SEPARATING OPERATIONAL FROM DSS: SOME CRITERIA

BY

W. H. Inmon

"Be not the first by whom the new are tried,
Nor yet the last to lay the old aside."
Alexander Pope

There are many reasons why the information systems architect needs to separate operational data and processing from DSS/data warehouse data and processing. The most obvious reason is that certain kinds of processing and requirements belong in one environment and other kinds of processing and requirements belong in another environment. When there is a misunderstanding as to which environment processing and requirements belong in, then there is potential for a major error of design to occur. Stated differently, when processing and requirements end up in the wrong environment, all sorts of problems begin to appear which can only be resolved by moving the processing and requirements elsewhere.

When processing and requirements are placed in the incorrect environment, no further amount of design and engineering can rectify the mistake. Throwing financial and technical resources at an incorrectly positioned process or requirement is a waste and serves to magnify the incorrect design choice.

As information systems organizations mature, they progress from a mind set of having one computer environment serve all the information processing needs to the realization that multiple, specialized environments are needed. The first two environments they discover that are needed are the operational environment and the DSS informational environment. The organizational maturation process is fostered by at least three factors:

- the size of the organization and the amount of processing being done. Small organizations or organizations just beginning their information processing systems do not see the need for the separation between the different types of processing. It is only after an organization has grown that the need becomes to be apparent,
- the complexity of processing. Organizations that do very simple and very straightforward processing often do not see the need for anything other than a single information system that serves all communities. As systems grow and as processing becomes more complex, there is the realization that - from the standpoint of complexity alone - a separation of environments is called for,
- aging. As long as systems are new and fresh, the single information processing environment appears to be sufficient for an organization. But as systems age, the requirements that were originally behind the systems in the first place change and the systems are maintained. The process of changing requirements and doing long-term maintenance to systems greatly increases the complexity of what originally started out as a simple information processing environment.

The normal progression for almost all organizations is that - over time - the systems grow, become maintained, experience changing requirements, and become complex. One day the organization wakes up and finds that the paradigm of a single set of information systems is inadequate to get the shop into the future. A new paradigm is

needed and that paradigm is one where there is a distinct separation between information systems that serve the day to day operational needs of the organization and an information system and its associated data warehouse that serve the informational needs of the organization.

When an organization reaches the point of maturity of making this observation, a very fundamental question must be asked:

"WHAT IS OPERATIONAL PROCESSING AND WHAT IS DSS/INFORMATIONAL PROCESSING?"

Figure 1 shows that requirements are divided into one or the other camp.

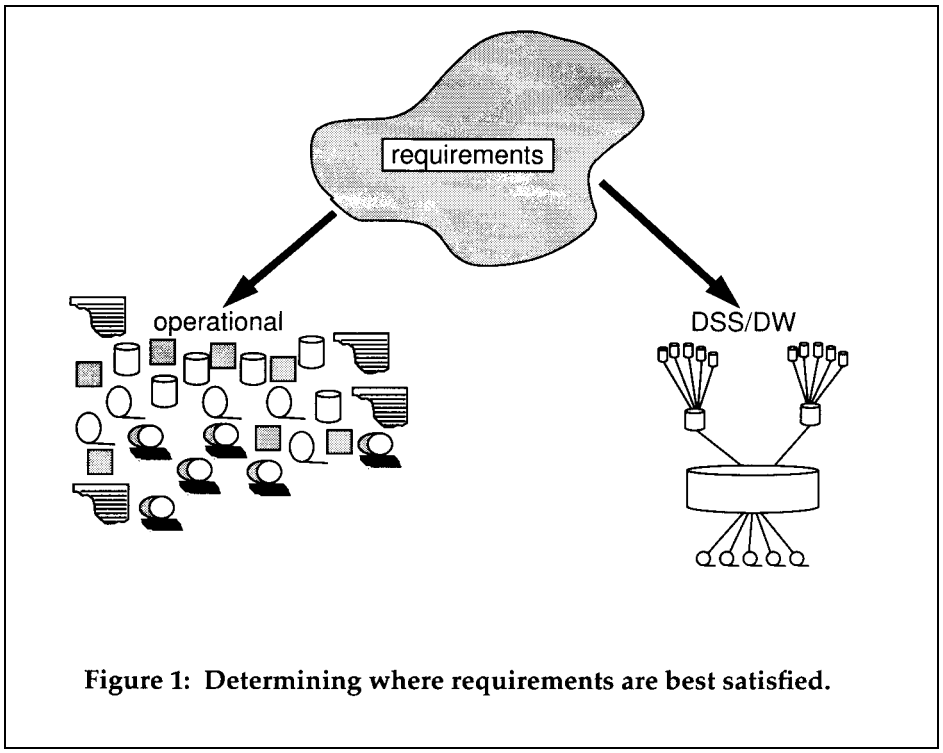

Figure 1: Determining where requirements are best satisfied.

The successful organization will answer this question properly. The unsuccessful organization will not answer the question properly and will put at risk much wasted development and much wasted opportunity. It is EXTREMELY IMPORTANT that a shop be able to answer the question quickly, astutely, and knowingly.

In many cases the differences between operational and DSS processing are obvious. In most instances the dividing lines between the two environments are very clear. But on other occasions there is a very narrowly defined line between the two worlds. Infrequently it may not be clear at all what is the difference between operational and DSS, informational processing.

This white paper will help to delineate the difference between the two environments. Explanations, examples, and diagrams will be used to point out the many differences between the two environments.

COMMUNITIES

The operational environment serves the clerical community. The DSS/DW environment serves the managerial community. Figure 2 shows these two different communities and how they are served.

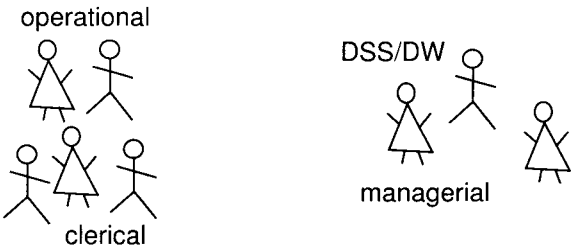

Figure 2: Operational systems feed the clerical community; DSS/DW systems feed the managerial community.

As an example, the bank teller (i.e., the financial clerical community) is served by the system that tells the teller whether a check can be cashed for a customer that is standing at the counter. The customer check cashing system then is an operational system.

The vice president of loans for the bank is served by another system that tells the vice president how account activities have been different for the bank this year from the past five years. The system that addresses lots of accounts and how they have been acting over the past few years is an informational, DSS system.

