PART I

IS Audit Process

CHAPTER 1

Technology and Audit

This chapter covers the basics of technology and audit. The chapter is intended to provide an understanding of the technology currently in use in business as well as knowledge of the jargon and its meaning. It also covers the components of control within an IT environment and explains who the main players are and what their roles are within this environment.

After reading this chapter you should be able to:

- Understand the technology currently in use in business
- Understand the jargon and its meaning
- Define the components of control in an IT environment
- Briefly explain who the players are and what their roles are
- Define the fundamental differences between batch and on-line systems
- Explain the principal business risks within each processing type
- Describe the components that make up the on-line system and the effect these have on control objectives
- Explain the controls within each type of computer system
- Contrast the basics of batch and on-line security
- Demonstrate an ability to:
  - Identify the differing types of database structures
  - Identify the principal components of each type of Database Management System (DBMS)
  - Identify the primary threats to each of these components
  - Relate DBMS components to the operating system environment in which they operate
  - Identify potential control opportunities and select among control alternatives
  - Identify the principal DBMS products in market

- Recognize vulnerabilities in multiple DBMS environments and make appropriate recommendations

TECHNOLOGY AND AUDIT

Some Computing Jargon

Before we can start to discuss the audit and control of computer systems, we must have a common understanding of the jargon used.

Hardware Hardware consists of those components that can physically be touched and manipulated. Principles among those components are:

- **CPU.** The Central Processing Unit is the heart of the computer. This is the logic unit that handles the arithmetic processing of all calculations.
- **Peripherals.** Peripheral devices are those devices that attach to the CPU to handle, typically, inputs and outputs. These include:
  - Terminals
  - Printers
  - Disk and tape devices
- **Memory.** Memory takes the form in modern computers of silicon chips capable of storing information. In commercial computers, this information takes the form of 1 and 0 in the notation known as ***binary***. Memory comes in various forms including:
  - **RAM.** Random Access Memory whose contents can be changed but which is vulnerable to loss of power where the contents of memory may also be lost. This type of memory is also known as ***dynamic*** or ***volatile*** memory.
  - **ROM.** Read-Only Memory is a form of memory whereby instructions are "burned-in" and not lost in the event of a power loss. These programs cannot be changed. This is also known as non-volatile memory.
  - **PROM.** Programmable Read-Only Memory is similar to ROM but can have the contents changed.
  - **EPROM.** Erasable Programmable Read-Only Memory is similar to PROM but the instructions can be erased by ultra-violet light. There is another version of memory known as

nonvolatile RAM. This is memory that has been attached to a battery so that, in the event of a power loss, the contents will not be lost.

- **Mainframe.** Mainframe computers are the large (physically as well as in power) computers used by companies to carry out large volume processing and concentrated computing.
- **Mini.** Minicomputers are physically smaller than mainframes, although the power of many minicomputers exceeds that of recent mainframes.
- **Micro.** Microcomputers are physically small computers with limited processing power and storage. Having said that, the power and capacity of today's micro is equivalent to that of a mainframe only five years ago.
- **LANs.** Local Areas Networks are collections of computers linked together within a comparatively small area.
- **WANs.** Wide Area Networks are collections of computers spread over a large geographical area.

Storage Data is stored in a variety of forms for both permanent and temporary retention:

- **Bits.** Binary Digits, individual ones and zeros
- **Bytes.** Collections of Bits making up individual characters
- **Disks.** Large-capacity storage devices containing anything from 10 Mb to 150 Gb of data
- **Diskettes.** Small-capacity removable disks containing from 360 k to 100 Mb of data
- **Optical Disks.** Laser-encoded disks containing between 650 Mb and 9 GB of data
- **Tapes.** Reel-to-Reel or cassette
- **Memory.** As above

Communications In order to maximize the potential of the effective use of the information on computers it is essential that isolated computers be able to communicate and share data, programs, and hardware devices.

