NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF BUSINESS AND HUMAN RESOURCE

COURSE CODE: BHM736

COURSE TITLE: E-Records Management

Course Code BHM736

Course Title E-Records Management

Course Team Gerald C. Okereke (Course Developer/Writer)

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island

Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published By:
National Open University of Nigeria

First Printed 2011

ISBN: 978-058-249-5

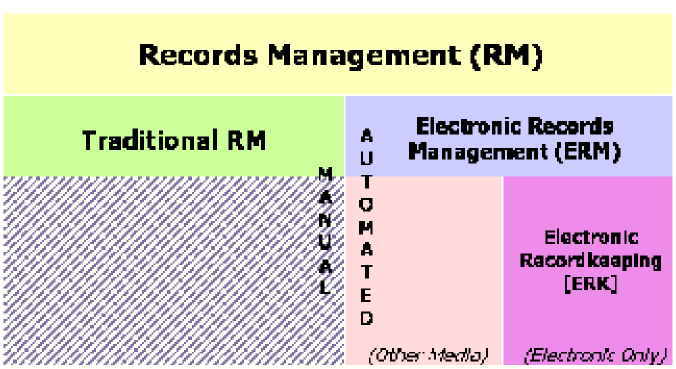
CONTENTS PAGE

MODULE 1

UNIT 1 INTRODUCTION TO E-RECORDS

CONTENTS

1.0 INTRODUCTION



The figure above is meant to place both the definitions that follow and the subsequent products in an appropriate context. A discussion of this follows.

As denoted in the above figure, records management is the rubric "umbrella term" for the discipline defined by an agency's policies and processes for managing the records that document its activities. This is denoted in the above figure as the top-level, overarching box.

Records management activities, subsumed under that umbrella, may be accomplished in either a manual or automated manner. This is denoted in the bifurcation at the next level down in the diagram. When records management activities are manual, this is the traditional approach to records management. When records management activities are automated, they are referred to as Electronic Records Management [ERM]. It is important to note that the electronic in ERM refers to automation, not to the nature of the record media. It is entirely possible to have manual processes and procedures to handle electronic records (e.g., shelf lists of boxes of floppy disks). Electronic records management simply implies automation of records management processes and procedures.

These first five Fast Track products focus specifically on automated processes and procedures used in the management of only those records maintained in electronic form; Fast Track has termed this arena as that of electronic record keeping [ERK]. This is denoted in the darkest box at the lower right of the figure above.

One of the first decisions that agencies must make when deciding to move towards automation of records management processes is to determine whether the system is to manage only paper records [TRM], only electronic records [ERK], or both [RM].

This unit provides a context for ERK, specifies the records management and information technology [IT] terminology associated with ERK.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define e-records management terminology
- define some information management terms related to records management
- explain how to retrieve records and documents
- discover record keeping responsibilities on a sheet of paper
- describe how to create taxonomies
- identify strategies to put in place to secure a record management system.

3.0 MAIN CONTENT

3.1 Context: Terminology

Shared terminology, scope, and framework are fundamental to our understanding of electronic records issues. This is especially true when more than one profession, each with its own vocabulary and point of view, must work collectively to solve the challenges electronic records present.

- Records

Records include all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organisation, functions, policies, decisions, procedures, operations or other activities of the Government or because of the informational value of the data in them.

Information technology uses the word record to describe a particular set of information. In this data processing context, records are composed of fields of information, and a collection of such fielded records comprises an electronic file.

- Electronic Records

Electronic, or machine-readable records, are records on electronic storage media (A Glossary for Archivists, Manuscript Curators, and *Records Managers, Society of American Archivists: Chicago, 1992 p.* 12). Electronic record, as defined in NARA regulations (36 CFR 1234.2), means any information that is recorded in a form that only a computer can process and that satisfies the definition of a Federal record per the Federal Records Act definition supplied above. Federal electronic records are not necessarily kept in a “record-keeping system” but may reside in a generic electronic information system or are produced by an application such as word processing or electronic mail.

- Metadata

Metadata is a term that describes or specifies characteristics that need to be known about data in order to build information resources such as electronic record-keeping systems and support records creators and users.

- Records Management

There are many (similar) definitions of records management. One common one is "...the field of management responsible for the systematic control of the creation, maintenance, use and disposition of records." (A Glossary for Archivists, Manuscript Curators and Records *Managers, Society of American Archivists: Chicago, 1992 p. 29.) From* the Federal perspective, it is the planning, controlling, directing, organising, training, promoting, and other managerial activities involved in records creation, maintenance and use, and disposition in order to achieve adequate and proper documentation of the policies and transactions of the Federal Government (36 CFR 1220.14).

- Electronic Records Management [ERM]

Electronic records management [ERM] is using automated techniques to manage records regardless of format. Electronic records management is the broadest term that refers to electronically managing records on varied formats, be they electronic, paper, microform, etc. Electronic record-keeping [ERK] is a subset of ERM, because ERK focuses on electronically managing electronic records.

- Electronic Record-keeping [ERK]

Electronic record-keeping [ERK] is the development of automated processes an agency uses to manage its electronic records. These automated processes support not only the preservation of an electronic record's content, but also its context and structure over time. These first Fast Track documents specifically address ERK issues.

- Electronic Information System (EIS)

"The collection of technical and human resources that provide the storage, computing, distribution, and communication for the information required by all or some part of an enterprise. A special form of information system is a Management Information System (MIS), which provides information for managing an enterprise."

Electronic information systems automate certain business functions. Other programmatic electronic information systems may automate agency mission-specific business functions, and as such, may produce Federal records in the process. These electronic information systems may or may not incorporate all aspects of appropriate electronic record-keeping, depending on their design characteristics.

- Electronic Record-keeping System [ERKS]

An electronic record-keeping system [ERKS] is an electronic information system that meets an agency's record-keeping needs. At a high level, NARA has defined an ERKS as an electronic information system in which records are collected, organised, and categorised to facilitate their preservation, retrieval, use, and disposition. From a records perspective, an ERKS will ensure that the records it maintains will have sufficient authenticity and reliability to meet all of the agency's record-keeping needs.

The International Council on Archives document Guide for Managing *Electronic Records from an Archival Perspective states "The reliability* of a record is its ability to serve as reliable evidence....Authenticity refers to the persistence over time of the original characteristics of the record with respect to context, structure and content. An authentic record is one that retains its original reliability."

- Electronic Document Management System [EDMS]

An electronic document management system [EDMS] is software that manages the creation, storage, and control of semi-structured documents. It consists of several technologies including, but not limited to document management, COLD (Computer Output to Laser Disk), imaging, and workflow. See Preliminary Planning for Electronic Record-keeping: Checklist for RM Staff for a more detailed description of these technologies.

In part, because an EDMS does not support the preservation of the business context of an individual record, (i.e., EDMS systems manage a content item as an individual unit, as opposed to preserving its relationship to a larger group of documents that provide evidence of the same particular organisational function), EDMS systems are not electronic record-keeping systems.

- Records Management Application [RMA]

Records Management Application [RMA] is the term used for software that manages records. Its primary management functions are categorising and locating records and identifying records that are due for disposition. RMA software also stores, retrieves, and disposes of the electronic records that are maintained in its repository.

3.2 Information Management Terms Related to Records Management

There is considerable confusion in the marketplace regarding the definition of various information management terms. The scope and role of specific information systems is particularly blurry, in part caused by the lack of consensus between vendors.

Content Management System (CMS)

Content management systems support the creation, management, distribution, publishing, and discovery of corporate information. Also known as 'web content management' (WCM), these systems typically focus on online content targeted at either a corporate website or intranet.