One way then to tell the difference between operational processing and DSS informational processing is by the community being served.

TYPES OF DECISIONS

The operational environment is used in the making of day-to-day decisions. The DSS informational environment is used to make long term, directional decisions.

Figure 3 illustrates the difference between the two environments with regard to the types of decisions being influenced.

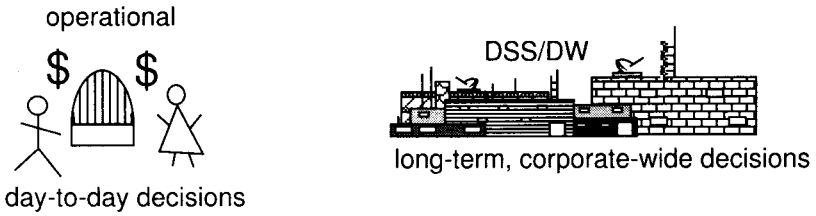

Figure 3: Day-to-day decisions are made on operational data and systems; long-term, corporate-wide decisions are made on DSS/DW data and systems.

As an example of day-to-day decisions, an insurance company has a system that tells - on the spot - how much and what kind of coverage one of its customers has. The system is used by agents and adjusters in the day-to-day handling of claims. Such a system would be an operational system.

The insurance company also has a system that looks at all policies in force and analyses the total risk that the company is insuring. The system is used to decide what rates policies should be sold at, whether a new line of business ought to be entertained, and whether the risk being taken by the company as a whole is too much. This system is an example of a DSS informational system.

A good way to tell the difference between operational processing and DSS informational processing is by the type of decision being made by the user of the system.

DETAILED / SUMMARY DATA

The operational environment contains detailed data, almost exclusively. There is very little other than detailed data that is found in the operational environment. The DSS informational environment contains a mixture of both detailed data and summary data. Indeed, there are different levels of summary data found within the classical architecture of the data warehouse, such as lightly summarized data and highly summarized data.

Figure 4 shows the difference between the two environments insofar as the level of detail is concerned.

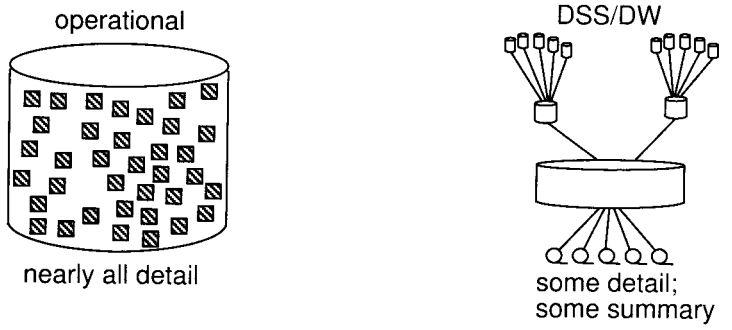

Figure 4: The operational environment contains nearly all detail; the DSS/DW environment contains a mixture of detailed and summarized data.

As an example of detailed operational data, an operational system may contain the details of how many parts are currently on hand for the storeroom of a manufacturer. An informational environment may contain how much total product has been built and shipped, part-by-part, for the past year.

Another approach to tell the difference between operational processing and DSS informational processing is by the level of detail. Operational environments contain almost exclusively detailed data; DSS informational environments contain a mixture of both detailed and summary data.

UPDATE/LOAD AND ACCESS PROCESSING

General-purpose record update - inserts, deletes, changes, and accesses of records of data - occurs in the operational environment. The ability to locate and alter a record if needed is a hallmark of the operational world. Data is loaded into the data warehouse. Once the data is loaded, it is then accessed there by the DSS community. But once loaded into the data warehouse, the data is not altered, under all normal circumstances.

Figure 5 outlines the difference in update characteristics between the two environments.

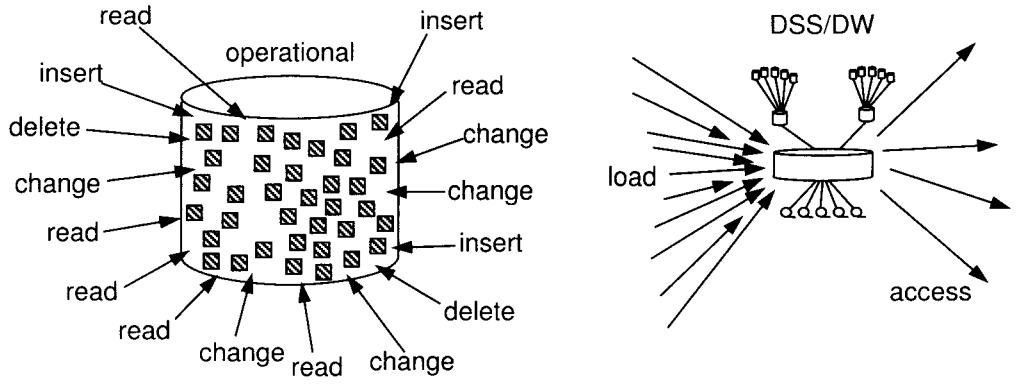

Figure 5: In the operational environment, data is updated (inserted, deleted, changed, and read); in the DSS/DW environment, data is loaded and accessed.

As an example of update in an operational environment, a customer can execute an ATM activity and on the spot change the bank balance for the account. The account balance is altered by means of an update to a single record.

A DSS informational system has snapshots of data loaded into the data warehouse. Once those snapshots are loaded (assuming they are loaded correctly), they are not altered again.

One technique then to clarify the difference between operational processing and DSS informational processing is to determine whether general purpose, record update is occurring as a normal matter of course. If update is occurring, then the system is operational.

CURRENT VALUE / HISTORICAL DATA

Operational processing operates on data that is assumed to be accurate as of the moment it is accessed. In other words, operational data is up to the second accurate. If a business condition changes, the operational data related to that business condition quickly changes. DSS informational data is quite different. The world of informational processing is one of historical data. Each record in the DSS informational environment is a "snapshot" of some aspect of the organization's environment. When a change occurs in the business environment, a new snapshot is added to the DSS informational environment.

Figure 6 describes the two kinds of data found in the different environments.

