such capabilities as precision, power, endurance, and the man such capabilities as judgment, recognition, creativity – so as to in the end analysis deliver the most efficient IS.

It is within above framework of designing IS for the reality of changing business and technology environments and for use in terms of higher level functions that one then recognizes proper use and design of man – IS interface environments as integral to IS, so as to introduce appropriate cognitive, i.e., judgmental, processing stages in IS for obtaining maximum IS efficiency.

3.3 More about Errors in IS due to Judgmental Factors

And even as one accepts man – IS interface as a reality in IS implementation, one has to grapple with the problem of errors in IS that are introduced due to judgmental factors thus entering IS (η_j) and systems factors present in IS environment (η_s).

Errors at man – IS interface due to judgmental factors and errors due to systems factors have their origin in a number of limitations of human mind such as : too much information, too little time, stress and fatigue, pressure of other demands, tendency of brain to filter information in line with predetermined patterns and beliefs, discomfiture and threat felt when information disagrees with current beliefs, lack of information literacy resulting in an inability to understand what might constitute relevant information, etc. Specifically, the load of information, which is common in continually changing business and technological environments of today, seems to cause stress and fatigue that affect information processing performance in the human mind. What this results in is barriers to information processing that individuals create to cope with an over abundance of information [1]; these include :

a) Failure to process some information because of sheer volume;
b) Faulty processing of information – which may take the form of overlooking what is most critical among relevant data;
c) Adding incoming data/information to a `to be read' pile which results in a backlog;
d) Filtering or systematic omission of certain categories of information;
e) Approximation or over simplification – cutting categories of data/information because there is no time in which to deal with the data or information;
f) Escape or withdrawal from the information situation altogether;
g) Redefining the situation, thus changing the quantity of or relevance of the information required.

All of the above have serious errors and, therefore, integrity implications in respect of information processed by information systems.

3.4 Basic IS Model

It is within the above framework of error components, also referred to as noise components (η) – singular (η_{sing}), general (η_g), judgmental (η_j) and systems (η_s) – present in IS, that the Basic IS Model as in Figure (1) can be defined .

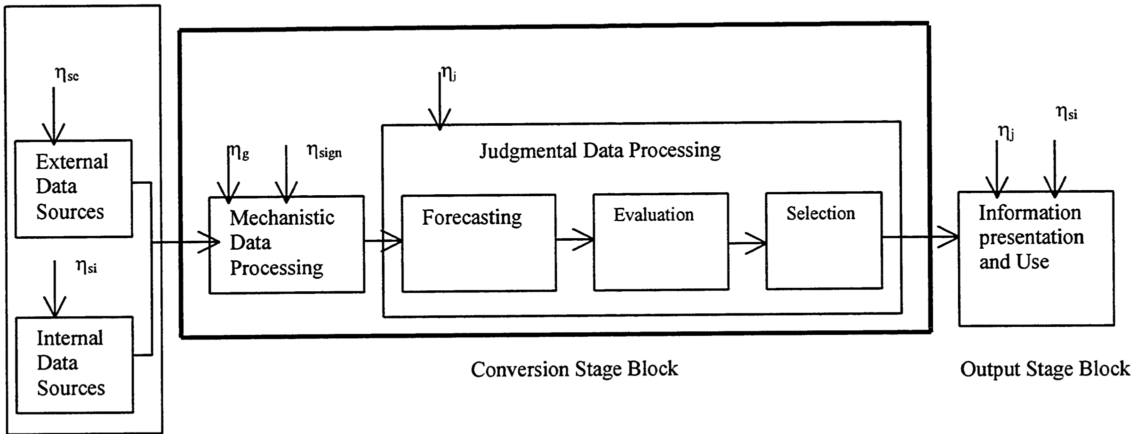

Figure 1 : Basic IS Model

In this model, "Data Origin Block" comprises data generation/preparation/collection process from external and internal sources and is affected by η_s (comprising η_{se}, systems errors due to external factors, and η_{si}, systems errors due to internal factors) indicating uncertainty generated by systems factors of government regulations, competitor's actions, new products, technology innovation, etc. Data originating from this Data Origin Stage may be classified under three broad categories: (a) certain and controllable data emanating from sources, mainly internal, operating in closed environments, (b) uncertain but controllable data emanating from internal sources affected by changing internal environment like software and hardware upgradation, and (c) uncertain and uncontrollable data mainly originating from external environment. Data thus collected is processed through "Conversion Stage Block" which represents mechanistic data processing and judgmental data processing.

Specifically mechanistic data processing stage is in a position to completely process data characterized by certainty (category a), the judgmental processing (necessitated by human intervention as through operator – machine interaction) playing minimal role. Information obtained from mechanistic data processing is subject to errors due to general factors (η_g) and singular factors (η_{sing}).

It is in respect of processing requirements for data characterized by uncertainty, i.e., categories b and c (and which is the most significant case (whether source is internal or external) in continually changing business and technological environments as IS is also designed to serve higher level of functionalities), that along with mechanistic data processing there is extensive use of judgmental data processing requirements which in turn constitute the basis for man – IS interface environments in IS. Judgmental data processing procedures comprise forecasting (F), evaluation (E) and selection (S) activities. Specifically, judgmental forecasting techniques include individual forecasts which are intuitive, adhock; group forecasts which could be committee based or scientific group interaction based; and aggregates' forecasts where forecasts are based on information aggregated from many individuals as in case of market research. As can be appreciated, a number of alternatives may emerge as a result of forecasts based on uncertain data or information and it is these alternatives that are evaluated to make a selection and complete the judgmental data processing. The selected alternative is then presented under the "Output Stage Block" in an identified format to constitute output of IS, i.e., the information which is used by the user say in an information function or in the command and instruction function or in the influence and persuasion function or in the integrative function or as an aid to make a control decision for improved performance, etc. as the case may be, once again signifying man – IS interface character of an IS. Finally it may be mentioned that this IS representation in terms of basic IS blocks is a core representation. In actual system these basic blocks are combined in a variety of ways. They may be repeated, paralleled, and interrelated. If the integrity implications of error components η_{sing}, η_g, η_j and η_s are not controlled, that is how then the IS would become infested by the problem of information pollution.

It is this Basic IS Model represented in its elementary block form that then provides the basis for developing Information Flow Model (IFM) compatible with requirements of integrity analysis.

4. Information Flow Model

In the Basic IS Model in Figure (1), Data Origin block comprises data generation/preparation/collection activities from external and internal data sources. In terms of the task of modeling datum as a triple <entity, attribute, value> denoted by < e, a v >, the Data Origin stage begins with identifying a view of the real world, where the view is denoted by{ < e, a, v>} [5]. As an example, the real world may represent say a simple business process of 'vendor supplying items to a company'; the IS being implemented for this business process. One may identify an universe of entity types or classes and entity relationships for this business process in the form of "vendors" and "items" representing entity types or classes and "vendors-items" representing relationship. Then attributes for entity type "vendors" may be identified as vendor code, vendor name and address; for entity type "items" as item code and item name; and for entity relationship "vendor-item" as vendor code, item code, order no., quantity supplied, date of supply and price per unit [11]. Further, based on the knowledge of the business environment, domains for each attribute may be identified based on which values for specified attribute of a specified entity from a given entity type or class or entity relationship are obtained.