Enterprise Content Management System (ECMS)

An enterprise content management system consists of a core web content management system, with additional capabilities to manage a broader range of organisational information. This often consists of document management, records management, digital asset management or collaboration features.

Document ManagementSystem (DMS)

Document management systems are designed to assist organisations to manage the creation and flow of documents through the provision of a centralised repository, and workflow that encapsulates business rules and metadata. The focus of a DMS is primarily on the storage and retrieval of self-contained electronic resources, in their native (original) format.

Records Management System (RMS)

The Australian Standard on Records Management defines record-keeping systems as 'information systems which capture, maintain and provide access to records over time'. This includes managing both physical (paper) records and electronic documents.

Digital Asset Management (DAM) System

Digital Asset Management (DAM) systems support the storage, retrieval and reuse of digital objects within an organisation. DAM differs from document management and content management in its focus on multimedia resources, such as images, video and audio. DAM also typically provides rights management capabilities.

Brand Management System

Brand management systems are specific applications of the more general DAM category of products to the management of advertising and promotional materials.

Library Management System (LMS)

Library management systems provide a complete solution for the administration of all library's technical functions and services to the public. This ranges from tracking the assets held by the library, managing lending, through to supporting the daily administrative activities of the library.

Digital Imaging System

Digital imaging systems automate the creation of electronic versions of paper documents (such as PDFs or TIFFs) and are used as an input to records management systems. By creating electronic resources, they can be manipulated directly by the records system, eliminating the need for physical filing.

Learning Management System (LMS)

Learning management systems automate the administration of training and other learning. This includes registering students, managing training resources, recording results, and general course administration. Learning management systems are designed to meet the entire needs of professional trainers and other educators.

Learning Content Management System (LCMS)

Learning content management systems combine the capabilities of a content management system (CMS) with that of a learning management system (LMS). This allows them to manage both the content of the training materials, and the administration of the course itself.

Geographic Information System (GIS)

Geographic information systems (GIS) are special purpose, computer-based systems for the capture, storage, retrieval, analysis and display of spatial (location-referenced) data.

SELF ASSESSMENT EXERCISE

Define brand management system.

3.3 How to Retrieve Records and Documents

Earlier we find that in rolling out a records management system, three critical success factors were identified: the software, classification scheme and message.

This sub unit focuses on a specific aspect of document management, records management and intranet projects.

How to look for records and documents

There are many situations which can make a staff in an office to look for documents. It is these kinds of situations that inform how suitable software's are designed and developed to make classification schemes easy.

Four ways to look for records and documents

Observation of typical working environments has identified that there (at least) four different situations in which staff members look for documents:

- known-item searching
- unknown-item searching
- own documents
- other people's documents

Note that these four situations are not mutually-exclusive, but they overlap in practice.

- Known-item searching. The staff member is looking for a specific document that he has seen before, or knows exists. For example, he might know the title of the document, or the 'document number'.

  Document management systems often offer an interface for directly entering a document number, which may be relevant for these types of searches. Alternatively, these documents may also been recently-accessed by the staff member, and therefore on a 'recent', or 'favourites' list.

- Unknown-item searching. In this situation, the exact details of the document are not known, beyond an expectation that a relevant document does exist. For example, the staff member might be looking for “precedents relating to the mining industry”.

  A search interface may be used to find these documents (using either a ‘simple’ or ‘advanced’ search). Users may also browse through the classification scheme to identify relevant topics.

- Own documents. Staff is often working intensively on a small number of documents that they ‘own’. For example, in a law firm, staff will be assigned to a number of 'client matters’.

  In this situation, the staff member is very familiar with the area and the documents within it. Features such as 'work areas', 'most used' or 'recently accessed' lists might be used when looking for these documents.

- Other people's documents. Finding documents written by other staff is much more difficult (particularly in a large, geographically-dispersed organisation). The specific staff managing the documents may be known (or not).

  Search strategies may involve looking up the documents owned by a staff member (or team), or conducting an advanced search.

Putting this into practice

Document management and records management systems are complex and flexible pieces of software, offering a wide variety of tools for finding documents.

Each of these searching and browsing tools will be more (or less) useful in each of the four situations outlined earlier. The challenge therefore is to provide staff with a suitable mix of tools to ensure that documents can be found, without overloading users with unnecessary complexity within the system.

Conducting user research

Implementation teams should conduct practical user research to identify how staff uses existing systems in each of the four situations identified. Prototypes or pilot installations of the new system can be used to explore how staff could use this.

Techniques such as staff interviews, workplace observation and contextual inquiry are particularly effective at obtaining this information. The systems and classification schemes can then be customised or configured in the light of the research results.

3.4 Record-keeping Responsibilities on a Single Sheet of Paper

With the move from paper to electronic documents, responsibility for record-keeping within organisations has shifted to individual staff and away from centralised records management specialists. Much is made of the need for all staff to understand their record-keeping responsibilities. To this end, many training and communication programs are conducted within government agencies (and elsewhere). To a large extent, this training has failed. While staff gains a general awareness of record-keeping, they are not provided with sufficiently concrete and detailed guidance to make their record-keeping successful and consistent.

This sub unit explores ways to help staff meet their record-keeping obligations by creating a single sheet of paper for each staff member with everything that they need to know.

Traditional record-keeping training

Most organisations have fairly well-established staff training programs on record-keeping, covering topics such as:

- what is a record
- why records need to be kept
- record-keeping obligations of all staff
- how to file records in corporate record keeping systems
- when and how to dispose of records

Crucially, this training only talks of records in general terms, outlining statements such as ‘records are any documents that provide evidence of a decision or activity’.

In practice, not every document or email should be kept, and these general statements do little to help staff make judgements about what to file.

The training also fails to tell staff where to file individual records, other than generally pointing to the corporate records systems.

Day to day record -keeping challenges

Record- keeping obligations will only be met when staff actually files everything that they should in the records management systems. Day to day, however, staff are left to make value judgements about when and how this should be done. Do emails get filed, and if so, which ones? Does internal correspondence get filed? Where does staff hiring records go?

Furthermore, staff members are not the same as each other. The specific documents that need to be filed by an HR team member vary greatly from those that are important for call centre or admin staff.
A single sheet

Staff members need to know exactly what they are supposed to do and this information should be provided to them in the simplest possible way. To this end, each staff member should be provided with a tailored, personalised single sheet of paper that covers:

- the top six (or a dozen) types of documents that they need to file
- the specific folders each of these documents is filed in
- the documents they don’t have to file
- guidance on common issues or challenges

In addition to listing specific documents to be filed, the guidance should be equally specific, on common issues (such as whether broadcast emails need to be filed by recipients). To highlight this is not a general training document, but an individualised ‘cheat sheet’ for each staff member. If it cannot be distilled down to this simple level, then staff members need to find the answers before they can expect other staff members to meet their obligations.

This also forces record -keeping staff to roll out record keeping approaches incrementally, identifying the key documents that have to be kept, and those that can be safely disposed of. While very simple, the ‘single sheet’ approach is quite different to traditional record-keeping training, and much more effective.

3.5 Creating Taxonomies

Taxonomies define the structure that underpins document and records management systems, knowledge management projects and more.

Considerable effort goes into developing these taxonomies, with the goal of creating a common structure that will benefit the whole organisation.

The challenge, however, is to ensure that these taxonomies work well for staff, beyond any organisational benefits that are sought. It is here that taxonomies often fail.