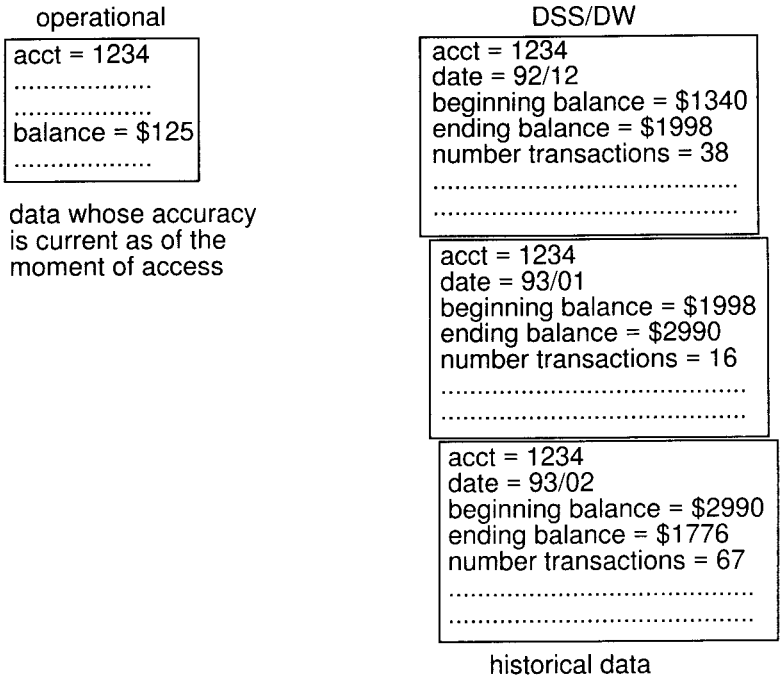

Figure 6: Operational data is data that is accurate as of the moment of access; data warehouse data is historical.

As an example of operational, current value data, an operational system has a record of the up to the second balance of an account. As an example of DSS informational data, an informational system has a record of all the money that has been invested in long-term bond accounts as of a given point in time.

A good way then to tell the difference between operational processing and DSS informational processing is by the nature of the data found in the systems in the different environments.

TRANSACTION PROFILE

The operational environment contains transactions that are small, use few resources, and have a predictable arrival (usually fast!) rate. In addition, many operational transactions are running at any moment in time. DSS informational transactions are large, use many resources, and have a very unpredictable arrival rate. In addition, there are seldom many DSS informational transactions that are being run at the same time. Typically an operational online transaction will use 20 or 30 I/O's. A DSS informational transaction, on the other hand may consume as many as 500,000 I/O's.

Figure 7 indicates the fundamental differences in transactions that are found in the different environments.

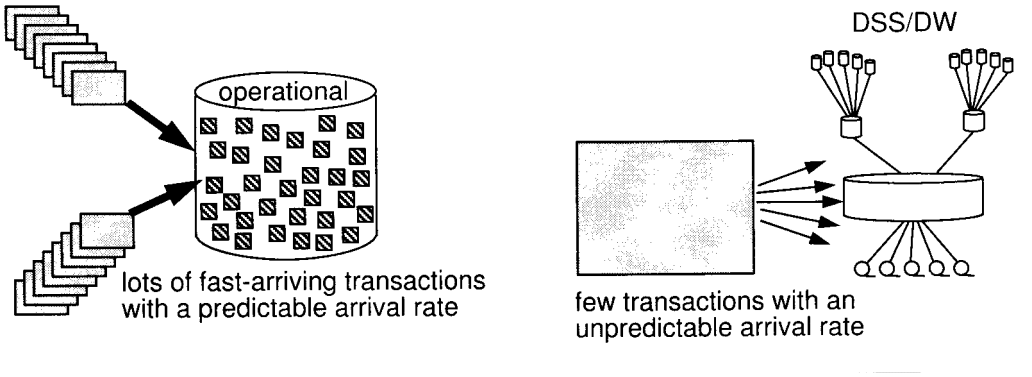

Figure 7: The operational environment is one that contains many small, fast-arriving transactions; the DSS/DW environment is one that contains few, large transactions with an unpredictable arrival rate.

As an example of an operational transaction, an online operational transaction:

- reads an account's record
- identifies who the account owner is
- looks at the personal identification used by the owner
- displays the information to the clerk verifying id

An example of a DSS informational transaction is:

- select all records whose balance is greater than 5,000
- look for all accounts where last payment is late
- add up the late balance amount
- add up the late days
- display the totals created

A method to tell the difference between operational processing and DSS informational processing is by the difference in the type of transaction being run in the environment.

TRANSACTION DIVERSITY

The same transaction (or set of transactions) is run through the operational environment. There is much repetition of processing in the operational world. Conversely, in the DSS informational environment there are many transactions that are run through the environment. Indeed, many transactions may be run once and only once in the DSS informational environment.

The difference in the diversity of the transactions is found in Figure 8.

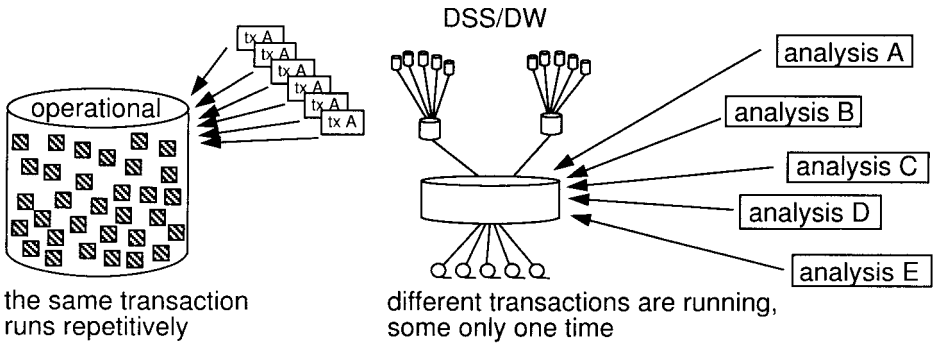

Figure 8: The operational environment is made up of the same transaction running repetitively; the data warehouse has many different transaction types running against it, many of which are run only once.

As an example, in an hours time in the operational environment, transaction A, which is used to verify the inventory on hand is run 1,000 times. Most executions of the transaction are for a different part whose inventory is being verified. As an example of a DSS informational transaction, a DSS analyst creates a transaction that looks at how many customers have opened accounts then have failed to use the account within six months of opening. Having found out how many of those customers there are, the DSS analyst issues another transaction that looks at all people who use their account more than six times per month but who pay only the minimum amount each month. The DSS analyst continues to issue one of a kind transaction until the analyst starts to have some insight into what business problem is being investigated.

The difference between operational processing and DSS informational processing is often typified by the difference in the repetitive nature of the transactions being run through the different environments.