4.1 Data Origin Stage

4.1.1 Modeling View Defining Activity

In the business environment, examining only transactions, or processes, or outputs, or data flows, or even a combination of all four produces a picture, which is correct as far as it goes, but which does not reflect a true or complete picture of the environment [9]. The power of data/information model as a triple <e, a, v> using E-R approach described above without reference to a specific DBMS model, lies in its requirement to focus on describing the entities of the real world of the business, and the relationships between them. However, this in turn requires identification and definition of universe of entities (i) where an entity is defined as a person, place, or thing which is (a) of interest to the business, or business firm if so is the case, (b) distinguishable from all other types of entities and (c) relevant within the context of the specific environment of the firm and (ii) where entity relationship is any association, linkage, or connection between the entities of interest to the firm, once again such that relationships describing business context are also (a) of interest to the firm and (b) relevant within the context of the specific environment of the firm.

Against this, attributes provide entity description by the way of identifying and defining aspect, quality, characteristic, or descriptor of either an entity or a relationship. As a result, an attribute must also be (a) of interest to the firm or business and (b) relevant within the context of the specific environment of the firm.

Within above framework then it is easy to appreciate why good understanding and knowledge of business environment and of systems concepts in development of E-R model is a must to correctly define view {<e, a, v>} and, for that purpose, correctly identify and define {e}. Indeed the view identification and definition process undertaken by the individual generating, collecting or preparing data at Data Origin Stage must specify the entity at the exact level of precision which ensures that it is not so general as to be meaningless and yet not so specific that it fragments into too many subsets. Finally, entities identified and defined must be so generic in their presentation of the real world, i.e., say business structure that to the extend possible they should be of interest to the firm or business and relevant to the specific context of the business even when business procedures and for that matter business changes; changes in entities identified becoming necessary only when business changes or is willed to be changed dramatically [9]. In other words in changing environments, entities identified should have much larger time constants than those of business and those of business procedures in that order.

Defining the view of the real world is thus a "problem solving" situation involving data originator(s) and calls for judgmental data processing characterized by man – IS interface and depends on the participant's understanding, experience, orientation, knowledge and perception of the real word. In a changing environment of the real world, the activity essentially takes the form of open-ended problem solving where uncertain data (due to presence of η_{se} and η_{si} ,i.e., error components due to systems factors from external and internal environments, respectively) is processed through stages of forecasting of alternatives, their evaluation and selection to define the view when no one definition can be uniquely correct. In the era of information overload this cognitive data processing implemented through man – IS interface is further characterized by barriers to data processing through human mind detailed elsewhere. And the very basis of E-R approach to data modeling requires that the view {<e, a, v>} so emerging under the Data Origin Stage must ensure Integrity [12].

Within above framework, Figure (2) gives a system's representation of Information Flow Model for View Defining Activity under Data Origin Stage of Basic IS Model alongwith its integrity analysis implications.

4.1.2 Modeling Data Representation Stage Activity

After defining view, the Data Origin Stage has to implement Data Representation step which comprises defining a set of rules for recording data on a prescribed medium and in the process identifying and defining format (f) for representing value part of data and a set of symbols (S) for recording value. This is an involved area of investigation as requirement is to

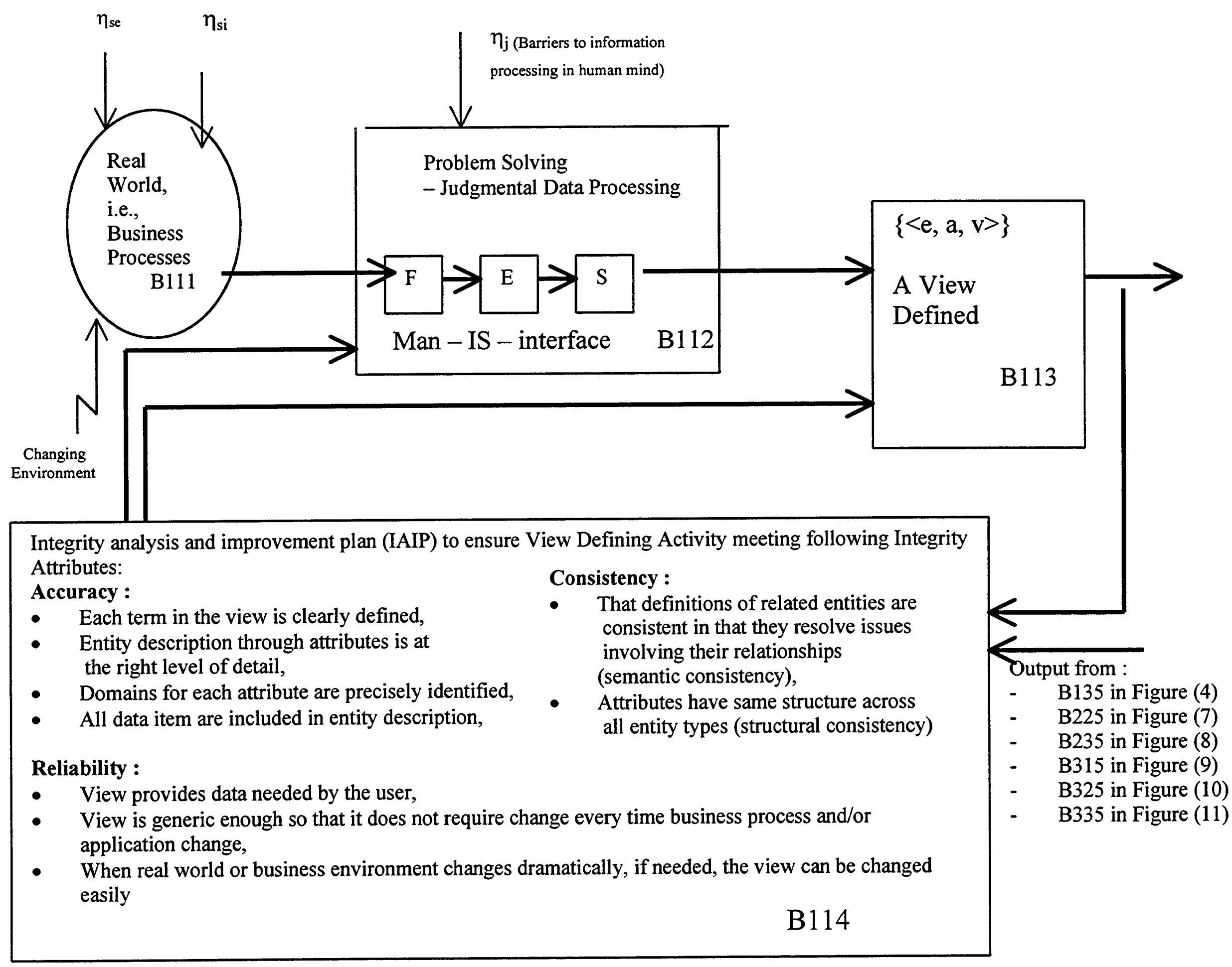

Figure 2 : Information Flow Model for View Defining Activity under Data Origin Stage of Basic IS Model alongwith accompanying integrity analysis implications

define formats (f) for data types like labels, categories, quantities representing entities and attributes and to transform their respective conceptual domains into appropriate sets of symbols which are well understood and are communicable to all those involved in generating/preparing/collecting data.

Needless to say there can be more than one representation of the same data item like date represented by European format dd/mm/yy or American format mm/dd/yy and one will have to evaluate these alternatives and select one; thereby suggesting that the Data Representation is also a judgmental data processing activity involving forecasting of alternatives, evaluation and selection and indicating presence of man – IS interface environment which, as elaborately explained through this paper, is prone to errors. And in spite of this the E-R Approach requires that Data Representation defined ensures integrity.