If not designed well, taxonomies can become 'white elephants', too hard to understand and too complex to use. At their worst, poorly designed taxonomies are the direct causes of project and system failure.

The field of Information Architecture (IA) has much to offer those creating taxonomies, including a range of structured techniques for building and testing their effectiveness.
This briefing outlines some of these approaches, and encourages creators of taxonomies to retain a clear focus on usability throughout the design process.

Building taxonomies

Taxonomies are typically drawn from a number of sources, including existing industry-wide classification schemes, business functions and structures already in place within sections of the organisation.

These are pulled together to create a larger or more complete taxonomy.
Testing of this taxonomy usually relies on internal review, discussing the taxonomy with staff, and gaining input on areas of strength and weakness. While effective for gaining broad user and stakeholder input, this kind of review is very shallow, and is not sufficient to ensure that the taxonomy can be used in practice. Instead, structured techniques must be used, getting beyond staff and expert opinions.

Three key purposes of taxonomy

There are three clear purposes of taxonomy:

- knowing where to file information correctly
- retrieving information easily when needed
- meeting legislative, compliance or business objectives

It is easy to create a taxonomy that meets the third goal, at least superficially. If the first two goals are not achieved, however, the taxonomy will certainly fail.

Information architecture

Information Architecture (IA) is a discipline that focuses on creating effective structures and navigation for websites and intranets. Through this work, the field of IA has built up a toolbox of techniques that can be applied equally well to taxonomy creation. Three of these techniques are outlined below, with pointers to further information.

Card -sorting

This is a very simple technique for building an understanding of how staff members think about information used as an early input when creating taxonomy. More on this technique are the following.

Card-based classification evaluation

It provides a rapid way of testing taxonomy to ensure that staff can correctly store information and find it again later.

Usability testing

This is designed to test the overall ease of use and effectiveness of not just the taxonomy, but the system used to implement it. It should be used throughout the design and implementation process.

Best-practice approaches

It is no longer sufficient to simply gather staff input to assess the effectiveness of taxonomies. Instead, practical IA techniques should be used to ensure that taxonomy works in practice.

3.6 Providing-Intranet Access to Records

Many organisations are attempting to clarify the relationship between the corporate intranet, and their document/records management system. While this is a broader issue of information management with an organisation, there are some short-term activities that can be taken to create a working relationship between these two platforms.

This briefing outlines a simple scenario in which the intranet helps staff find key corporate information, while the documents accessed are stored in the document/records management system.

Usage scenario

The following scenario outlines one of the typical ways an intranet is used:

- A staff member browses into the HR section, and then to the 'HR *policies' page. Scanning through the list of documents available,* *the leave policy is selected, and the PDF opens up in Acrobat Reader.*

This is a very common scenario, but more could be happening behind the scenes than is apparent to the end user:

- The list of policies in the HR section is actually dynamically created, generated by asking the records management system to return a list of all the documents stored in a particular folder.
- The list returned is then formatted, and presented as a simple list of document titles (as links), along with supporting descriptions and other details.
- Clicking on a document link invisibly requests the file from the records management system, which is then automatically opened up in the user's copy of Acrobat Reader.

In this way, the intranet provides a simple mechanism for browsing for required information, while the files themselves are stored in the records management system. What is happening 'under the hood' is then invisible to the end user.

Benefits of this approach

Implementing a solution as outlined in the scenario provides a number of direct benefits:

- The intranet provides a simple and user-friendly way of finding required information, without having to expose staff to the complexities of the records management system.
- Information is accessed from a single location, regardless of whether it is provided as a web page, or as a file from the records management system.
- Storing the documents themselves in the records management system provides all the benefits relating to better document management, security, versioning, and record-keeping compliance.
- Providing simpler access to information via the intranet demonstrates the value of the records management system, by allowing information to be easily retrieved from the system, and not just stored for record-keeping reasons.

Taking this approach therefore delivers a ‘win-win’ outcome that benefits the intranet, the records management system, as well as the end user.

Making it happen

Implementing this solution can be quite straightforward. Almost all of the document/records management systems provide a well-developed (and well documented) API for accessing files within the repository.

A day (or two) of programming, in a language of your choice, is generally sufficient to create the requests to return a list of documents within the records management system.

As long as the records management system is then configured (and licensed) to return the documents directly to the user for display in Acrobat Reader, the job is then done.

Better yet, a small (but increasing) number of content management systems provide this capability built-in, without requiring additional development or customisation.

4.0 CONCLUSION

Electronics management systems have gone a long way to facilitate the search, storage, and retrieval of documents and records. Several definitions, related concepts and applications of e-records are all target at the same goal of processing data and making information available at the right time. The advantages are very obvious especially when compared to the hitherto manual systems. Though there are rooms for improvement.

5.0 SUMMARY

- One of the first decisions that agencies must make when deciding to move towards automation of records management processes is to determine whether the system is to manage only paper records [TRM], only electronic records [ERK], or both [RM].
- Shared terminology, scope, and framework are fundamental to our understanding of electronic records issues. This is especially true when more than one profession, each with its own vocabulary and point of view, must work collectively to solve the challenges electronic records present.
- There is considerable confusion in the marketplace regarding the definition of various information management terms. The scope and role of specific information systems is particularly blurry, in part caused by the lack of consensus between vendors.
- Earlier we find that in rolling out a records management system, three critical success factors were identified: the software, classification scheme and message.

- With the move from paper to electronic documents, responsibility for record-keeping within organisations has shifted to individual staff and away from centralised records management specialists
- Taxonomies define the structure that underpins document and records management systems, knowledge management projects and more.
- Many organisations are attempting to clarify the relationship between the corporate intranet, and their document/record management system.

6.0 TUTOR-MARKED ASSIGNMENT

1. List ten information systems related terms and concepts
2. Briefly discuss how to create taxonomies in e-records management

7.0 REFERENCES/FURTHER READING

Bruce, Miller (2004). “Electronic Record Keeping: The State of the Art, and How to make it Work”, IBM Software Group, Ottawa Software Lab.

Robertson, J. (2004). Definition of Information Management Terms, Step II Design. Australia.

Robertson, J. (2007). Applying IA Technology When Creating Taxonomies, Step II Design. Australia.

Robertson, J. (2008). Record Keeping Responsibilities, Step II Design. Australia.

Robertson, J. (2004). Evaluating Caloundria City Councils EDMS Classification, Step II Design. Australia.

The National Archives. (2008). Typical Records Management (RM) Functions and T ypical RM Programme.

UNIT 2 ACTIVITIES, FUNCTIONS AND MANAGEMENT

CONTENTS

1.0 INTRODUCTION

An effective records management program is the foundation on which an agency can begin automating records management activities or implementing systems to manage electronic records. Some of the major activities of a typical records management program are described in the following. In conjunction with agency Records Management Officers, these activities are performed on an ongoing basis in order to establish and maintain an effective records management program.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- identify the typical activities associated with e-records management system
- list the functions of e-records management systems
- describe some examples of functions of e-records on display
- reconise the logical steps to take in building an e-records management system
- discover some questions on the security of e-records management system.

3.0 MAIN CONTENT

3.1 Typical RM Program Activities

- Identify Records and Records Sources

Distinguish records from non-records and uniquely identify the records found. Determine how, how many, and by whom records are being created or received and identify their relationship to the agency's business operations or functions.

- Develop File Plan

Specify how records are to be organised once they have been created or received. Identify the classes of records (record series) that an organisation produces, and establish how to associate given records within a class to other records in that same class. For example, put all the travel vouchers in a file marked "travel vouchers" or, alternatively, put all travel vouchers for person X in a file marked "person X travel."