HARDWARE UTILIZATION

The operational environment uses hardware - I/O, CPU, memory - in a static, predictable pattern. There are peaks and valleys in the usage of the hardware, but the pattern of usage is predictable, within the bounds of reasonability. The DSS informational environment uses hardware in a binary fashion. In the DSS informational environment either hardware is being used at its peak or the hardware is not being used at all. Typical of the DSS informational hardware usage pattern is the personal computer, which is being used at peak load or is not being used at all.

The difference between the patterns of hardware utilization is shown by Figure 9.

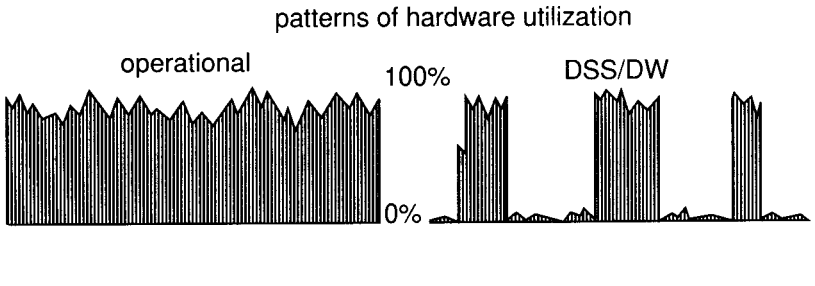

Figure 9: The pattern of hardware utilization is very different for the two environments.

As an example of operational hardware utilization, a mainframe processor experiences utilization between 76% and 99% during peak hours, with a mean of 84%. As an example of DSS informational processing hardware utilization, an executive's EIS workstation is turned off all day except for her 7:45 am review of critical numbers and for the 5:30 p.m. wrap-up.

Hardware utilization is a good way then to tell the difference between operational processing and DSS informational processing.

REPORTING

Reports in the operational environment feature contents where the line item is of importance, and where summary information is of incidental importance. The clerical community is most interested in the operational reports that are issued. Reports in the DSS informational environment feature line items that are of little or no interest, and summarizations that are of a great deal of interest. Management is most interested in DSS informational reports.

Figure 10 portrays these differences between reporting in the different environments.

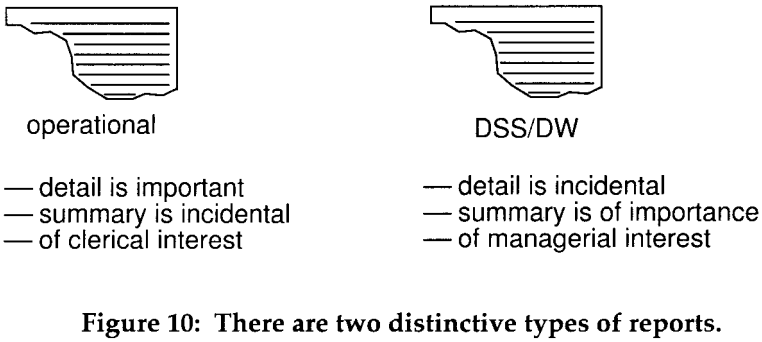

Figure 10: There are two distinctive types of reports.

As an example of an operational report, a daily audit is done of financial transactions. This audit is done against financial activities that have occurred during the day. Each

and every transaction on the report goes under scrutiny. As an example of a DSS informational report, the accountant prepares a quarterly statement that declares profits and losses. The details of how those profits and losses were made are of little consequence once the report has been issued.

One trait of the difference between operational processing and DSS informational processing is difference between the reports that are issued in the different environments.

METADATA

Metadata in the operational environment has been "optional" for as long as there has been an operational environment. The professional IS technician has been ambivalent to data dictionaries, libraries, and the like. Conversely, metadata in the DSS informational environment is THE cornerstone of success. Metadata is at the heart of making the DSS analyst successful. Metadata is the glue that holds the data warehouse together, across its many platforms.

The role of metadata in the two environments is depicted in Figure 11.

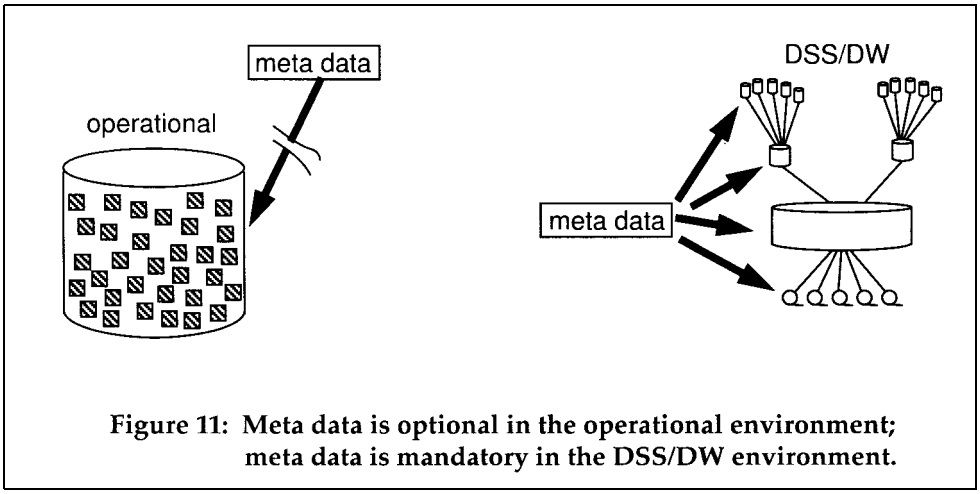

Figure 11: Meta data is optional in the operational environment; meta data is mandatory in the DSS/DW environment.

Metadata in the operational environment is typically contained in a data dictionary or a system library. Usually only the scantiest of metadata is stored, such as control block information, indexing information, and so forth.

The source of data, the transformation that has occurred, and the actual contents of the DSS informational data is only the start of the metadata to be found in the DSS informational environment.

Looking at metadata and its role is a good way to discern the difference between operational processing and DSS informational processing.

STABILITY OF CONTENTS

The contents of the operational environment are in a constant state of flux. As data is being updated, the contents of the operational data are changing all the time. On the other hand, for the DSS informational environment, once the contents of the data warehouse are written, they are not changed (except under the most exceptional of circumstances!)

The stark difference between the rate of change of data content in the two environments is seen in Figure 12.