- **Terminals.** Remote devices allowing the input and output to and from the computer of data and programs.

- **Modem.** MOdulator/DEModulator, which translates digital computer signals into analog signals for telephone wires and retranslates them at the other end.
- **Multiplexer.** Combining signals from a variety of devices to maximize utilization of expensive communication lines.
- **Cable.** Metallic cable, usually copper, which can carry the signal between computers. These may come in the form of "twisted pair," where two or more cables are strung together within a plastic sleeve, or in the form of coaxial, where a cable runs within a metallic braiding in the same manner as a television aerial cable.
- **Fiber Optics.** These consist of fine strands of fiberglass or plastic filaments that carry light signals without the need for electrical insulation. They have extremely high capacity and transfer rates but are expensive.
- **Microwave.** This form of communication involves sending high-power signals from a transmitter to a receiver. They work on a direct line-of-sight basis but require no cabling.

Input Inputs to computer systems have developed rapidly over the years. The IS Auditor will still occasionally encounter some of the earlier types:

- **Cards.** Rarely seen nowadays, punch cards were among the first input and output media and consisted of cardboard sheets, some 8 inches by 4 inches with 80 columns, where rectangular holes could be punched in combinations to represent numeric, alphabetic, and special characters.
- **Paper Tape.** Another early input/output medium, paper tape was a low-cost alternative to punch cards and consisted of a 1-inch wide paper tape with circular holes punched to form the same range of characters.
- **Keyboards.** The most common input device today (although that is changing). Most keyboards are still based on the original typist's QWERTY keyboard design.
- **Mouse.** An electromechanical pointing device used for inputting instructions in real time.
- **Scanners.** Optical devices that can scan pictures into a digitized computer-readable form. These devices may be used in combination with OCR (Optical Character Recognition) software to

allow the computer to interpret the pictures of data into actual characters.
- **Bar Codes.** Optically recognizable printing that can be interpreted by low-cost scanners. Common in retail operations.
- **Voice.** Perhaps the future of computer input whereby the computer user, programmer, or auditor simply dictates into a microphone and the computer responds appropriately.

Output As with inputs, outputs are changing rapidly. In the earliest of computing times, output came in three basic forms. The most common of these was paper; however, quantities of cards and paper tape were output for subsequent reprocessing. Nowadays most outputs are via screens or directly onto magnetic media.

- **Paper.** Still a popular output medium, paper may be in continuous stationery form, cut sheet form, or preprinted business stock such as invoices or negotiable instruments such as checks.
- **Computer.** Output directly to another computer is a growing trend with the coming of age of electronic data interchange (EDI).
- **Screen.** Output to screen is the current norm for the majority of outputs with graphics, tables, and charts, and three-dimensional forms possible.
- **Microfilm/fiche.** For permanent, readable recording of outputs with a small storage space required, microfilm is a popular output medium. Each frame contains one page of printed output. An alternative is the creation of microfiche measuring approximately 6 inches by 4 inches and containing some 200 pages of printout.
- **Magnetic Media.** Output to disks, diskettes, and tapes is commonly used to store large volumes of information.
- **Voice.** Another new output medium is voice, where a permanent record is not required.

Control Within the computer systems, control is exercised at a variety of points within the overall architecture. At each stage, opportunities exist to vary the manner in which the computer systems perform to meet the needs of the users.

- **Operating System.** The Operating System is the set of programs that control the basic operations of the computer. All other soft-

ware runs under the direction of the Operating System and rely on its services for all of the work they undertake.

- **Applications.** These systems perform the business functions required of the computer. They run under the direct control of the Operating System but may contain many powerful control elements themselves.
- **Parameters.** These are user-defined variations adjusting the manner in which programs normally operate.
- **Run Instructions.** These are instructions to operators of computers instructing them on the jobs to be run and responses to machine questions to be entered.
- **JCL.** Job Control Language is a means of automating the job-running process by giving the computer the instructions in a form of batch programming language.
- **Human Element.** Ultimately control is exercised by the people who use, operate, program, and manage computers.