Within above framework, Figure (3) gives a system's representation of Information Flow Model for Data Representation Activity under Data Origin Stage of Basic IS Model alongwith its integrity analysis implications.

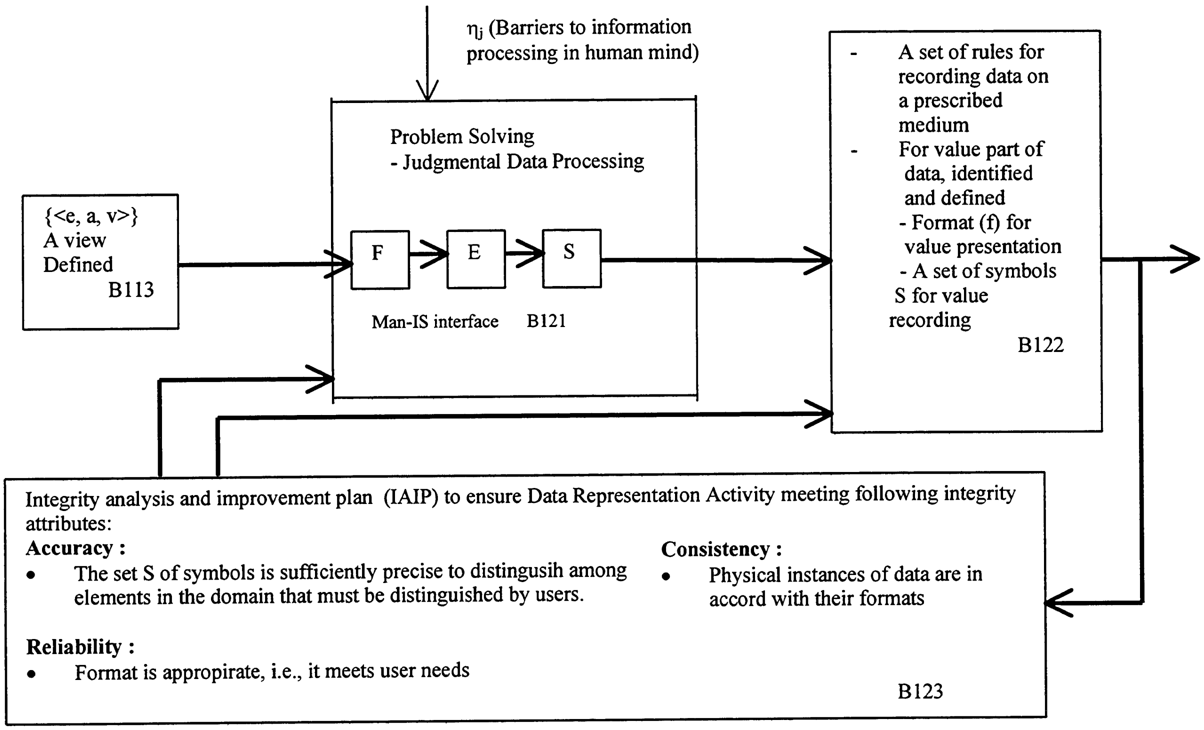

Figure 3 : Information Flow Model for Data Representation Activity under Data Origin Stage of Basic IS Model alongwith accompanying integrity analysis implications

4.1.3 Modeling Data Value Obtaining Activity

After deciding on a set of rules for recording data on a prescribed medium and in the process on a format (f) for representing value part of data and on a set of symbols (S) for recording value, Data Origin Stage implements "value obtaining" activity. Though conceptually straightforward, like "view defining activity", the importance of this activity has been underestimated by both practitioners and theoreticians. At the one end, use of diverse yet common manual forms of data acquisition like measuring, surveying, observing, and copying from another source (which are characterized by sources of errors such as measurement errors, sampling errors, hidden information, poorly designed forms and questionnaires, data aggregation, classification and definition, time factors, etc.,), and, at the other end, use of electronic data acquisition systems (once again with manual data entry alongwith analog and digital inputs and comprising building blocks of analog multiplexer, signal conditioning, A/D converter, output buffer and mini-computer and its peripherals, which are further characterized by systems errors), provide an appreciation of potential difficulties in obtaining values. Yet another source of error in obtaining data is that the characteristics of a small population may change, with the result that different data is reported or obtained than would be the case without these changes.

An observation about the data-obtaining task is in order here. As mentioned earlier, facts in the real world are modeled by a view {<e, a, v>}, where each entity type or relationship `e' is described by attributes (a_1, a_2, . . . , a_n); e.g., entity type "vendors" being described by attributes (vendor code, vendor name, address). Here each attribute (field) represents data type which can be classified as "label". To elaborate, for every specific vendor from the entity type "vendors", there will be a specific code number (which will be a number but does not represent an arithmetic quantity but a label), a specific name and a specific address (both of which are labels), all of which will go to form a record. Now what is submitted is each field, which could be a label or a category or a quantity as the case may be, has characteristics and whenever data record is generated or prepared or collected or recorded, it is important that these characteristics are correctly processed, or otherwise the value obtained will be incorrect, inconsistent and unreliable. These characteristics are :

a) type of data (numeric, alphanumeric, or binary),
b) length of data (in characters or bytes),
c) purpose and use of data,
d) method used to gather the data,
e) operations performed on the data (for example, coding, formatting, combining with other data).

And as can be seen, all through these Data obtaining activities, which consist of data observation, interpretation and recording, there are man – IS interface environments; thereby once again making judgmental data processing integral to Data Obtaining stage, and thereby once again exposing the Data Origin Stage to errors that accompany man- IS interface environment implementation. Here it may be mentioned that this interface is present say every time when an individual is observing, measuring, surveying or copying data for recording value part of the data or when an electronic data acquisition system ,with its manual data entry operations, is entrusted with the value obtaining function. It is in the face of these occasions for errors in processing of data items at the `value obtaining' stage that E-R information flow model should ensure integrity of obtained value of data item.

Within above framework, Figure (4) gives system's presentation of Information Flow Model for Value Obtaining Activity under the Data Origin Stage alongwith the accompanying integrity analysis implications.

4.1.4 Modeling Data Storage Activity

The next and final activity under Data Origin Stage is storing of data (values) so obtained. Indeed it is in Data storage that phenomenal progress has taken place over last few decades. What with unimaginable volume of data stored in many databases – and the likelihood that this volume will grow by several orders of magnitude when technologies such as optical storage are widely used – and what with modern DBMS designs implemented allowing users working in distributed and shared environments to access and alter the same data almost simultaneously, there are problems to maintain Accuracy, Consistency and Reliability of data/information stored and accessed; in turn leading to loss of integrity in databases [5, 12].

Figure (5) gives systems representation of Information Flow Model for Data Storage Activity under the Data Origin Stage alongwith the accompanying integrity analysis implications.