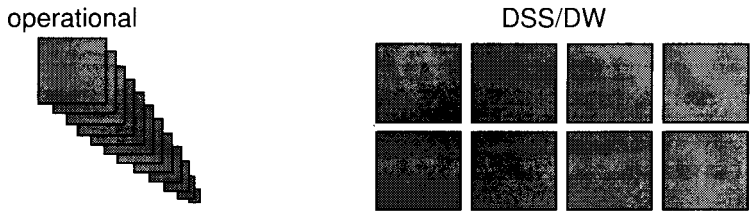

Figure 12: The contents of the operational environment are constantly changing; the contents of the data warehouse are static.

An example of operational data content that changes is a customer bank account balance. Every time a check is written and clears the bank, every time an ATM is executed, every time a deposit is made into the account, the balance is altered.

As an example of a DSS informational data content that is stable, once a month a profile snapshot is made of the account activities for a customer for a bank. Once the record is written (correctly!), the record is never rewritten.

Examining the stability of the content of data is a way to tell the difference between operational processing and DSS informational processing.

PUBLIC / PRIVATE ACCESS

Operational data is typically available across a wide network, available to many people who may need access to the data. DSS informational data is typically privately held and is available to only a few people. There are usually good reasons why DSS informational data should not be made publicly available.

Figure 13 denotes the difference in access to data in the different environments.

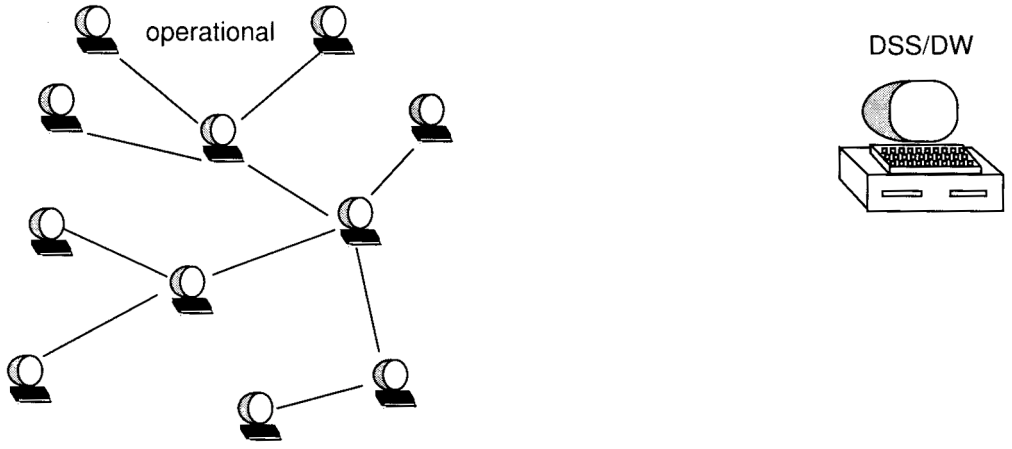

Figure 13: Data is available across a network in the operational environment; data is successively private in the data warehouse environment.

As an example of an operational network, consider an airlines reservation system. In an airline reservation system data concerning flights and passengers needs to be available in many locations, as flights are taking off and landing in many places, and as connecting information is needed on a widespread basis as well.

As an example of DSS informational access to data, an executive calculates his estimation of next quarter's revenues and profitability. The executive does not want this information being available on a widespread basis.

The right to public access of data and the need to keep data private delineate one way to tell the difference between operational processing and DSS informational processing.

SEMANTIC STABILITY

Operational data is typically very semantically stable. Once a data structure is defined in the operational environment, the data structure remains stable. On the other hand, in the DSS informational environment, when a data structure is defined, it usually is redefined as the DSS analyst goes through the iterative process of analysis.

Figure 14 describes the difference in semantic stability between the two environments.

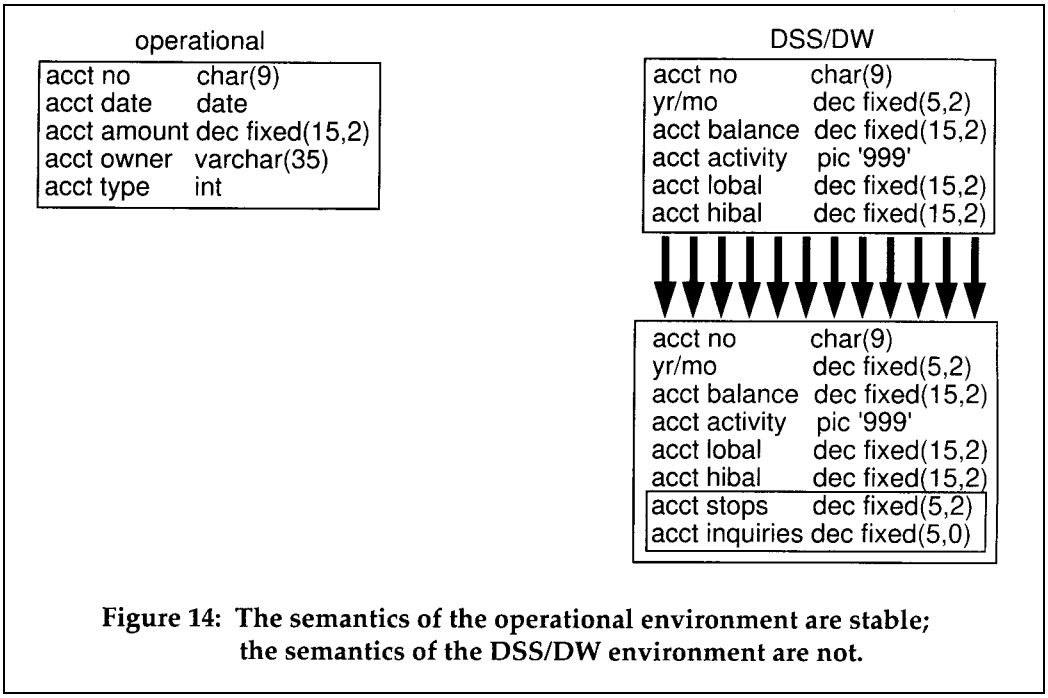

Figure 14: The semantics of the operational environment are stable; the semantics of the DSS/DW environment are not.

As an example of operational semantic stability, an operational table for account balances is defined. The operational table has accounted for many requirements, which have been thoroughly gathered. It is ten years before the table needs a redefinition.

As an example of DSS informational semantic instability, a DSS analyst creates a report that summarizes monthly account balances and account profiles. Upon creating the data structure and running a report, the DSS analyst discovers that more data is needed. The DSS analyst adds the new types of data that are needed to the database and recalculates the contents.

Semantic stability differs markedly between operational processing and DSS informational processing.