People As pointed out in the Criteria of Control (CoCo) report referenced in Chapter 15, control is exercised by people and, as such, the auditor must understand the roles and responsibilities of the individuals involved in the development and processing of computer systems.

- **Operators.** Run the computers on a day-to-day basis.
- **Programmers.** Write the application programs that run on the computer.
- **Systems Designers.** Design the overall structure of the application systems and specify the programs required.
- **Systems Analysts.** Analyze the business structures, applications, and procedures to determine what, if any, contribution IS can make. They will also design the outline business specifications of new systems.
- **Systems Programmers.** Are responsible for the well-being of the Operating Systems and programmers, the related systems software components.
- **Database Analysts.** Are responsible for maintaining the Database Management System (DBMS), which is the systems software controlling access to and format of the data.
- **Network Analysts.** Are responsible for ensuring availability; performance standards and security are achieved on networks.

- **Management.** Plan, organize, and direct to ensure corporate objectives are achieved.

Data Data consists of:

- Fields held in
  - Records held in
    - Files held on
      - Disks

BATCH AND ON-LINE SYSTEMS

Batch versus On-line

In the early days of commercial computing, and up to the late 1960s, most processing took place on a batch basis only. This meant that all inputs were collected centrally and input together in "batches" of documents. This would typically take place using a centralized data preparation function to convert the data from written form into holes punched into either cards or continuous paper tape. The process was highly error prone and the input medium was fragile. In later batch systems the data was entered via a terminal onto a file, which would later be processed in batch mode. In this type of system, the primary control objectives were the ***accuracy*** and ***completeness*** of capture.

Many highly effective controls were designed and implemented to ensure completeness of data capture of batches of data, complete capture of all batches, and accurate capturing of batches of input data. These controls included the manual preparation of batch header documents for later comparison to computer-generated information, and double keystroke verification, whereby an operator entered the data into a batch of cards or directly onto a file containing a batch of input transactions. This data was then re-input by an independent data capture clerk and system-compared to ensure accuracy and completeness.

With the advent of on-line systems such controls fell away because they were deemed to be no longer appropriate. In many cases within an on-line environment very few alternative controls were implemented and frequently the auditor would find that large

assumptions were made as to the adequacy of the controls surrounding the accuracy and completeness of data input.

In today's systems, capture and processing will normally take place using on-line, real-time data capture with a small batch component. Input is typically via a terminal with instantaneous update. Overnight report production in batch mode is common. The terminals may be local or remote and the remote terminals may be either dial-up or dedicated. The terminals themselves may be of differing types but the principal control objectives remain:

- Availability
- Security
- Confidentiality
- Accuracy

In on-line systems there is an additional component to the system that comes complete with its own concerns and that is the ***Communications*** component. This may take the forms of microwave links, satellite hook-ups, or the more basic cables, which themselves may be either dedicated or dial-up.

Computers communicate in a digital form where a signal is either on or off, whereas normal telephone cables operate in an analog mode where the signal is moderated either by changing the height of the curve (amplitude modulation or AM) or by changing the frequency of the signal (frequency modulation or FM). Communications may operate in a ***Simplex*** mode where traffic is one way only. This means effectively that a circuit must make a complete circle to get there and get a reply back. This form of circuit is inexpensive but vulnerable. ***Half-duplex*** communications allows two-way traffic, but only one way at a time. This is the type of signal used in CB radio. ***Duplex*** communications involves simultaneous two-way communication. Computer systems typically use half-duplex communications.