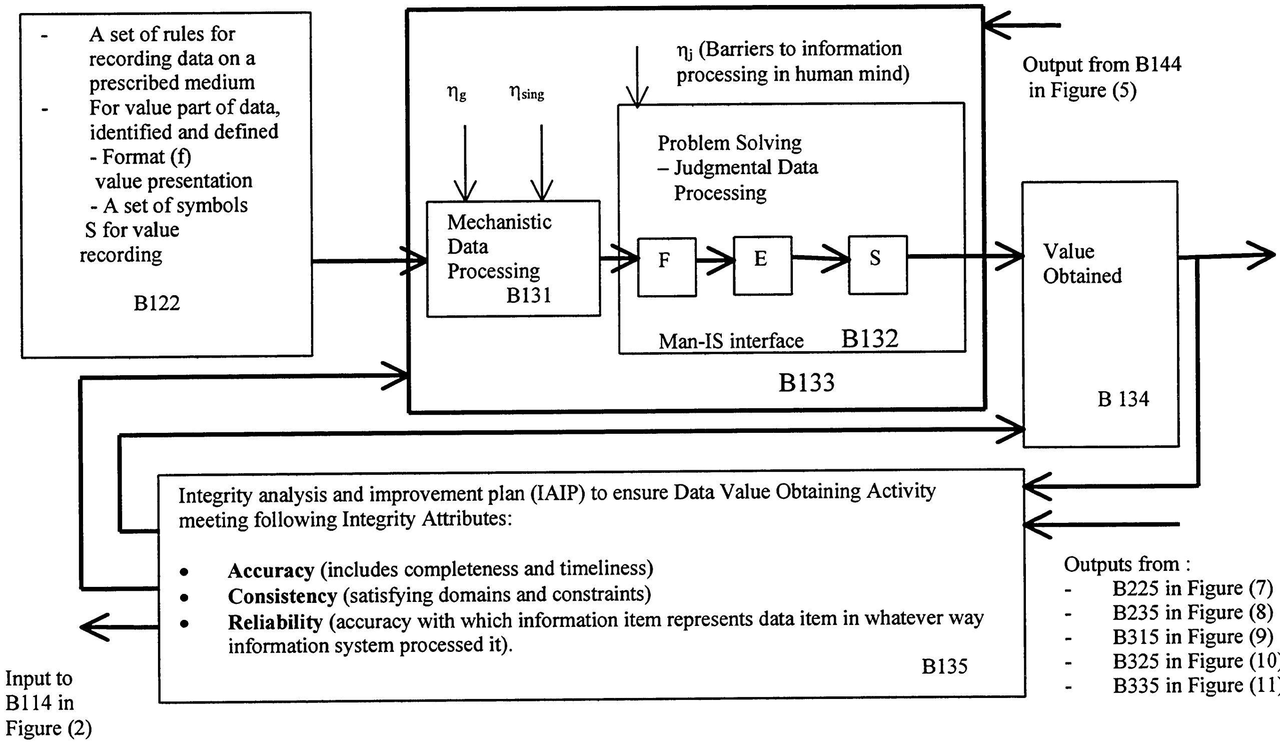

Figure 4 : Information Flow Model for Data Value Obtaining Activity under Data Origin Stage alongwith accompanying integrity analysis implications

4.2 Conversion Stage

Upto this stage investigation has considered the Data Origin Stage in its totality, and now one focuses attention on the "Conversion Block" in the Basic IS Model of Figure (1). Specifically, the "Conversion Stage" begins with user, consistent with her or his application need, defining a subset from the set of data as collected and stored under the Data Origin Stage. Literature defines this as sub-view [3] and it may be denoted by $\{<e_a, a_a, v_a>\}$, where "a" denotes application area of user interest.

4.2.1 Modeling Sub-view Defining Activity

It is important to realize that while view defined in the Data origin Stage constituted model of the real-world, i.e., business environment, the sub-view is based on the data collected based on that view, and is dependent on the user's knowledge, understanding, orientation and perception of the data so collected and stored. Thus the process of defining sub-view $\{<e_a, a_a, v_a>\}$ is implemented within the confines of the cognitive boundary governing user's thinking. Otherwise structurally this "sub-view defining" stage has similar model as that for view defining stage, i.e., it is characterized by judgmental data processing and has same integrity attributes though of course in the context of user's application area. Even in the face of possibility of inaccurate, inconsistent and unreliable data/information stored and even in the face of errors subsequently introduced due to judgmental data processing through user specific sub-view defining stage, the sub-view defined should meet requirements of concerned Integrity Attributes.

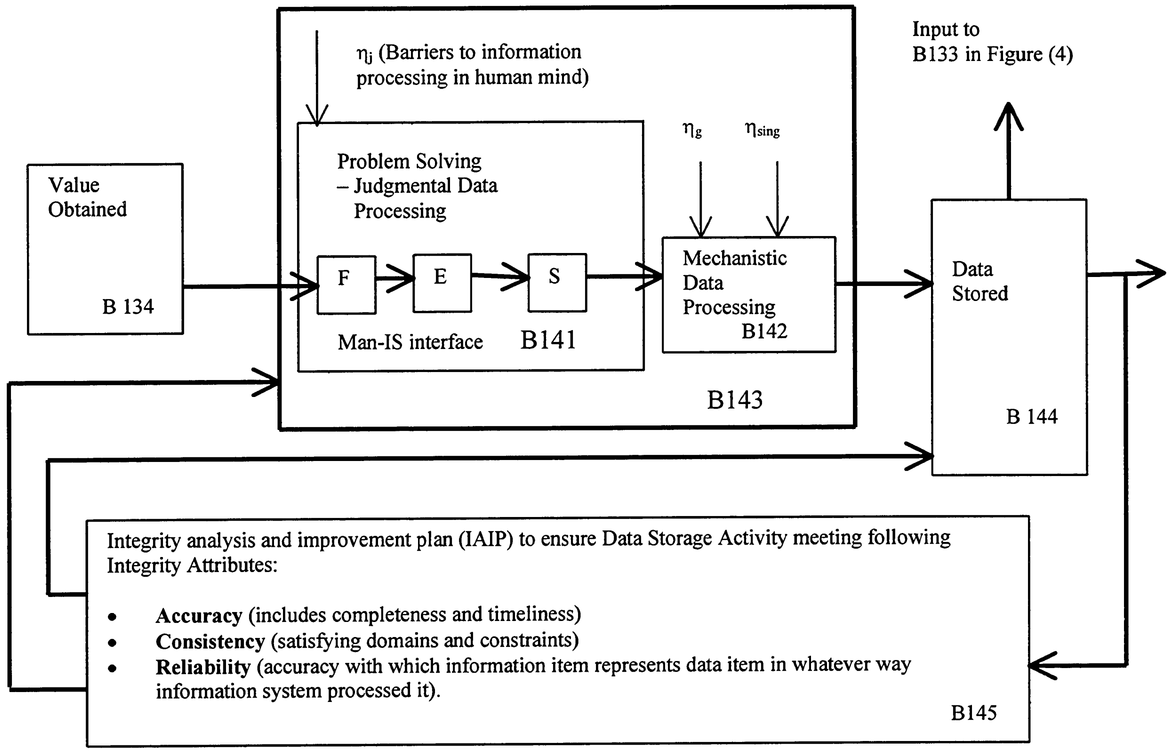

Figure 5 : Information Flow Model for Data Storage Activity under Data Origin Stage alongwith accompanying integrity analysis implications

Accordingly, Figure (6) presents systems representation of Information Flow Model for Sub-view Defining Activity under the Conversion Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications.