- Develop Records Schedules

Determine where and how long records need to be retained and what their final disposition will be. Records typically go through a period of active use at the time of, and shortly after, their creation. As records are used less frequently, they become inactive and are often transferred to less expensive storage media or locations. All Federal records must be assigned a final disposition, which must be approved by NARA (44 USC 3303). Records schedules call for the disposition of records based on time, or event, or a combination of time and event. Records are either scheduled for destruction or for transfer to the National Archives for permanent preservation. Determining how long records need to be retained - and under what conditions - can dramatically reduce agency resource outlays and at the same time ensure that agencies are maintaining adequate and proper documentation of agency activities.

- Provide Records Management Guidance

Develop and distribute agency-specific policies and procedures for implementing records management activities, including for records created or maintained by contractors for the Federal government. Establish agency-specific record-keeping practices, specifically establishing what records need to be created in order to conduct agency business. Identify parties within the agency with records management responsibilities, such as records officers or liaisons.

3.2 Typical Functions for Managing Records

The management and control of organisational records involves a number of actions taken with respect to individual records or series of records to ensure their authenticity, integrity and reliability, and to preserve their usability over time. The following is a list of some of the major functions most typically performed in the process of managing records. This list of functions may serve as a guide or checklist in defining requirements or specifications for an ERK or ERM system, or in seeking to improve an existing system, to assure that major functional requirements have been addressed.

- Declare a Record: Recognise a record to be a record, identify it as a record.

- Capture Records: Include a record in a system that manages records.

- Maintenance and Use of Records

- **Organise Records: Group records according to a predefined** structure to meet business needs.
- **Maintain Records Security: Protect the integrity of records** against unauthorized alteration or destruction.
- **Manage Records Access: Grant or limit the ability of** individual(s) to examine records or record groupings.
- **Facilitate Records Retrieval: Provide or enable the ability to** collect records relevant to a query.
- **Preserve Records: Ensure the physical state of records so they** remain usable.
- **Audit/Oversight: Ensure compliance of agency record-keeping** practices with existing statutes and internal and external regulations.

- Final Disposition of Records

- **Destruction: Eliminate from a system a group of records in** compliance with the appropriate records retention schedule so that they cannot be accessed, retrieved or recovered.
- **Transfer: Change legal custody of agency records to NARA**

3.3 Examples of Functions

The following checklist gives examples of more detailed system functions that an ERK or ERM might need to perform in order to satisfy each of the basic records management functions. There are a number of ways that each of these records management functions can be automated or implemented in an ERK or ERM, so this checklist is meant only to be a starting-point for agencies to develop system specifications.

Users should note that these specifications are more general than other, more detailed requirements statements like those in the DoD 5015.2-STD. This checklist would apply not only to Records Management Applications (RMAs), which is the scope of DoD 5015.2, but more broadly to any agency initiative to automate records management functions, establish records management control over electronic records, or otherwise improve existing records management systems.

Not all of these requirements will necessarily be implemented as automated "system functions" in an ERK, ERM, or other records management system. Some functional requirements may be implemented through non-automated organisational policies, practices, or records management procedures.

In all requirements that follow, the term "user" refers to authorised users only. Different functions are permitted to different groups of users -- administrative functions to records managers, retrieval functions to end-users, etc.

1. General Requirements

For all records systems (ERM or ERK) the system should allow users to print and view:

- **all system management and control information: file plans,** security assignments, disposition instructions, etc.
- **the records themselves.**

2. Declare a Record

For all records systems (ERM or ERK) the system should:

- **assign unique identifiers to records and their associated metadata**
- **capture as much metadata automatically as possible, and reliably** link metadata to the records
- **some metadata captured for records in ERKs or ERMs might** include: creator, creating organisation, author, recipients, subject matter, format, various dates (date created, date filed), a "vital records" indicator, and others.

3. Capture Records

For all records systems (ERM or ERK) the system should:

- **allow import of records from other sources. This may involve** format conversion for records that are imported from external information systems. In this case records are physically captured and transported to a record-keeping system
- **allow establishing a link from the ERK/ERM system to a record** in an external information system in order to establish records management control. In this case, physical transport of the records from one system to another is not required.

4. Maintenance/Use

Organise Records: For all records systems (ERM or ERK) the system should:

- allow implementation of an agency-specific scheme for how records are organised
- allow users to create an organisation-specific file plan and link the file plan to records retention schedules, including disposition instructions and specification of "cut offs"
- allow users to select categories in which records are filed and assign records to these categories
- allow records to be linked to other records (e.g., a redacted record with its non-redacted counterpart, a final report with its earlier draft versions)
- allow users to create file folders, and to add, edit and delete categories or file folders
- prevent deletion of non-empty folders from the file plan
- allow users to add, edit and delete records retention schedules, and to "freeze" or execute records retention schedules
- execute disposition instructions (e.g., move a group of records from active to inactive status)
- allow users to assign a status to records that prevents their destruction
- allow users to specify to the system the organisational structure, organisational locations, and staff or unit to which records management responsibility may be assigned.
- import information from other sources (e.g., pre-existing file plans, box indices)
- allow users to specify identifiers for boxes, their contents, locations, and related accession information for ERMs, which may also manage paper records.

5. Maintains Records Security: For all records systems (ERM or ERK) the system should:

- prevent over-writing of a record. (Usually, a record 'copy' is checked out of the system and a re-filed record is written as a new record or a new version of an existing record)
- prevent any modification of a record's unique identifier, once it is defined
- prevent deletion of indexes, categories, and other 'pointers' to records
- calculate and maintain a checksum for records and their metadata, or use some similar technological means of detecting any alteration of record or metadata
- provide audit trails of all add, update, deletion, and retrieval activities
- maintain appropriate backup copies of records and record-keeping systems
- provide adequate recovery/rollback procedures and rebuild procedures, so that records may be recovered or restored following a system malfunction.

6. Manage Records Access: For all records systems (ERM or ERK) the system should:

- control access so that only an authorised individual is able to retrieve, view, print, copy, or edit a record or other entities (e.g., metadata, file plan) in the record keeping system
- permit the identification of individual users and groups of users, and enable different access privileges to be assigned to individuals or groups. Access privileges may limit access to selected records or groups of records, and may limit access by selected individuals
- maintain the integrity of redrafted records and assure that redrafted material is not accessible.

7. Facilitate Records Retrieval: For all records systems (ERM or ERK) the system should:

- allow searching on metadata, record content, or assigned subject categories (using a controlled vocabulary)
- ensure that all access privileges (permissions and restrictions) are enforced on all retrievals.
- allow searching based on a combination of metadata, content, and subject categories within a single query. Query results that may

be a list of records and their locations, or may be the records themselves

- allow retrieval of records and associated metadata, and allow retrieval of records based on defined links (e.g., between versions of the same record)
- provide a sufficiently powerful range of search features and options, as needed to meet various agency requirements. These might include: wild-card or exact-match searching, proximity or adjacency searching, relevance ranking of search results, use of stop-words, limits on maximum size of results set from a search, query by image content, or others.

8. Preserve Records: For all records systems (ERM or ERK) the system should:

- ensure that all records can be read and accurately interpreted throughout their useful life in that system
- enable migration of records to new storage media or formats in order to avoid loss due to media decay or technology obsolescence
- ensure that all captured metadata remains linked to appropriate records and is unchanged throughout the useful life of the records, including after migration to new media or technology
- monitor storage capacity and utilisation and alert system operators when action is needed (e.g., to increase capacity, backup system files, etc.).