Other communication concepts that will be of interest to the auditor are:

- ***Synchronous communications.*** High-speed transmission and reception of long groups of characters
- ***Asynchronous communications.*** Slow, irregular transmissions, one character at a time with start and stop bits

- ***Encryption.*** Scrambling of data into unreadable forms such that it can be unscrambled
- ***Protocol.*** A set of rules for message transmission in the network

Networks themselves may be of varying types including Private Networks; Public Switched Networks (PSNs), such as the telephone systems; Value Added Networks (VANs), such as Beltel, where the service provider adds on additional services to simply providing point-to-point connection; and Local Area Networks (LANs), where the connections are both private and nearby. Where there is a significant physical distance involved the network may be referred to as a Wide Area Network (WAN). In recent years, the Internet has become of increasing concern as well as use to the Internal Auditor. The Internet is a collection of computers worldwide connected together loosely and provides both a source of information as well as a source of external risk.

Networks may be configured as point-to-point with separate direct links. An alternative configuration could be a multidrop configuration with multiple terminals sharing a single line. Ring Networks have no central computer; each machine is classed as a "node" on the network, and Star Networks have a single, central computer coordinating all communications.

Where an on-line system exists, there may be capabilities for:

- **On-line inquiry**, which allows a remote user to retrieve data directly. In this case the primary concern should be ***Confidentiality*** of information.
- **On-line data entry** permits remote entry of data and allows concurrent processing of data. In this case the primary concerns would be ***Transaction Authenticity*** as well as ***Accuracy*** and ***Completeness***.
- **On-line update** is similar to on-line data entry but with immediate effect of transactions. The primary concerns here would be ***Concurrency Control*** (prevention of two users updating the same record at the same time) and ***Availability***.

The basic on-line concerns are:

- Availability
- Security

- Unauthorized access
- Accidental or intentional changes

Security-threatened areas would include the operating system and particularly its management features, intercomputer communication including dial-up access and gateways, as well as poor network performance.

In any networked operation, availability is a major concern. This includes availability of the hardware components, the software, the data, the networking capability, and the human resources.

Typical controls in this area to protect against unavailability are the ensuring of:

- An adequate physical environment
- Adequate backups
- Multiple redundancies in equipment to ensure no reliance on a single piece
- Peer-to-peer networking to permit mutual back-up
- Adequate disaster recovery planning
- Training of the appropriate sort

Security itself is a factor of the hardware, the software, and the human element. ***Hardware*** is liable to theft, sabotage, and penetration. On the ***software*** side, the operating system software may itself be stolen, corrupted, or bypassed, while applications software may suffer a similar fate and may additionally be substituted by an alternative application.

Data is one of the organization's most valuable assets and may be liable to theft, corruption, substitution, or manipulation.

Such security threats may come from normal users of the systems, deliberately or accidentally, specialist insiders such as the IT staff, legitimate outsiders such as computer engineers, or even customers and suppliers who have been granted access to the site, or outside hackers who attempt to penetrate the organization's security for fun or profit.

Database Management Systems

A Database Management System is a software or hardware structure controlling the nature of, and access to the information required by a

user application system. Given the manner in which the systems developed over the years, it helps to have a clear understanding of what each component is.

Definition of Terms

Access Methods: Software logic procedures used to retrieve, insert, modify, and delete data on a storage device.

Data Dictionary/Data Directory Systems (DD/DS): The software that manages a repository of information about data and the database environment.

Data Independence and Data Sharing: Data independence is a technique allowing diverse users with different logical views to access the same data in different ways. This is achieved by divorcing the definition of the nature and location of the data from the programs using it. The definitions, views, access rules, locations, logical views, and other information describing the actual data are located in one file of *Metadata*, or data about the data. This enables new users with new logical views to be accommodated as well as changing logical views and changing physical representation.

Data Structure: The interrelationships of data.

Database: A collection of data logically organized to meet the information requirements of a universe of users.

Database Administration: A human function involved in the coordination and control of data-related activities.

Database Management System (DBMS): A hardware/software system that manages data by providing organization, access, and control functions.

Storage Structures: Methods and techniques used to physically represent data structures on storage devices.

User System Interfaces: Components of the database environment that request, manipulate, and transform data into information for an end user.