4.2.2 Modeling Data Retrieval Activity

Once sub-view is defined, the Conversion stage would need to retrieve the data identified under the sub-view. One situation is that the retrieval requirements may be met with the help of same DBMS as for Data Storage Activity. Yet another possibility is that completely different people and organization than those dealing with Data Storage Stage may handle the retrieval requirements. And still another possibility is that, to facilitate data retrieval, user is given on-line access, as the user unit or an individual user is likely to have in many cases. All these situations in increasing order point towards programme and data errors due to factors of singular type rendering data retrieved inaccurate, inconsistent and unreliable. Further particularly activities of data entry also introduce operator – machine interface environments at Data Retrieval Stage, in turn resulting in judgmental type data processing involving forecasting, evaluation and selection stages which also cause errors that result in loss of integrity at Conversion Stage in terms of attributes of accuracy, consistency and reliability.

Within this framework, Figure (7) presents systems representation of Information Flow Model for Data Retrieval Activity under the Conversion Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications.

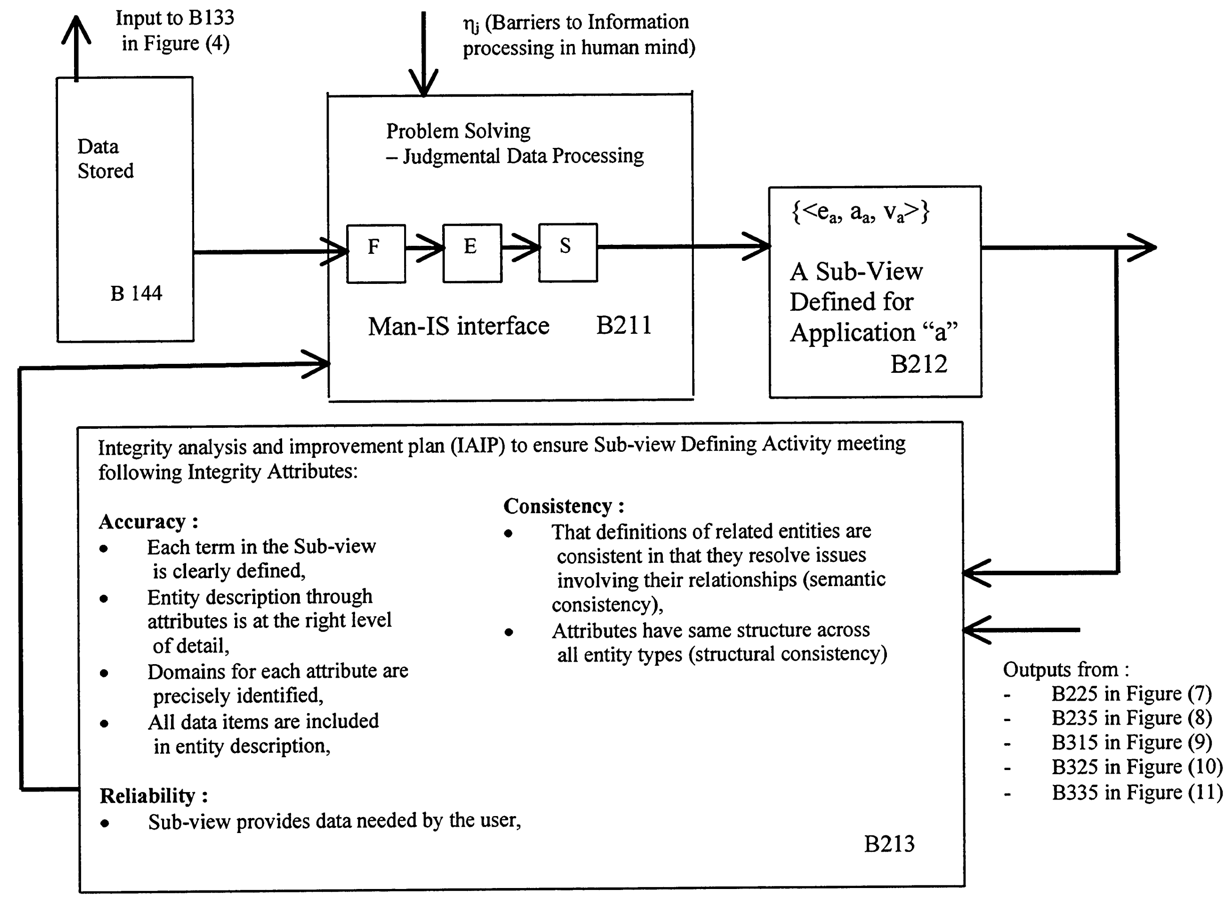

Figure 6 : Information Flow Model for Sub-view Defining Activity under the Conversion Stage alongwith accompanying integrity analysis implications

4.2.3 Modeling Functional Processing Activity

Following the Data Retrieval Activity, the Conversion Stage in Figure (1) implements the "Functional Processing Activity". To elaborate, as discussed elsewhere, functional processing of data covers a spectrum of functions from processing of transactions to office automation functions to management information functions to decision support functions. Specifically, transaction processing functions include classifying, sorting, adding, deleting, updating or merely transmitting, while office automation process functions cover scheduling, word-processing or merely data storage and retrieval. Coming to management information processing functions, they include activities of report generation, data management, simple modeling, statistical methods, query response, etc., while information processing for decision support functions covers activities of query response, optimization techniques, modeling, simulation, etc.

These process functions are applicable to all business areas like production, finance, marketing, personal, R&D, strategy planning, etc. and at various levels – planning, monitoring, supervisory, and operational, and use quantitative as also judgmental data processing involving forecasting, evaluation and selection, and are characterized by man- IS interface environments, which as discussed elsewhere are prone to errors.