SUMMARIZATION

The operational environment contains primarily detailed data. However summarization can be done in the operational environment. However, the summarization that is done in the operational environment is accurate ONLY as of the moment of summarization. Conversely, summarization is a normal and regular part of the data warehouse DSS informational environment. Both detail and summarization exist as a regular and normal part of the data warehouse. The summarization that is done in the data warehouse represents a total that is calculated for some given moment in time and can be easily recalculated again if needed.

Summarization is shown to be very different in the two environments, as seen in Figure 15.

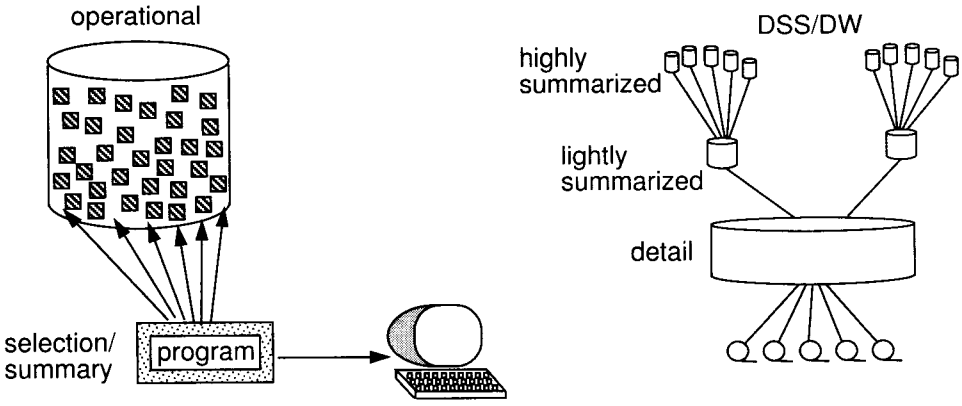

Figure 15: Operational summarization is accurate as of the moment it is summarized, but is not able to be repeated; DSS/DW summarization is accurate as of some one moment in time, is stored for repeated use, and is able to be repeated.

An example of operational summarization is the calculation of the total daily activity of all accounts taken at 10:15 am. While there is some use to the data calculated at 10:15 am, it must be recognized that if the same calculation is done at 10:30 am, that a different result will be calculated.

An example of DSS informational summarization is the monthly revenue attributable to interest. The value is stored and is never needed to be recalculated.

One way then to tell the difference between operational processing and DSS informational processing is by the difference in the handling of summary data.

REQUIREMENTS

In developing operational systems, the requirements for development are gathered and identified before the design process commences. In the DSS informational environment, in many cases the requirements for processing are only understood at the end of the DSS development process.

Figure 16 illustrates the difference between requirements and the development process in the two environments.

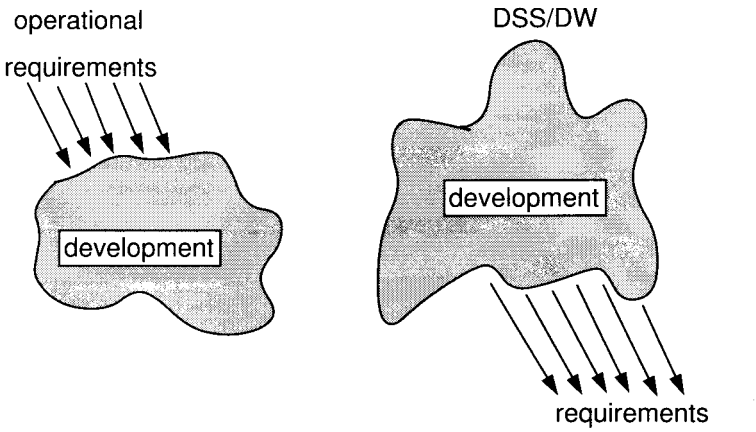

Figure 16: Requirements go into the operational development life cycle; requirements are identified by the DSS/DW development life cycle.

As an example of operational requirements, the system developer assimilates such information as expected response time, data needed for processing, and expected availability long before the operational system goes into detailed design.

As an example of DSS informational processing, it is only after much analysis over a lot of data that the DSS analysts discovers that the marketplace in the southeast is price inelastic to a new product that has been introduced, but is price elastic in the west.

The role of requirements and the relationship of requirements to the development process is a good way to tell the difference between operational processing and DSS informational processing.

SYSTEM DEVELOPMENT LIFE CYCLE

The operational environment is developed under a life cycle called the SDLC. The DSS informational environment is developed under a very different life cycle, called the CLDS.

Figure 17 shows both life cycles.

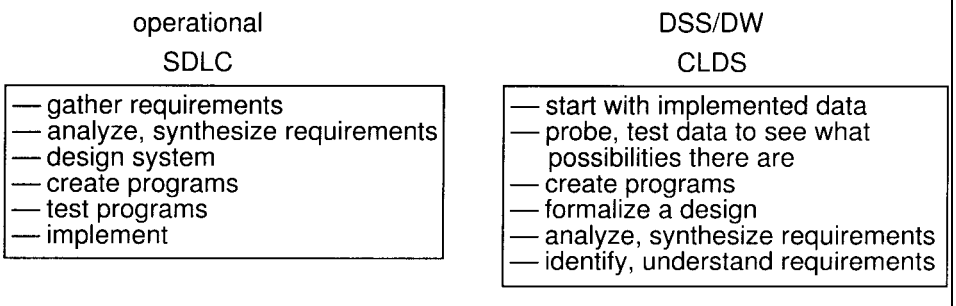

Figure 17: The development life cycle of the two environments is almost perfectly the reverse of the other.

The life cycles are essentially the reverse of each other.

TOOLS

The operational environment is served by tools such as CASE.

Figure 18 shows the different tools and how they serve their different audiences.

Figure 18: The tools that are appropriate to one environment are very different from the tools that are appropriate to the other environment.

DATA MODEL

There is an enterprise data model that serves as a basis for both the operational data model and the DSS/DW data model. Figure 19 illustrates the relationship of these different data models. (NOTE: there is a white paper on this relationship).