9. Audit/Oversight: For all records systems (ERM or ERK) the system should:

- create and maintain an audit trail (also called use-history metadata) for all records activity and system functions
- provide access to audit trail information at the fully detailed level (e.g., each individual record access, including record identifier, time, date, and user)
- provide summary reports of audit trail information (e.g., number of accesses)
- track failed attempts of all records activity and system functions
- treat audit trail information (e.g., number of accesses, details of individual record retrievals, attempts to delete a record, etc.) so that it can be managed as a record.

10. Final Disposition of Records

Destruction: For all records systems (ERM or ERK) the system should:

- identify records eligible to be destroyed, based on records retention schedules and disposition instructions
- delete records in a manner that they cannot be physically reconstructed or otherwise retrieved
- enable a record to be kept of all record destructions, providing certifiable proof of destruction.

Transfer: For all records systems (ERM or ERK) the system should:

- identify records eligible to be transferred, based on records retention schedules and disposition instructions
- export records and metadata (i.e., copy and subsequently remove them from the system), in a format acceptable for transfer to NARA
- enable a record to be kept of all record transfers, providing certifiable proof of transfer.

3.4 Record-Keeping Responsibilities on a Single Sheet of Paper

1. Build the Organisational Structure Required to Support E-Records

- Link IT to RM
- Involve legal counsel, corporate compliance and corporate risk management
- Define mission, mandate, roles and responsibilities
- Establish a strong project team
- IT, RM, business process people, legal counsel, corporate compliance and corporate risk management

2. Build Corporate Awareness

- Launch an education plan/strategy
- Up (to management) educate management on importance of e-records
- Across (to stakeholders)
- Down (to end users)
- Tie to business ethics training (when appropriate)
- Build the business case (if required)
- Tangibles
- Intangibles

3. Establish Underlying Corporate Policies

- Definition of a record. Official vs Transitory records. When to declare a record. (This will vary depending on the record and the business process.)
- Requirements of systems and processes that generate/store records, including quality and reliability of such systems
- Email policy (usage and application of e-records)
- Mainframe and client / server applications and database policy
- Policy's strategic place and role within the greater organisation's policies/framework
- Compliance, privacy, legal, public's right to information (FIOA) etc.
- Implementation plan
- Resources required
- File plan and retention schedule
- Access control (security)
- Policy review/Audit process

4. Enshrine the New Policies

- Communicate clearly from the top. This is really education.
- Provide the policy to all stakeholders and everyone who has a role to play

5. Build/Strengthen RM Foundation

- Creation of update/revise file plan & retention plan
- Use Skupsky's retention method or other 3rd party retention method or consultation.
- Ensure adequate skills and resources (funding, people)
- Conduct a records inventory (physical as well as electronic)
- Establish metadata requirements and standards (leverage existing corporate or industry standards)

6. Develop Implementation Strategy/Plan

- Decide on relationship between ECM and RM
- Why/When/How/Where to combine the two
- Identify target pilot group
- Define the target business processes to records-enable
- Set clear expectations, goals and objectives
- Timeframe, roles & responsibilities

7. Map Business Processes

- Clearly define and map every target business process
- Revise process to include e-records
- Secure stakeholder agreement on all updated processes
- (Paper systems only) define and map paper-based processes
- Folder storage/retrieval/tracking
- Box storage/retrieval/tracking

8. Implement RM Technology

- Install/configure
- Bulk load file plan, retention schedule
- (Paper Systems Only) Bulk load boxes, folders
- (Paper Systems Only) Define/build any required “custom” paper handling processes
- Train records administrators
- Establish operational physical records management (Paper Systems)
- Install, Configure any/all connectors/existing applications
- Design, build, test, deploy, audit, refine any new enablers

9. Conduct the Initial Pilot

- Select a target business application
- Design/build declaration and classification methods for target host application
- Design and deliver user training:
- Training on the policies and procedures
- Declaration/Classification procedures
- Conduct pilot for 60 days
- Measure against pre-established customer requirements, expectations, goals and objectives of pilot project
- Review the results
- Apply changes/corrections
- Measure results again

10. Enterprise Roll-Out

- Develop roll-out strategy/plan
- Identify target group(s)
- Define the target business processes to records-enable
- Set clear expectations, goals and objectives
- Timeframe, roles and responsibilities

SELF ASSESSMENT EXERCISE

Mention two points for building the organisational structure required to support e-records.

3.5 Document Management Security

Why your business needs to secure its documents

The threat of identity theft and other privacy breaches had escalated in recent years, increasing both awareness and the need to ensure that confidential information remains private. The right EDM solution can ensure that your information is protected.

Security and privacy go hand in hand, and many businesses are required by regulations (HIPAA, SEC, SOX, ISO, etc.) to ensure the privacy of customers' and employees' personal information. With a paper-based filing system, there is no efficient way of securing files beyond locking the filing cabinet.

Locking a filing cabinet is not practical because:

- physically restricting access can hamper legitimate access to files for normal business operations
- this method fails to differentiate employee access to various types of files (credit information, shipping records)
- multiple, locked file cabinets restrict legitimate searches for information as files become spread among cabinets and departments.

The need for security must be weighed against the ability for authorised personnel to have quick access to information. Some document management systems take an "all or nothing" approach, while others may weigh down the organisation with cumbersome procedures to address even simple security setting needs.

With comprehensive document management security, you can control not only who can access the virtual file cabinet, but also who can access individual folders and documents. This approach enables all documents to be filed in one place, providing a complete, unified view of records for authorised personnel.

Document management systems provide security based on a number of factors, while allowing instant, anytime access from any location. A complete document management solution should provide the following:

- A mechanism for classifying documents when they are entered in the system.
- A mechanism for defining access classes or groups, so employees have uniform access to the documents they are authorised to use.
- Ability to apply retention policies that electronically preserve documents.
- Security mechanisms that define access at every level of the document system. (Folder- and document-level access is a minimum requirement.)
- An audit trail that records which users have accessed which documents, and what modifications, if any, they have made. This is a requirement for HIPAA legislation.
- The ability to view documents from multiple office locations.
- Workflow to route files electronically to the appropriate person or group based on standard rules and procedures.
- An efficient method for quick access to documents without the need for labor-intensive searches.
- Management for both scanned images and electronic documents.
- Integration with other business software and systems such as word-processing, accounting, and email. This allows users to file and access documents from applications they are already comfortable using.

4.0 CONCLUSION

The functions of any electronic management system do vary within organisations and businesses in terms of specifics. But generally, an electronic record keeping or electronic records management systems helps organisations and businesses to keep track of data/information within the organisations. Depending on the individual creativity, the specific function continues to enlarge.

5.0 SUMMARY

- An effective records management program is the foundation on which an agency can begin automating records management activities or implementing systems to manage electronic records.
- The management and control of organisational records involves a number of actions taken with respect to individual records or series of records to ensure their authenticity, integrity and reliability, and to preserve their usability over time.
- There are a number of ways that each of these records management functions can be automated or implemented in an ERK or ERM, so this checklist is meant only to be a starting-point for agencies to develop system specifications.

- The threat of identity theft and other privacy breaches had escalated in recent years, increasing both awareness and the need to ensure that confidential information remains private.
- Document management systems provide security based on a number of factors, while allowing instant, anytime access from any location

6.0 TUTOR-MARKED ASSIGNMENT

In the maintenance/use of e-records, identify 10 items an ERM system needs to assist in organising records.