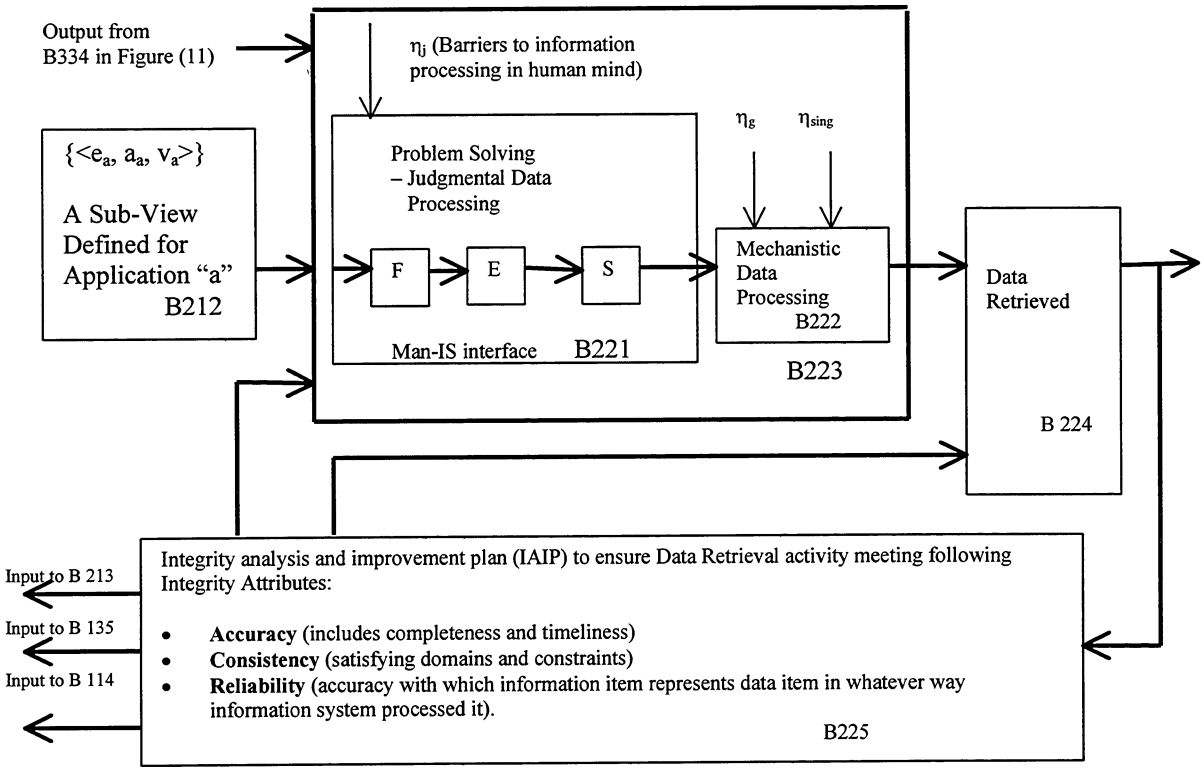

Figure 7 : Information Flow Model for Data Retrieval Activity under Conversion Stage alongwith accompanying integrity analysis implications

As output of Data Retrieving Activity, which itself is susceptible to inaccuracy, inconsistency and unreliability, is processed through the "Functional Processing Stage" under the Conversion Block, these errors have further integrity implications on the output of the Functional Processing Stage, which now is the functionally processed data termed as "Information". It is in the presence of all error components as above that, from the point of view of emerging Information Flow Model, the requirement is to ensure that information obtained as output of the Functional Processing Stage, which is the final stage in the Conversion Block, is Accurate (includes Completeness and Timeliness), Consistent (Domains and Constraints satisfied) and Reliable. Inability to assure this requirement then leads to loss of integrity in information as at this stage of information flow model building.

Accordingly, Figure (8) presents systems representation of Information Flow Model for Functional Processing Activity under the Conversion Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications.

4.3 Output Stage

With information flow modeling investigation completed upto Conversion Block, one is left with only the third and last block in the Basic IS Model in Figure (1) which pertains to "Output Stage" and covers Information (Output) Presentation and Information Use. Specifically, Information Presentation comprises following activities :

a) Identification of format (f) for information presentation.

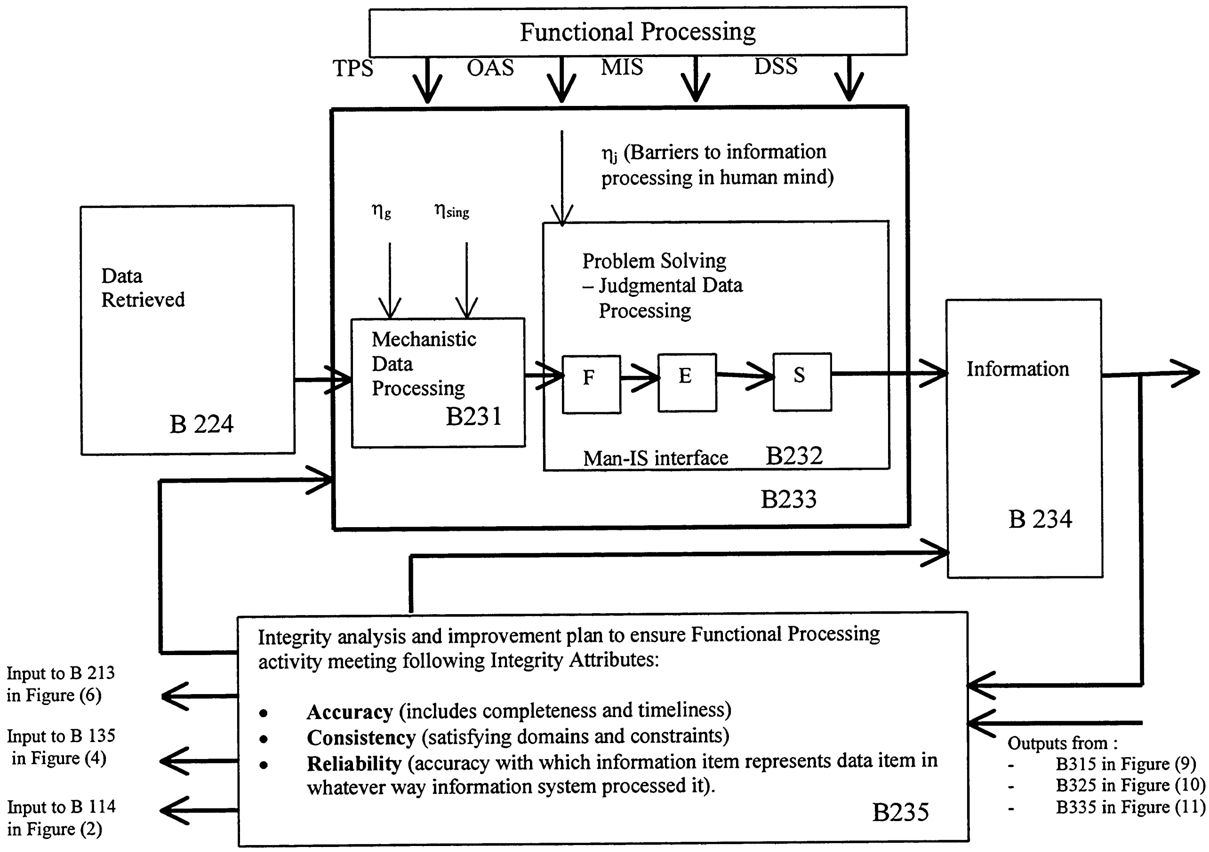

Figure 8 : **Information Flow Model for Functional Processing Activity under Conversion Stage alongwith accompanying integrity analysis implications**

b) Processing of output of Conversion block (information) in the format identified.

4.3.1 Modeling Information Presentation Activity

As can be seen, identifying information presentation format involves judgmental data processing resulting from presence of man – IS interface environments and, therefore, this activity will carry accompanying errors discussed earlier. Even then the Information Flow Model requirement is that representation is sufficiently precise and distinguishable by users (Accuracy), that representation symbols used are in accord with their formats (Consistency) and that format is appropriate, i.e., it meets user needs (Reliability). If these requirements are not ensured there would be integrity loss at this part of the Information Presentation activity.

Accordingly, Figure (9) presents systems representation of Information Flow Model for Information Presentation Activity under the Output Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications.