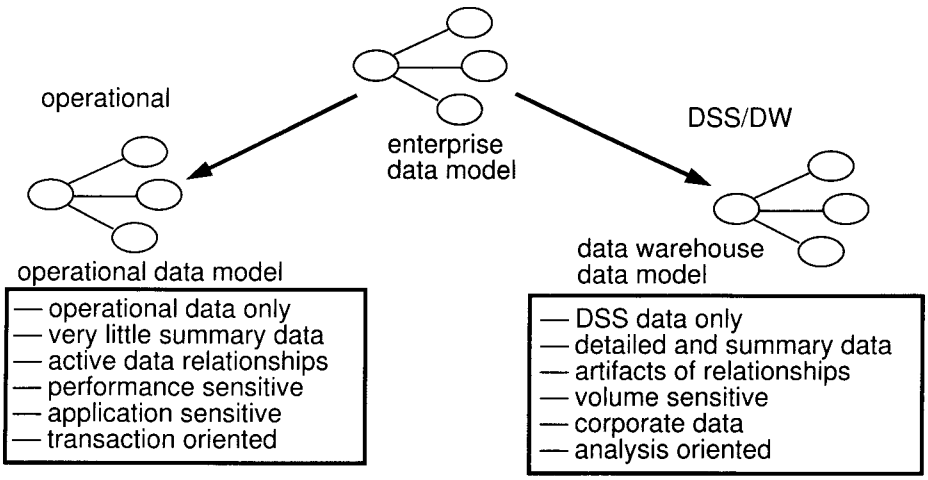

Figure 19: The enterprise data model feeds both the operational data model and the DSS/DW data model.

As an example of operational only data, an operational data model may contain a field - "call upon delivery". The field is used to specify a phone number to call once delivery of a shipment is made. This field is quite useful and necessary for operational activities. But a phone number to be called upon delivery is not useful for DSS processing.

As an example of DSS informational data that might be found in the DSS informational data model, there is the field average account balance during the month. The average account balance is of use primarily for account analysis and is not of interest to the day to day processing that occurs for an account.

APPLICATION / SUBJECT ORIENTATION

The operational environment is application oriented. Operational designs are done by applications. Operational projects are organized by applications. Operational programs are written by applications. The DSS informational environment on the other hand is oriented towards subjects. DSS informational design is done by subject. DSS informational projects are organized around subjects. Figure 20 shows the different orientation of the environments.

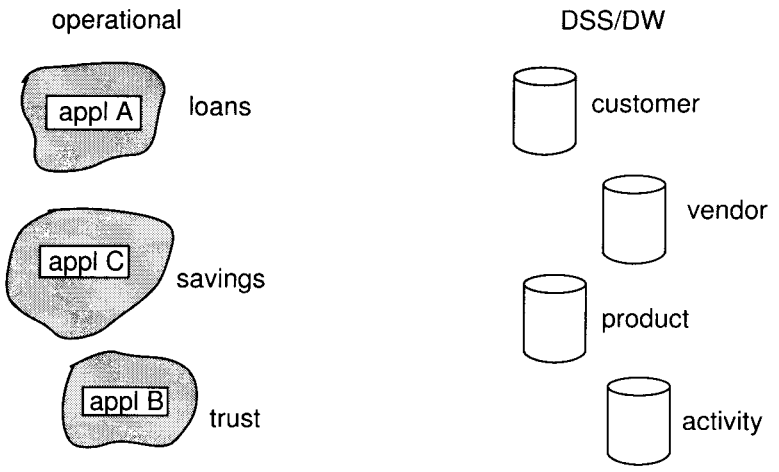

Figure 20: The operational environment is application oriented; the DSS/DW environment is corporate wide, integrated and subject oriented.

As an example of an operational application, there might be a loans application. As an example of a DSS informational subject, there might be a customer subject database.

he fundamental orientation of the way work is done is one way then to tell the difference between operational processing and DSS informational processing.

OWNERSHIP

Operational data ownership centers on the issues of the right to update data. In the operational environment one or the other group is responsible for the creation and maintenance of data. But in the DSS informational environment, since there is no update and there is only snapshots of data that exist, ownership of update is not an issue. Instead the issues of ownership of data in the DSS informational environment center on stewardship of data.

Figure 21 depicts the differences between the two environments.

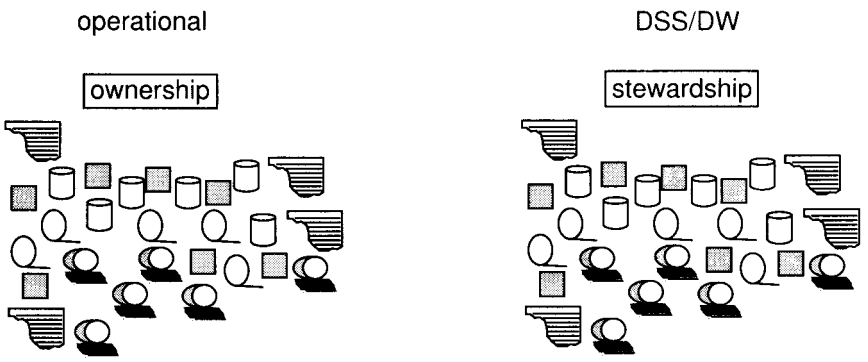

Figure 21: The approaches to data management are quite different in the two environments.

As an example of operational ownership, the accounting department has the responsibility for the definition and maintenance of the chart of accounts. No one other than an accountant can create or delete the definition of an accounting category.

As an example of DSS informational stewardship, after the snapshots are taken and the data is integrated into the data warehouse, the data architect monitors the usage of the data.

Determining whether ownership or stewardship is at stake is a way to tell the difference between operational processing and DSS informational processing.

LEVELS OF DATA

The operational environment contains data that is essentially flat. Little or no distinction is made between data detail and summarization (since there is no summarization that is permanently stored) in the operational environment. The operational environment makes little or nor distinction between the age of data, since all data is current. But those distinctions are made in the DSS informational environment.

Figure 22 shows the difference between levels of data in the two environments.

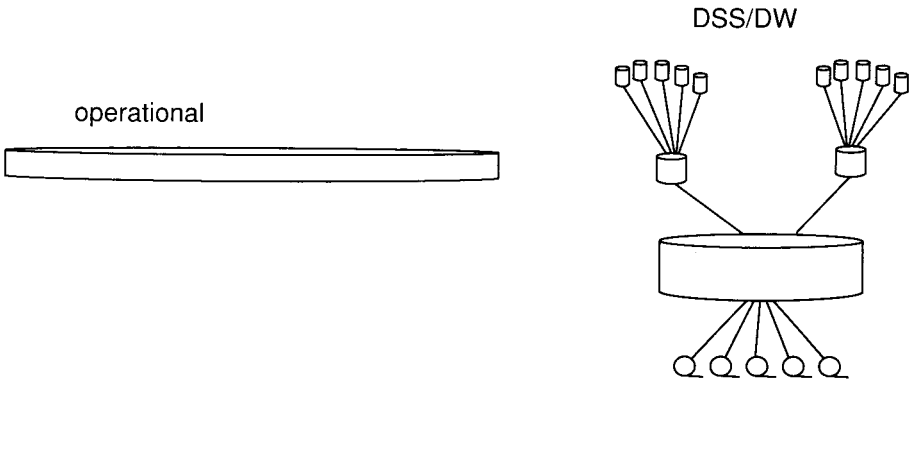

Figure 22: The level of detail in the operational environment is flat; the levels of data in the DSS/DW environment are many.