Conceptual Level of Database Design Individual Database Management Systems vary widely in data structuring capabilities. Selection among these will depend on both the ***Entry Access Methods*** (Randomizing,

Indexing) and the ***Navigational Access Methods*** (Read the First, Read the Next, Embedded Links, Inverted Index).

Principals of Data Structures Data Structures are used to model a business (function) in terms of information and follow the general business structure, namely:

- Sequential
- Hierarchical
- Network
- Relational Model

Data Structures then become the basis for Database Type selection:

- Sequential
- Hierarchical
- Network
- Relational Model

All of these database types have generic components although each component is different for each branded product:

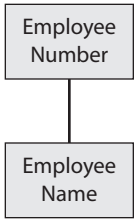
- Data Definition Language (DDL)
- Storage Structure Definition Language (SSDL)
- Data Manipulation Language (DML)
- DBMS Nucleus and Utilities

Database Structuring Approaches Over the years the form in which we have looked at data has evolved from the original *sequential approach* to today's *relational approach*. The auditor may still find examples of all such database approaches in the course of auditing.

Sequential Approach

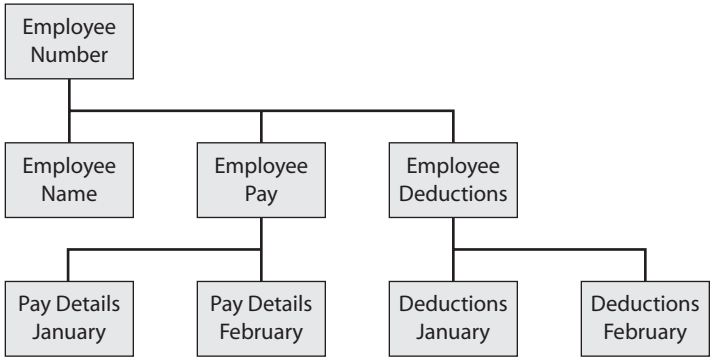
- Fundamental Assumption

  There is a Direct Relationship between data:



Hierarchical Approach

- Fundamental Assumption
  There is some Hierarchical Relationship between data

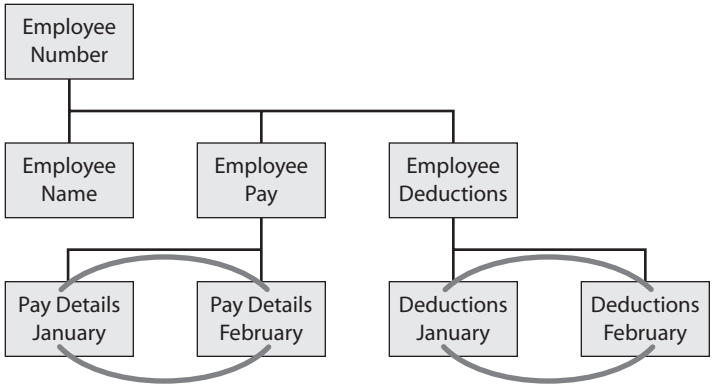


Terminology

- Root Segment
- Parent Segment
- Child Segment
- Twins

Network Approach

- Fundamental Assumption
  There is some General Relationship between data:



Terminology

- Records
- Pointers

Note

- Any structure may be defined
- Records may contain multiple fields

Relational Model

- Fundamental Assumption

  There is some Mathematical Relationship between data:

Emp No	Dept No	Name
12	15	F Bloggs
25	43	J Smith

Employee Table

Dept No	Dept Name
43	Internal Audit
47	IT

Department Table

Data Manipulation

SELECT	— All Retrieval
UPDATE	— Change
INSERT	— Create new Tuple
DELETE	— Delete Tuple
FROM	— Specifies Table
WHERE	— Conditions
AND	— Conjunction of Conditions
OR	— Disjunction of Conditions