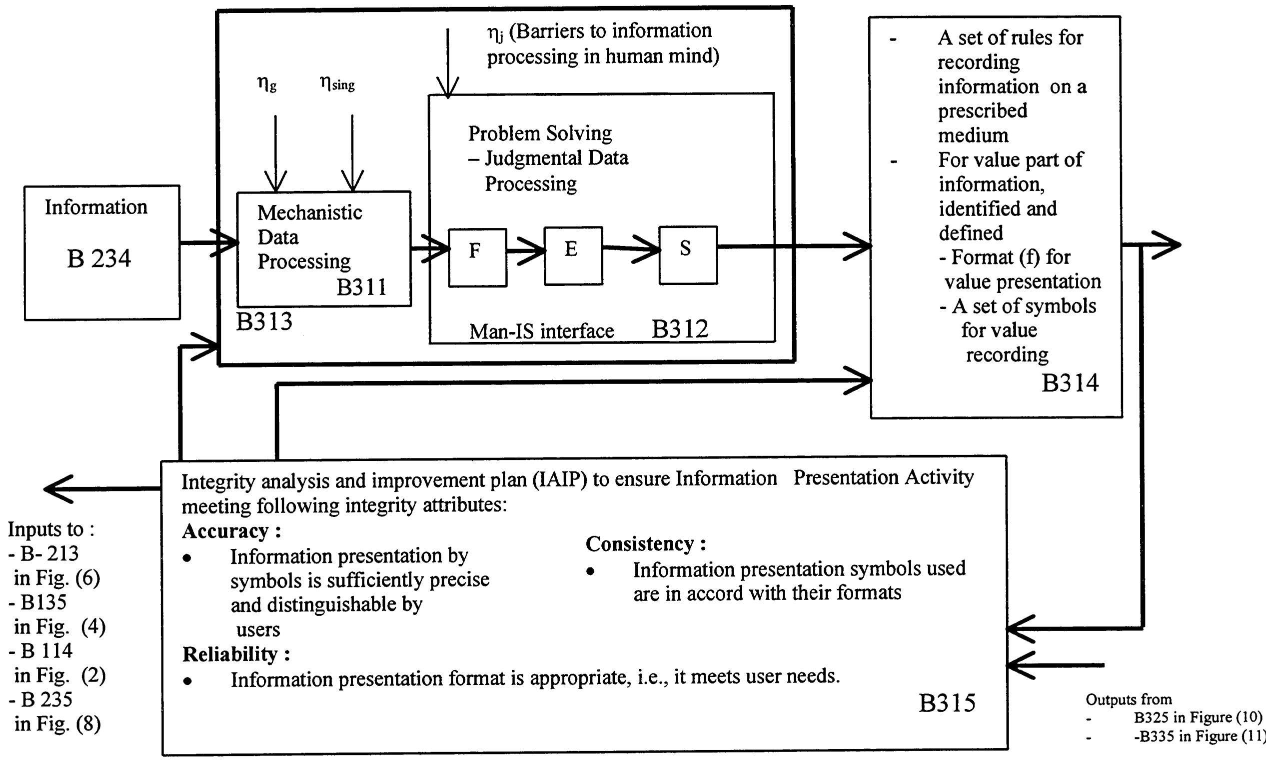

Figure 9 : **Information Flow Model for Information Presentation Activity under Output Stage of Basic IS Model alongwith accompanying integrity analysis implications**

4.3.2 Modeling Information Obtaining Activity

Coming to the processing of output of Conversion Block (i.e., Information) in the format identified, it is likely that in many situations users will be on-line involved in this activity. Thus there are possibilities of programme errors due to factors of singular type discussed earlier and data entry errors due to judgmental data processing involving man – IS interface environments. Even then requirement is to ensure that information presented at this stage of the Information Flow Model is Accurate (includes Completeness and Timeliness), Consistent (satisfies Domains and Constraints) and Reliable. Inability to ensure this would result in loss of integrity at this stage of the "Output Block" in the Basic IS Model of Figure (1).

Accordingly, Figure (10) presents system representation of Information Flow Model for Information Obtaining Activity under the Output Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications.

4.3.3 Modeling Information Use Activity

The final activity in the "Output Block" is use of information by the user. Users can be individuals, organizations as also software applications and machines and all of them are to use the information in shared environments of today characterized by on-line, rapid access. Once

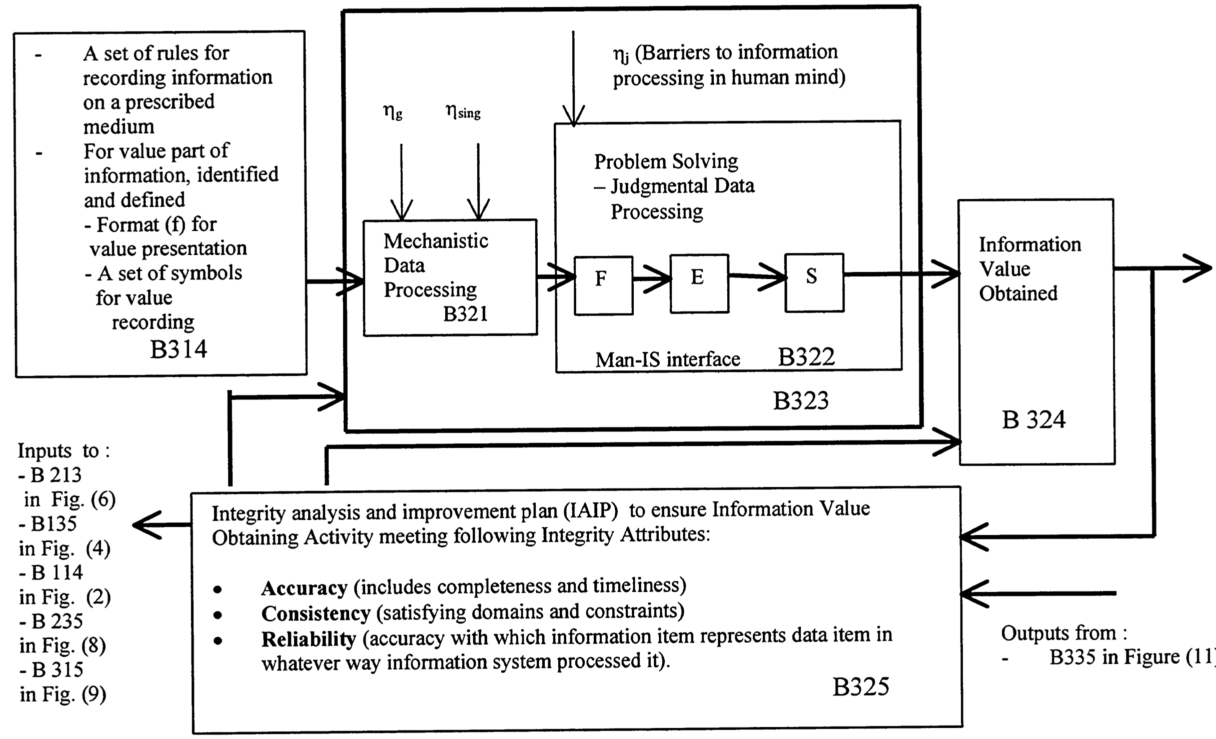

Figure 10 : Information Flow Model for Information Obtaining Activity under Output Stage alongwith accompanying integrity analysis implications

again all this may contribute programme and data errors due to singular factors as also judgmental factors and also errors due to factors of general type defined earlier. Finally, in changing business environments, users may even come with requirements not anticipated and when frustrated in the process take irrational actions like attempts to beat the system. This will also add to errors in information when under use. And in spite of all these errors, Information Flow Model requirement is information that is Accurate, Consistent and Reliable. Inability to meet these requirements as the information is used would then result in loss of integrity.

Accordingly, Figure (11) presents systems representation of Information Flow Model for Information Use Activity under the Output Stage in Basic IS Model of Figure (1) alongwith the accompanying integrity analysis implications

This then completes each data processing activity wise Information Flow Model construct in respect of the Basic IS Model.