7.0 REFERENCE/FURTHER READING

The National Archives (2008). Typical Records Management (RM) Functions and T ypical RM program.

UNIT 3 DATABASE MANAGEMENT SYSTEM

CONTENTS

1.0 INTRODUCTION

A database management system (DBMS) is computer software designed for the purpose of managing databases based on a variety of data models.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define a database management system
- describe the database management structure
- list the benefits of database management system
- describe the features and capabilities of a typical DBMS
- identify and differentiate the different types and models of DBMS.

3.0 MAIN CONTENT

3.1 Description

A DBMS is a complex set of software programs that controls the organisation, storage, management, and retrieval of data in a database. DBMS are categorised according to their data structures or types, sometime DBMS is also known as Database Manager. It is a set of prewritten programs that are used to store, update and retrieve a database. A DBMS includes:

- A modeling language to define the schema of each database hosted in the DBMS, according to the DBMS data model.
- The four most common types of organisations are the hierarchical, network, relational and object models. Inverted lists and other methods are also used. A given database management system may provide one or more of the four models. The optimal structure depends on the natural organisation of the application's data, and on the application's requirements (which include transaction rate (speed), reliability, maintainability, scalability, and cost).
- The dominant model in use today is the ad hoc one embedded in Structured Query Language (SQL), despite the objections of purists who believe this model is a corruption of the relational model, since it violates several of its fundamental principles for the sake of practicality and performance. Many DBMSs also support the Open Database Connectivity API (Application Program Interface) that supports a standard way for programmers to access the DBMS.

Data structures (fields, records, files and objects) optimised to deal with very large amounts of data stored on a permanent data storage device (which implies relatively slow access compared to volatile main memory).

A database query language and report writer to allow users to interactively interrogate the database, analyse its data and update it according to the users privileges on data.

- It also controls the security of the database.
- Data security prevents unauthorised users from viewing or updating the database. Using passwords, users are allowed access to the entire database or subsets of it called subschemas. For example, an employee database can contain all the data about an individual employee, but one group of users may be authorised to view only payroll data, while others are allowed access to only work history and medical data.
- If the DBMS provides a way to interactively enter and update the database, as well as interrogate it, this capability allows for managing personal databases. However, it may not leave an audit trail of actions or provide the kinds of controls necessary in a multi-user organisation. These controls are only available when a set of application programs are customised for each data entry and updating function.
- A transaction mechanism, that ideally would guarantee the Atomicity, Consistency, Isolation, Durability (ACID) properties,

in order to ensure data integrity, despite concurrent user accesses (concurrency control), and faults (fault tolerance).

- It also maintains the integrity of the data in the database.
- The DBMS can maintain the integrity of the database by not allowing more than one user to update the same record at the same time. The DBMS can help prevent duplicate records via unique index constraints; for example, no two customers with the same customer numbers (key fields) can be entered into the database.
- The DBMS accepts requests for data from the application program and instructs the operating system to transfer the appropriate data.

When a DBMS is used, information systems can be changed much more easily as the organisation's information requirements change. New categories of data can be added to the database without disruption to the existing system.

Organisations may use one kind of DBMS for daily transaction processing and then move the detail onto another computer that uses another DBMS better suited for random inquiries and analysis. Overall systems design decisions are performed by data administrators and systems analysts. Detailed database design is performed by database administrators.

Database servers are specially designed computers that hold the actual databases and run only the DBMS and related software. Database servers are usually multiprocessor computers, with Redundant Array Inexpensive Disks (RAID) disk arrays used for stable storage. Connected to one or more servers via a high-speed channel, hardware database accelerators are also used in large volume transaction processing environments.

DBMSs are found at the heart of most database applications. Sometimes DBMSs are built around a private multitasking kernel with built-in networking support although nowadays these functions are left to the operating system.

3.2 DBMS Benefits

- Improved strategic use of corporate data
- Reduced complexity of the organisation's information systems environment
- Reduced data redundancy and inconsistency
- Enhanced data integrity
- Application-data independence

- Improved security
- Reduced application development and maintenance costs
- Improved flexibility of information systems
- Increased access and availability of data and information
- Logical & Physical data independence
- Concurrent access anomalies.
- Facilitate atomicity problem.
- Provides central control on the system through DBA.

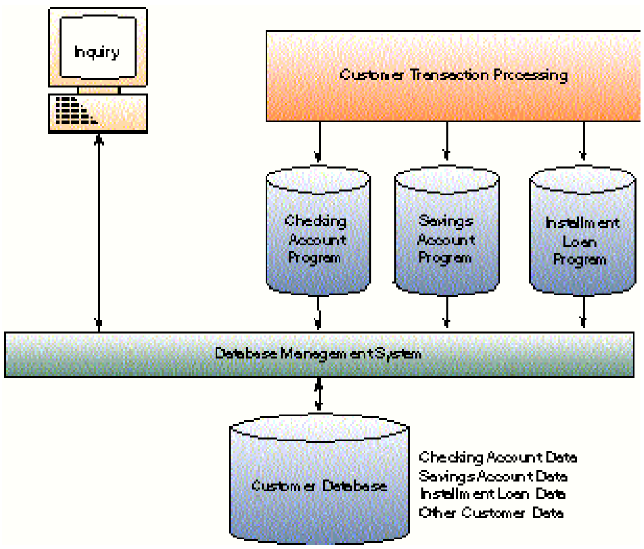


Fig. 1: An Example of a Database Management Approach in a Banking Information System

Note how the savings, checking, and installment loan programs use a database management system to share a customer database. Note also that the DBMS allows a user to make a direct, ad hoc interrogation of the database without using application programs.

3.3 Features and Capabilities of DBMS

A DBMS can be characterised as an "attribute management system" where attributes are small chunks of information that describe something. For example, "colour" is an attribute of a car. The value of the attribute may be a color such as "red", "blue" or "silver".

Alternatively, and especially in connection with the relational model of database management, the relation between attributes drawn from a specified set of domains can be seen as being primary. For instance, the database might indicate that a car that was originally "red" might fade to “pink” in time, provided it was of some particular "make" with an inferior paint job. Such higher relationships provide information on all of the underlying domains at the same time, with none of them being privileged above the others.

Throughout recent history, specialised databases have existed for scientific, geospatial, imaging, document storage and like uses. Functionality drawn from such applications has lately begun appearing in mainstream DBMSs as well. However, the main focus there, at least when aimed at the commercial data processing market, is still on descriptive attributes on repetitive record structures.

Thus, the DBMSs of today roll together frequently-needed services or features of attribute management. By externalising such functionality to the DBMS, applications effectively share code with each other and are relieved of much internal complexity. Features commonly offered by database management systems include:

- Query ability

a. Querying is the process of requesting attribute information from various perspectives and combinations of factors. Example: “How many 2-door cars in Texas are green?”
b. A database query language and report writer allow users to interactively interrogate the database, analyse its data and update it according to the users privileges on data. It also controls the security of the database.
c. Data security prevents unauthorised users from viewing or updating the database. Using passwords, users are allowed access to the entire database or subsets of it called subschemas. For example, an employee database can contain all the data about an individual employee, but one group of users may be authorised to view only payroll data, while others are allowed access to only work history and medical data.
d. If the DBMS provides a way to interactively enter and update the database, as well as interrogate it, this capability allows for managing personal databases. However it may not leave an audit trail of actions or provide the kinds of controls necessary in a multi-user organisation. These controls are only available when a set of application programs are customised for each data entry and updating function.

