Project SL.NO. (6): Based on the Frameworks for Projects at Sl. Nos. (4 -5), develop Information Integrity Attributes.

Framework:

Information Integrity Attributes: A Heuristic Treatment

As can be seen, errors in information systems result in loss of integrity at each stage of the information system, and thereby, in the loss of overall system integrity. This integrity loss is in terms of the attributes (not to be confused with entity attributes referred to in this Section and Paper) of: accuracy (purity), completeness (entireness, wholeness), data/information being up-to-date (i.e. timeliness implying accuracy inspite of time related changes in data/information), security and privacy (unimpaired meaning undamaged; purity).

Let us consider data/information modeled as a triple < e, a, v > as suggested in Framework for Project at SL. No. (4) above. As explained earlier, this affords a very meaningful approach whereby integrity attributes for an information system can be considered by studying them (integrity attributes) for the components of the triple , i.e., entity, attribute and value.

To elaborate, a universe for a company may comprise "employees", "products" and "customer orders". Employees and products represent "entity classes or types" and customer order represents "relationship" in the universe.

One can consider entity class, namely, employees class. As explained elsewhere, it may be represented by attributes as follows:

EMPLOYEE = (Employee Number, Name, Department Number, Salary, Date of Birth, Sex)

An entity is a specific occurrence of an entity type. Categories that make up an entity type are called attributes. The entity type shown above lists the following attributes: Employee number, Name, Department Number, Salary, Date of Birth, Sex. Attributes either identify specific entities (for example, Employee Number), or specify particular facts or properties about entities (for example, Date of Birth).

Each entity attribute has a domain assigned to it; a domain is a set of permissible values. For example, the Salary attribute may have the domain from $10,000 through $20,000. In addition, one or more constraints can be imposed on admissible attribute values. For example, values for the Salary attribute may be subject to the constraint that the employee's salary can not exceed that of his or her superior.

Finally, a value provides information for specified attribute of a specified entity. For example, for employee entity, Albert, entity representation may be as follows:

Albert = (94256, Albert, 9, $15000 p.a., 6.5.75, M)

In such a case, Date of Birth attribute has the value 6.5.75 and Salary attribute has the value $15,000p.a. and so on.

With data/information model (in terms of triple) illustrated above, a clearer picture of integrity attributes could be obtained. Suppose the Information System Model delivers information about an organization - a company. Let "view" of this information comprise, as explained above, entity types "employees" and "products" and relationship "customer orders". Further, let entity class, namely, employee class, have view model as illustrated above.

Then to study accuracy of the information on the company, one may study accuracy of the information on entity types, namely, employees, products and on relationship customer orders. Further, to study accuracy of information on entity type employees, one may study accuracy of information about attributes corresponding to entity type employees. Finally, to study accuracy of information about attributes, one may study accuracy of values for attributes; thereby making the exercise of studying accuracy of information on the company a viable exercise.

In specific terms, given the employee entity representation, namely:

Albert = (94256, Albert, 6.5.75,M)

the accuracy of information in respect of the entity type could be studied by studying accuracy of information about value of its each attribute. Above representation gives value of Date of Birth is known, then it is possible to ascertain the accuracy of the information item from the information system. In this case, it is further possible to quantify the accuracy by computing the difference between the actual and the obtained Date of Birth.

However, suppose the information obtained gave values on the specific entity Albert as follows:

Albert = (94256, Albert, 9,15000, 6.5 --- ,M)

As can be seen, here in the value of the Date of Birth attribute, year is missing. Thus, the information is incomplete. Incidentally, information, therefore, is also inaccurate. What thus emerges is for information to be accurate, it should also be complete; but every complete information is not necessarily accurate.

Let us consider another situation, this time concerning Salary attribute. Suppose Albert's Salary history is as follows:

Table (6.1) : Salary History for Entity Albert

Year	Salary $
1992	12000
1993	13000
1994	14000
1995	15000
1996	16000

If the information obtained for entity Albert as above is as of 1996, then value of $15000 pa.m for Albert's Salary attribute is not up-to-date and, hence, inaccurate for the year 1996, though the value was correct for the year 1995. This once again suggests that requirement of "timeliness" is also necessary for "accuracy", though not sufficient.

Finally, from the point of view of information model, requirement of security, meaning undamaged information, is analogous to accuracy, as any damage to information i.e., say to the value of an attribute will only result in inaccuracy of the value.

Security also has an aspect of confidentiality. Further, security of information is also important from the point of view of privacy. However requirements of confidentiality and privacy, though they emerge as implications of errors in the information system, can not be considered central requirements for all information systems, as there can be information where confidentiality , and, for that purpose, security, and privacy may not be required.

Thus from the set of integrity attributes of accuracy, completeness, timeliness, security and privacy identified above, attribute of accuracy is central to an information system and attributes of completeness and timeliness are necessary for the attribute of accuracy. In other words, attributes of accuracy, completeness and timeliness are intrinsic to an information system irrespective of use of the information derived from the system. Against this, requirements of security in the sense of confidentiality and of privacy are optional to an information system and depend on the context and nature of use of information.

There are two other requirements that have not emerged in the integrity analysis so far and they are consistency and reliability of data/information.

Specifically, like completeness and timeliness requirements, consistency requirement is also a part of accuracy requirement, i.e.., if data/information is accurate, then it is also consistent, but otherwise is not true. To clarify, consider the example of entity class employees with value for specific entity Albert as follows :

Albert = (94256, Albert, 9,16000, 6.5.75,M)

with salary attribute having domain through $10,000 to $20,000 and with Albert's Salary history from 1992 to 1996 starting with initial salary of $12,000 and ending with $16,000 with yearly increment of $1,000 as given in Table (1) earlier. Further, there is, say, a constraint that Alberts's Salary may not exceed his superior's salary which happens to be $17,000.

As can be seen, in the above example, value of Alberts's Salary attribute which is $16,000 as also all values under his salary history are within the domain range of $10,000 to $20,00. Further, Albert's Salary attribute value $16,000 is less than his superior's salary value which is $17,000. Hence information on Albert's Salary attribute value satisfies the domain as well as the constraint. Therefore, the information can be seen to be consistent.

Coming to the requirement of reliability its origin may be seen in the very choice of the

Information System Model, wherein system output, i.e., information, is defined as what user receives as an aid in action or in management or in decision. Here no user is defined as such, but "utility" or "use" role of information is brought out. It is in this context, that the requirement emerges that information obtained be reliable.

Various statistical methods are possible to study the above question- analysis of Variance (AOV) technique can be adopted. each question for a student. Of course, what is suggested here, is perceiving reliability as an accuracy with which the information obtained represents the data item in whatever respect the information system processed it and a methodology to quantify reliability for those attributes for which values lie on the real line. If it is not so and which often would be the situation in information systems one comes across, then one will have to once again look into the question of how reliability can be quantified (indeed so is the case with examples given earlier in respect of other integrity attributes, too).

In other words, immediate effort here is not to offer an all pervasive definition of reliability, but to state significance and centrality of such integrity requirement in respect of an information system. The attribute of consistency, being part of accuracy, also comes under this category.

Thus, taking entire discussion together, irrespective of the nature of use of the information obtained from the information system, attributes of accuracy, completeness, timeliness, consistency and reliability emerge as intrinsic attributes that an information system must meet, while attributes of security and privacy are optional depending on the context and nature of use.

There is more to the intrinsic integrity attributes mentioned above. As pointed out earlier, attributes of completeness and timeliness are necessary for accuracy. That is to say when checked for accuracy, the value of the information item also gets checked for its