It can be seen that data/information records flowing through each data processing activity stage can be analyzed to see if respective integrity attributes are met. If not, the model development facilitates integrity improvement plan which depending on need can correct the record or have the record re-obtained or improve on the data/information representation or even improve the sub-view or view as the case may be. This is achieved by essentially incorporating feedback loop for integrity analysis and improvement plan (IAIP) for each activity under different stages of the IFM. As a result, the Information Flow Model has the convenience

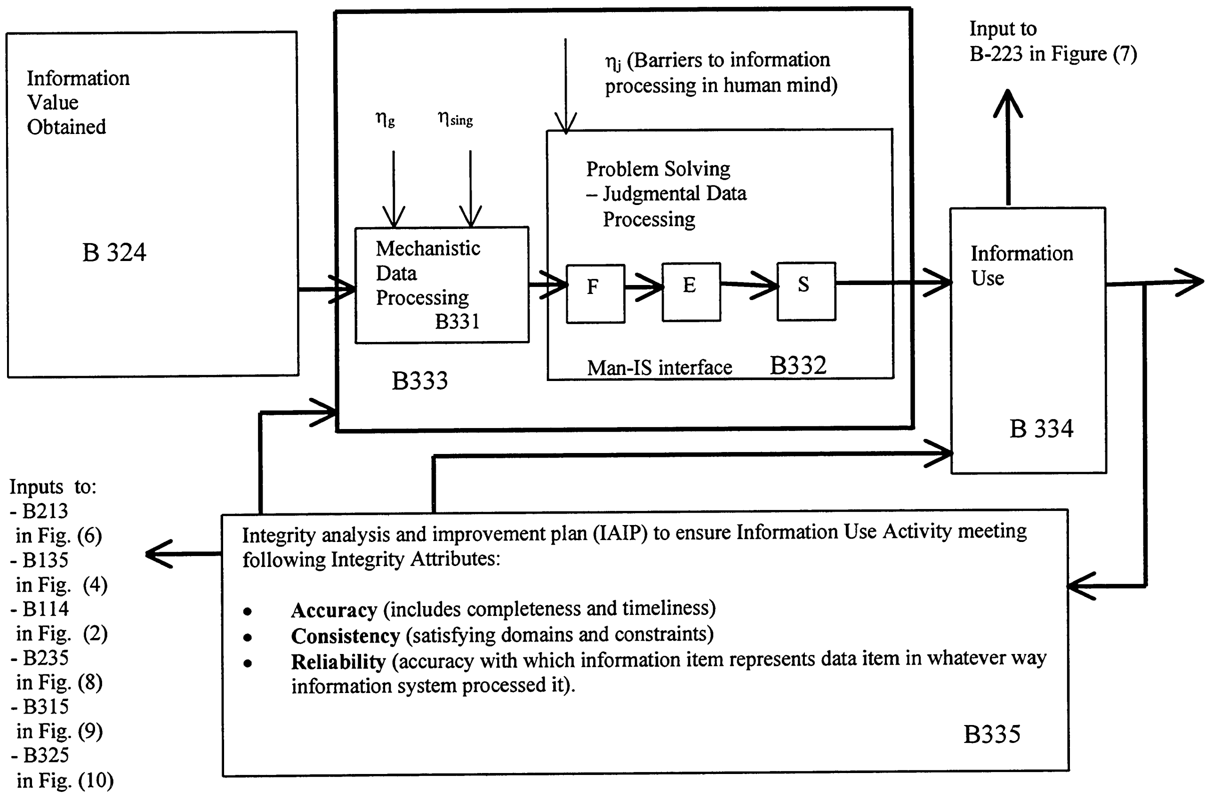

Figure 11 : Information Flow Model for Information Use Activity under Output Stage alongwith accompanying integrity analysis implications

of immediate correction for integrity improvement, at whatever stage the integrity loss might have occurred.

Finally, the Information Flow Model suggested incorporates the impact of people and changing business and technological environments on the information integrity. Towards this, the approach taken is to view IS in its comprehensive entirety in the sense all the three stages, namely, Data Origin Stage, Conversion (Application) Stage and Output (Information Use) Stage are seen together as a whole in developing the Information Flow Model.

5. Conclusion

A workable approach to developing Information Flow Model (IFM) with capability for integrity analysis and improvement plan as above is to view IFM in its totality by taking Data Origin Stage, Data Conversion (Application) Stage and Output (Information Use) Stage of the IFM together. For each of these stages then the Information Flow Model can be developed. This requires defining intrinsic integrity attributes of Accuracy, Consistency and Reliability for each of the activities from Data Origin Stage (namely, View Defining, Data Representation, Data Value Obtaining and Data Storage), from Data Conversion Stage (namely, Sub-view Defining, Data Retrieval and Functional Processing), and Output Stage (namely, Information Presentation, Information Value Obtaining and Information Use), so as to provide the basis for integrity analysis.

Within this framework Section (4) has presented information flow models for data processing activities under three stages of IFM alongwith accompanying integrity analysis and improvement plan (IAIP) implications. Understandably, these information flow models provide basis for generating error databases for implementing integrity analysis leading to integrity technologies discussed in literature [7].

On immediate basis, amongst other aspects, development of information flow models as here presents a question of quantifying respective integrity attributes in respect of various IFM activities. While this aspect has been studied in literature [7] for "Value Obtaining Activity" and that, too, for a simplistic case when data/information item takes numerical value, for all other situations there is a need for further rigorous research as it is only then the integrity analysis techniques as proposed in literature [7] could be effectively applied to the Information Flow Model developed here.

References

1. Bather H., "Meeting Managers' Information Needs", A Managing Information Report, Published by The Association for Information Management", London (1998)
2. Kroeber D.W. and Watson H.J., "Computer-Based Information systems", Macmillan Publishing Co., NY (1984).
3. Langefors B., Infological Models and Information User Views", Information Systems, Vol. 5, No. 1, pp. 17-32 (1980)
4. Lyon J.K., "The Database Administrator", A Wiley – Interscience Publication, NY (1976).
5. Mandke Vijay V., and Nayar M.K., "Information Integrity – A Structure For its Definition, Procedings of 1997 Conference on Information Quality, Edited by Diane M. Strong and Beverly K. Kahn, MIT, Cambridge, Massachusetts, USA (1997).
6. Mandke Vijay V., and Nayar M.K., "Design Basis for Achieving Information Integrity – A Fedback Control System Approach", IFIP TC 11 WG 11.5 second Working Conference on Integrity and Control in Information Systems, Edited by S. Jajodia, W. List, A.W. McGregor and Leon A.M. Strous, Kluwer Academic Publishers, London (1998), pp. 169-190.
7. Mandke Vijay V., and Nayar M.K; "Information Integrity Technology Product Structure", Procedings of 1998 Conference on Information Quality, Edited by Indu Shobba Chengalur Smith and Leo L. Pipino, MIT, Cambridge, Massachusetts, USA (1998).
8. Matthews Don Q., "The Design of the Management Information System", Auerback Publishers, NY (1971).
9. Modell, Martin E., "Data Analysis, Data Modeling, and Classification", McGraw-Hill Inc., NY (1992).
10. Myers W., "Uncover Data Aquisiton Errors",. Electronic Design (May 10, 1975).
11. Rajaraman, V., "Analysis and Design of Information Systems", PHI New Delhi (1991).
12. Redman T.C., "Data Quality : Management and Technology", Bautam Books, NY (1992).
13. Vandoren Arnold, "Data Acquisition Systems",. Reston Publishing Company Inc., Virginia (1982).
14. Willcocks Leslie and Mason David, "Computerising Work – People, Systems Design and Workplace Relations", Paradigm Publishing, London (1987).