Example

```
SELECT    — EMPLOYEE = NAME FROM EMPLOYEE-DB WHERE
          —DEPT = "B03" AND POSITION = "MANAGER"
```

The result is always a table:

Packages and Vendors

DB2	—IBM
DATACOM	—ADR

Inverted List

Record	Make	Color	Model
1	BMW	Red	528I
2	Ford	Blue	Laser
3	Ford	Red	Laser
4	BMW	Blue	328i

Can be indexed by Make, Color, Model

Make	Records
BMW	1,4
Ford	2,3

Color	Records
Red	1,3
Blue	2,4

Model	Records
328I	4
528I	1
Laser	2,3

Terminology

- Indexes and Pointers

Data Dictionary/Directory Systems The *Data Dictionary* tells "what" is in a database/file. It deals with the description of the logical view (that is, the partial view of the data that a user has and includes such items as the data Name, Description, Synonym, etc.).

The *Data Directory* tells "where and how to access" data. It deals with the description of the physical aspects of data such as Location, Address, and Physical Representation.

Entities of the DD/DS are defined by *Attributes*, which describe data's:

- Identification
- Source
- Classification
- Usage
- Qualification
- Relationship

A DD/DS can be a useful tool independent of the need for a DBMS in the areas of documentation support, coordination of shared data usage, and control over modification of programs and files. The DD/DS has become popular with the advent of DBMS packages with the greater recognition of opportunities to share data, leading to a greater need to control data usage, the future introduction of computer privacy legislation, as well as the complexity of relationships involved.

Who Looks After the Database System?

Database Administrator

Functions of the DBA include coordinating the information content of the database. This does not mean that the actual data itself is the DBA's concern, but rather that the DBA is responsible for deciding the storage structure and access strategy. The DBA will liaise with computer users and, following their business requirements, will define authorization checks and validation procedures as well as a strategy for back-up and recovery. The DBA is also responsible for monitoring performance and responding to changes in requirements.

In order to achieve all these objectives DBAs have at their disposal tools specifically designed to facilitate these tasks. These tools include *utility programs* to permit the loading of raw data onto the database, *reorganization routines* to keep the database access efficient as well as effective, and *statistical analysis* to determine when maintenance is required. In addition *journaling* (e.g., logs) can keep records of who did what and when on the database. Although these logs are optional, they can be of major assistance to the DBA in database recovery in the event of problems. The *Data Dictionary* itself, in addition to *database analyzers*, will also assist in recovery.

Database Recovery The objective of database recovery is to reinstate databases to a known state while minimizing lost work. This means that it must permit recovery on a transaction basis and provide fast recovery in order to minimize manual work while ensuring the safety of recovery data. At the same time, it is inevitable in such a process that some data will eventually go missing and recovery must provide a mechanism to inform users of "lost" transactions. Recovery must cater for various types of failures including hardware as well as software disasters.

Recovery procedures come in various forms to cater for the various forms of failure and include:

- **Checkpoints.** The DBMS will either determine or force a point in time where all transactions have been effected, committed to disk, and all memory buffers have been flushed. At this point a record is made that the database is "quiet" or has been "quiessed" and that any recovery only needs to go back this far. It is used to define a "stable" state of the database for recovery.

- **Roll Back.** The log is processed backward and completed transactions are rolled out to the last checkpoint.
- **Roll Forward.** Log is processed forward to reinstate the system.
- **Update Back-up copy.** (e.g., Media Failure)
- **Compensating Transactions.** (e.g., Journal Entries)
- **Salvation Routine.**

For effective log recovery, before the database is updated a log of the *before* image is made to enable undoing the change if necessary (e.g., failure before update). After the database is updated a log of the *after* image is made to redo change if necessary (e.g., media failure). These logs contain audit trail information—for manual follow-up (e.g., Program Id, Transaction Id).

Auditing Databases With the advent of DBMSs, these have been a migration of controls from individual application to the general database environment. This migration of controls improves control opportunities overall and permits the centralized administration of the control environment.