Backup and replication

Copies of attributes need to be made regularly in case primary disks or other equipment fails. A periodic copy of attributes may also be created for a distant organisation that cannot readily access the original. DBMS usually provide utilities to facilitate the process of extracting and disseminating attribute sets.

When data is replicated between database servers, so that the information remains consistent throughout the database system and users cannot tell or even know which server in the DBMS they are using, the system is said to exhibit replication transparency.

Rule enforcement

Often, one wants to apply rules to attributes so that the attributes are clean and reliable. For example, we may have a rule that says each car can have only one engine associated with it (identified by Engine Number). If somebody tries to associate a second engine with a given car, we want the DBMS to deny such a request and display an error message. However, with changes in the model specification such as, in this example, hybrid gas-electric cars, rules may need to change. Ideally, such rules should be able to be added and removed as needed without significant data layout redesign.

Security

Often, it is desirable to limit who can see or change which attributes or groups of attributes. This may be managed directly by individual, or by the assignment of individuals and privileges to groups, or (in the most elaborate models) through the assignment of individuals and groups to roles which are then granted entitlements.

Computation

There are common computations requested on attributes such as counting, summing, averaging, sorting, grouping, cross-referencing, etc. Rather than have each computer application implement these from scratch, they can rely on the DBMS to supply such calculations and all arithmetical work to be performed by computer which is called a computation.

Change and access logging

Often, one wants to know who accessed what attributes, what was changed, and when it was changed. Logging services allow this by keeping a record of access occurrences and changes.

Automated optimisation

If there are frequently occurring usage patterns or requests, some DBMS can adjust themselves to improve the speed of those interactions. In some cases, the DBMS will merely provide tools to monitor performance, allowing a human expert to make the necessary adjustments after reviewing the statistics collected.

3.4 Uses of DBMS

The four major uses of database management systems are:

- database development
- database interrogation
- database maintenance
- application development

Database Development

Database packages like Microsoft Access, Lotus Approach allow end users to develop the database they need. However, large organisations with client/server or mainframe-based system usually place control of enterprise-wide database development in the hands of database administrators and other database specialists. This improves the integrity and security of organisational database. Database developers use the data definition languages (DDL) in database management systems like oracle 9i or IBM's BD2 to develop and specify the data contents, relationships and structure of each database, and to modify these database specifications called a data dictionary.

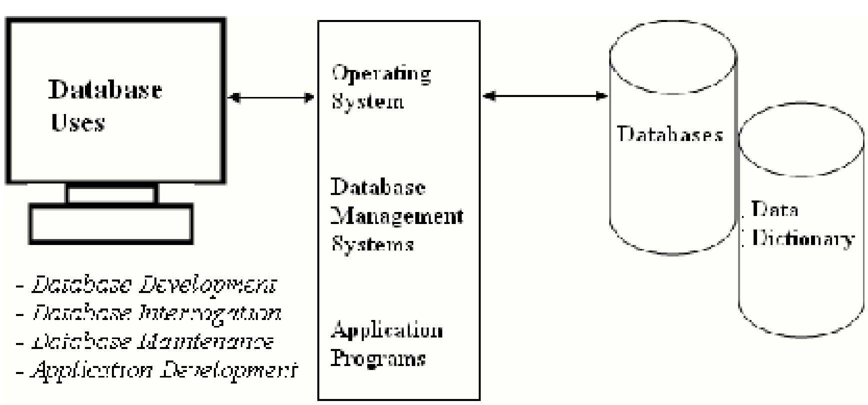


Fig. 2 : The Four Major Uses of DBMS

Database Interrogation

The database interrogation capability is a major use of database management system. End users can interrogate a database management system by asking for information from a database using a query *language or a report generator. They can receive an immediate response* in the form of video displays or printed reports. No difficult programming is required.

Database Maintenance

The databases of organisations need to be updated continually to reflect new business transactions and other events. Other miscellaneous changes must also be made to ensure accuracy of the data in the database. This database maintenance process is accomplished by transaction processing programs and other end-user application packages with the support of the database management system. End-users and information specialists can also employ various utilities provided by a DBMS for database maintenance.

Application Development

Database management system packages play major roles in application development. End-users, systems analysts and other application developers can use the fourth Generational Languages (4GL) programming languages and built-in software development tools provided by many DBMS packages to develop custom application programs. For example you can use a DBMS to easily develop the data entry screens, forms, reports, or web pages by a business application. A database management system also makes the job of application programmers easier, since they do not have to develop detailed data handling procedures using a conventional programming language every time they write a program.

3.5 Models of Database Management Systems

The various models of database management systems are:

- hierarchical
- network
- object-oriented
- associative
- column-Oriented
- navigational
- distributed
- real time relational
- SQL

These models will be discussed in details in subsequent units of this course.

4.0 CONCLUSION

Database management system has continued to make data arrangement and storage to be much easier than it used to be. With the emergence of relational model of database management systems, much of the big challenge in handling large database has been reduced. More database management products will be available in the market as there will be improvement in the existing ones.

5.0 SUMMARY

- A database management system (DBMS) is computer software designed for the purpose of managing databases based on a variety of data models.
- A DBMS is a complex set of software programs that controls the organisation, storage, management, and retrieval of data in a database
- When a DBMS is used, information systems can be changed much more easily as the organisation's information requirements change. New categories of data can be added to the database without disruption to the existing system.
- Often, it is desirable to limit who can see or change which attributes or groups of attributes. This may be managed directly by individual, or by the assignment of individuals and privileges to groups, or (in the most elaborate models) through the assignment of individuals and groups to roles which are then granted entitlements.
- A DBMS can be characterised as an “attribute management system” where attributes are small chunks of information that describe something. For example, “colour” is an attribute of a car. The value of the attribute may be a color such as “red”, “blue” or “silver”.
- Querying is the process of requesting attribute information from various perspectives and combinations of factors. Example: “How many 2-door cars in Texas are green?”
- As computers grew in capability, this trade-off became increasingly unnecessary and a number of general-purpose database systems emerged; by the mid-1960s there were a number of such systems in commercial use. Interest in a standard began to grow, and Charles Bachman, author of one such product, IDS, founded the Database Task Group within CODASYL.

6.0 TUTOR-MARKED ASSIGNMENT

1. List 10 benefits of database management system.
2. Briefly outline how DBMS executes its backup and replication functions.

7.0 REFERENCES/FURTHER READING

Codd, E. F. (1970). “A Relational Model of Data for Large Shared Data Banks”. In: Communications of the ACM 13 (6): 377–387.

O’Brien, James A. (2003). Introduction to Information Systems. (11th ed.). McGraw-Hill.

UNIT 4 COMPUTERS /ELECTRONIC DATA STORAGE

CONTENTS

1.0 INTRODUCTION

Computer data storage, often called storage or memory, refers to computer components, devices, and recording media that retain digital data used for computing for some interval of time. Computer data storage provides one of the core functions of the modern computer, that of information retention. It is one of the fundamental components of all modern computers, and coupled with a central processing unit (CPU, a processor), implements the basic computer model used since the 1940s.

In contemporary usage, memory usually refers to a form of semiconductor storage known as random access memory (RAM) and sometimes other forms of fast but temporary storage. Similarly, storage today more commonly refers to mass storage - optical discs, forms of magnetic storage like hard disks, and other types slower than RAM, but of a more permanent nature. Historically, memory and storage were respectively called primary storage and secondary storage.