DATA RELATIONSHIPS

Operational data relationships are based on a single business rule and are implemented by such constructs as referential integrity or logical relationships. Only one business rule can be represented by an operational data relationship. The business rule that is represented is the one that is current or active. The DSS informational environment is one that contains "artifacts" of relationships. A snapshot is taken and the relevant business relationships are captured as of the instant in time when the snapshot occurred. As such through artifacts multiple business rules can be represented, and the relationship and all of its changes can be represented over a lengthy spectrum of time.

Figure 23 shows the difference between the two environments in terms of data relationships.

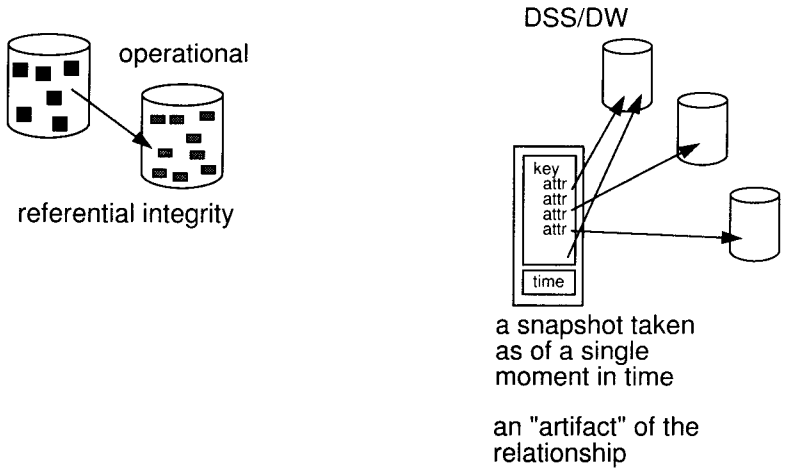

Figure 23: Data relationships look very different in the two different environments.

As an example of an operational data relationship, s part may show that it has a supplier. As an example of a DSS informational artifact of a relationship, the shipment of a part shows the data of shipment, the part, and the supplier who made the shipment.

The representation of data relationships is yet another way to tell the difference between operational processing and DSS informational processing.

DATA RELATIONSHIPS OVER TIME

The operational world is able to represent data relationships that are active and current. As such, the operational environment is able to represent only one data relationship between any two objects at the same time. But the DSS informational world requires a different treatment of data relationships. Given that the data warehouse contains data relationships (and implicitly, their underlying business rules) over a lengthy spectrum of time, there is a need to be able to represent many different manifestations of a data relationship over a lengthy period of time.

Figure 24 illustrates the differences between the representation of a data relationship over a lengthy period of time.

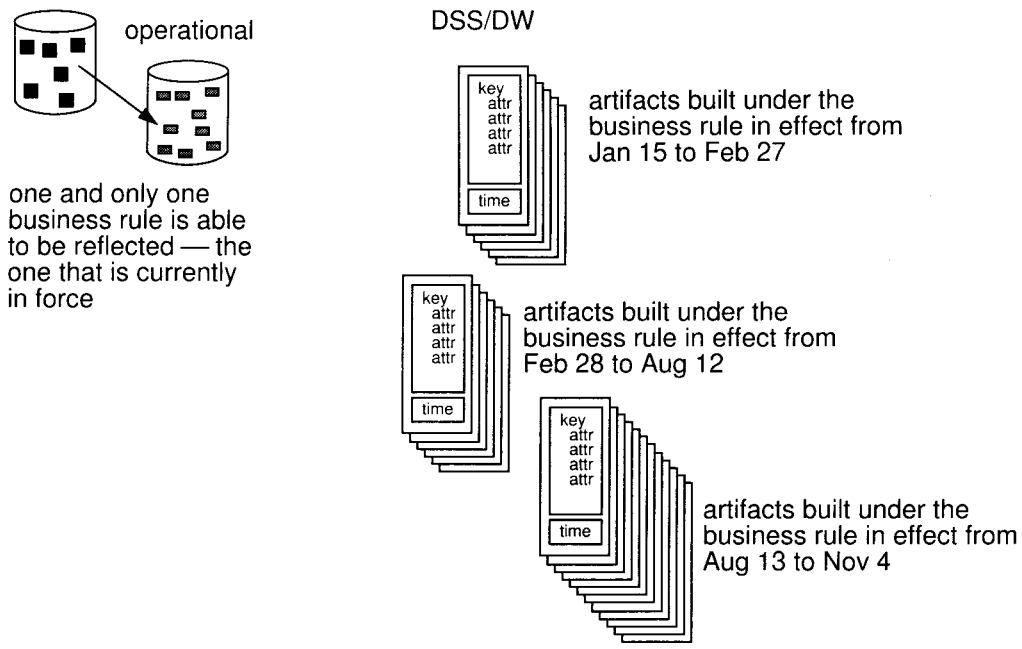

Figure 24: The ability to represent business rules is quite different from one environment to the next.

As an example of an operational data relationship, a customer may have met with a sales representative because the sales representative is assigned the sales territory. As an example of a DSS informational artifact of a data relationship, a customer may have met with different sales representatives from different organizations over a period of

time, as well as other people from the sales organizations, such as sales representatives.

IN SUMMARY

There are many different ways that the operational environment and the DSS informational environment distinguish themselves from each other. The ways that have been discussed in this Tech Topic have been:

- the community being served,
- the immediacy of the decision being made,
- the amount of detail and summarized data there is,
- whether general purpose record update is occurring,
- whether current value data or historical data is stored,
- the difference in transactions being processed,
- whether the same transaction is being run repetitively,
- the pattern of hardware utilization,
- the types of reports that are produced,
- the role of metadata,
- the stability of the contents of data,
- the network access,
- the semantic stability of the data,
- the types of summarization that can be done,
- how requirements fit with development,
- the system development life cycle,
- the tools that are used,
- the role of the data model,
- the orientation towards applications or subjects,
- the ownership of data,
- the levels of data,
- how data relationships are treated, and
- how business rules are managed over time.