In order to review relevant database designs, the auditor would:

- List all the record types
- Read and analyze their descriptions and names
- Identify the key of each record and verify requirement for uniqueness
- Study the relationships (and sets)
- Identify all the relationships, which are:
  - One:Many
  - Many:Many
- Evaluate the strength of each relationship
- Verify consistency of the design with business information needs

Documentation of the Database Environment The Data Dictionary/Directory system can be used to accept, record, and generate a variety of documentation, including:

- Requirements and specifications
- Data documentation
- Metadata generation

It can automate the documentation process and enable cross-referencing of programs to individual data elements as well as report generation. It can also assist in enforcing *change control*.

The Data Dictionary can be effective in *active* or *passive* mode. In active mode, all database access must be made via the DD/DS. In passive mode the DD/DS is there as a record, but has very little effective control. The obvious advantages of an Active DD/DS include improved accuracy, timeliness, completeness, and control over updates.

Administration and Coordination Functions The auditor would also review the Administration and Coordination Functions by examining the review and monitoring activities of the DBA, which would include reviews of database designs, reviews of system design, reviews of program design, and coding, monitoring the general quality of data, and monitoring overall database performance.

Organizational issues such as the roles that require segregation would also be examined. These would typically include:

- Database Administration
- Systems Development
- Programming
- Operations
- End Users
- Internal Audit

This can only be effectively done by assigning responsibility for data ownership. This can mitigate the control concern of uncoordinated sharing of data update responsibility by assigning definitional responsibility. Such coordinating of shared usage ensures the uniform application of appropriate level of controls.

Operational Controls for the Database Environment In the operation of a database-structured business system the auditor must ensure the existence and effectiveness of controls for ensuring against unauthorized access, controls over ensuring accuracy and completeness, Recovery and Restart tools and techniques, and controls over access to data.

The impact of a database environment on privacy and security is complicated by the need for the assessing of requirements among mul-

tiple users. Sharing of data may cause control concerns; however, it is possible to describe security specifications using the Declarative Data Definition Language (DDL), thus centrally ensuring clearer specifications and making the environment easier to audit. This is brought about by the migration of the implementation of controls from the application to the environment.

Impact of Database on Completeness and Accuracy Issues Database technology can have a marked effect on the quality of information provided. This is a concentration of risk due to sharing of data and an increased cost of error correction due to the system's complexity. In addition, there can be a deteriorating effect on user reliance and confidence due to database erosion and cascading errors.

Mitigating the concerns involving completeness and accuracy means that not only the risks but also the controls must migrate. These may be generalized in the environment and implemented in the DBMS, DD/DS, and so on.

The benefits from the auditor's viewpoint include the potential for:

- Consistency of data
- Enhance quality of audit by increased accessibility
- More accurate systems development process
- Data resource management will accrue benefits through formalized discipline
- Migration of controls

Disadvantages from the auditor's viewpoint include:

- New Technology/Pioneer syndrome (technology lust)
- Implementation cost control
- Access to DBMS managed data
- Data Integrity trade-offs
- Change in scope/timing of audit

The role of the auditor in such a changed environment is to consult with the database user on requirements, check the edit and validation rules, determine whether there is partial acceptance or rejection on error, and who has the responsibility for correctness and to consult with DBA on the implementation plan.

Qualities that assist the auditor in these tasks include the capability for edit and validation element by element, the error response, the procedures for edit and validation maintenance, and the procedures for adding new data elements.

Common DBMS edit and validation controls include:

- Uniqueness checking for Key and Nonkey
- Structural/Relational checking
- Picture string and simple format checking
- Edit and validation performed according to declaratively specified criteria

Metadata generation can provide the desired control environment.

Controlling Initial Database Content Usually the user is responsible for providing initial database content and should spot-check loaded database for correctness. The auditor will determine whether there has been sufficient checking using statistical methods when appropriate.