The contemporary distinctions are helpful, because they are also fundamental to the architecture of computers in general. As well, they reflect an important and significant technical difference between memory and mass storage devices, which has been blurred by the historical usage of the term storage.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define the concept of electronic storage
- identify the purposes of storage of data
- differentiate the hierarchies of data storage
- identify the characteristics of storage
- describe the fundamental storage technologies.

3.0 MAIN CONTENT

3.1 Purpose of Storage

Various forms of storage, based on various natural phenomena, have been invented. So far, no practical universal storage medium exists, and all forms of storage have some drawbacks. Therefore a computer system usually contains several kinds of storage, each with an individual purpose.

A digital computer represents each datum using the binary numeral system. Text, numbers, pictures, audio, and nearly any other form of information can be converted into a string of bits, or binary digits, each of which has a value of 1 or 0. The most common unit of storage is the byte, equal to 8 bits. A piece of information can be handled by any computer whose storage space is large enough to accommodate the *binary representation of the piece of information, or simply data. For* example, using eight million bits, or about one megabyte, a typical computer could store a small novel.

Traditionally, the most important part of every computer is the central processing unit (CPU, or simply a processor), because it actually operates on data, performs any calculations, and controls all the other components.

Without a significant amount of memory, a computer would merely be able to perform fixed operations and immediately output the result. It would have to be reconfigured to change its behavior. This is acceptable for devices such as desk calculators or simple digital signal processors. Von Neumann machines differ in that they have a memory in which they store their operating instructions and data. Such computers are more versatile in that they do not need to have their hardware reconfigured for each new program, but can simply be reprogrammed with new in-memory instructions; they also tend to be simpler to design, in that a relatively simple processor may keep state between successive computations to build up complex procedural results. Most modern computers are Von Neumann machines.

In practice, almost all computers use a variety of memory types, organised in a storage hierarchy around the CPU, as a tradeoff between performance and cost. Generally, the lower a storage is in the hierarchy, the lesser its bandwidth and the greater its access latency is from the CPU. This traditional division of storage to primary, secondary, tertiary and off-line storage is also guided by cost per bit.

3.2 Hierarchy of Storage

Various forms of storage are divided according to their distance from the central processing unit. The fundamental components of a general-purpose computer are arithmetic and logic unit, control circuitry, storage space, and input/output devices. Technology and capacity were in common home computers around 2005.

Primary storage

Primary storage, presently known as memory, is the only one directly accessible to the CPU. The CPU continuously reads instructions stored there and executes them. Any data actively operated on, is also stored there in uniform manner.

Historically, early computers used delay lines, Williams's tubes, or rotating magnetic drums as primary storage. By 1954, those unreliable methods were mostly replaced by magnetic core memory, which was still rather cumbersome. Undoubtedly, a revolution was started with the invention of a transistor that soon enabled then-unbelievable miniaturisation of electronic memory via solid-state silicon chip technology.

This led to a modern random access memory (RAM). It is small-sized, light, but quite expensive at the same time. (The particular types of RAM used for primary storage are also volatile, i.e. they lose the information when not powered).

Processor registers are located inside the processor. Each register typically holds a word of data (often 32 or 64 bits). CPU instructions instruct the arithmetic and logic unit to perform various calculations or other operations on this data (or with the help of it). Registers are technically among the fastest of all forms of computer data storage, being switching transistors integrated on the CPU's chip, and functioning as electronic "flip-flops".

Processor cache is an intermediate stage between ultra-fast registers and much slower main memory. It was introduced solely to increase performance of the computer. Most actively used information in the main memory is just duplicated in the cache memory, which is faster, but of much lesser capacity. On the other hand, it is much slower, but much larger than processor registers. Multi-level hierarchical cache setup is also commonly used—primary cache being smallest, fastest and located inside the processor; secondary cache being somewhat larger and slower.

Main memory is directly or indirectly connected to the CPU via a *memory bus, today sometimes referred to as a front side bus. It is* actually comprised of two buses (not on the diagram): an address bus and a data bus. The CPU firstly sends a number through an address bus, a number called memory address that indicates the desired location of data. Then it reads or writes the data itself using the data bus. Additionally, a Memory Management Unit (MMU) is a small device between CPU and RAM recalculating the actual memory address, for example, to provide an abstraction of virtual memory or other tasks.

As the RAM types used for primary storage are volatile (cleared at start up), a computer containing only such storage would not have a source to read instructions from, in order to start the computer. Hence, non-volatile primary storage containing a small startup program (BIOS: Basic Input Output System) is used to bootstrap the computer, that is, to read a larger program from non-volatile secondary storage to RAM and start to execute it. A non-volatile technology used for this purpose is called ROM, for read-only memory (the terminology may be somewhat confusing as most ROM types are also capable of random access).

Many types of “ROM” are not literally read only, as updates are possible; however it is slow and memory must be erased in large portions before it can be re-written. Some embedded systems run programs directly from ROM (or similar), because such programs are rarely changed. Standard computers do not store non-rudimentary programs in ROM, rather, they use large capacities of secondary storage, which is non-volatile as well, and not as costly.

Recently, primary storage and secondary storage in some uses refer to what was historically called, respectively, secondary storage and *tertiary storage.*

Secondary storage

Secondary storage, or storage in popular usage, differs from primary storage in that it is not directly accessible by the CPU. The computer usually uses its input/output channels to access secondary storage and transfers desired data using intermediate area in primary storage. Secondary storage does not lose the data when the device is powered down—it is non-volatile. Per unit, it is typically also in an order of magnitude less expensive than primary storage. Consequently, modern computer systems typically have an order of magnitude more on the secondary storage than primary storage and data is kept for a longer time there.

In modern computers, hard disks are usually used as secondary storage. The time taken to access a given byte of information stored on a hard disk is typically a few thousandths of a second, or milliseconds. By contrast, the time taken to access a given byte of information stored in random access memory is measured in billionths of a second, or nanoseconds. This illustrates the very significant access-time difference which distinguishes solid-state memory from rotating magnetic storage devices: hard disks are typically about a million times slower than memory. Rotating optical storage devices, such as CD and DVD drives, have even longer access times.

Some other examples of secondary storage technologies are: flash memory (e.g. USB sticks or keys), floppy disks, magnetic tape, paper tape, punch cards, standalone RAM disks, and Zip drives.

The secondary storage is often formatted according to a filesystem format, which provides the abstraction necessary to organise data into files and directories, providing also additional information (called metadata) describing the owner of a certain file, the access time, the access permissions, and other information.

Most computer operating systems use the concept of virtual memory, allowing utilisation of more primary storage capacity than is physically available in the system. As the primary memory fills up, the system moves the least-used chunks (pages) to secondary storage devices (to a swap file or page file), retrieving them later when they are needed. As more of these retrievals from slower secondary storage are necessary, the more the overall system performance is degraded.

Tertiary storage

Large tape library. Tape cartridges are placed on shelves in the front, robotic arm moving in the back. Visible height of the library is about 180 cm.

Tertiary storage or tertiary memory provides a third level of storage. Typically, it involves a robotic mechanism which will mount (insert) and *dismount removable mass storage media into a storage device according* to the system's demands; this data is often copied to secondary storage before use. It is primarily used for archival of rarely accessed information since it is much slower than secondary storage (e.g. 5-60 seconds vs. 1-10 milliseconds). This is primarily useful for extraordinarily large data stores, accessed without human operators. Typical examples include tape libraries and optical jukeboxes.

When a computer needs to read information from the tertiary storage, it will first consult a catalogue database to determine which tape or disc contains the information. Next, the computer will instruct a robotic arm to fetch the medium and place it in a drive. When the computer has finished reading the information, the robotic arm will return the medium to its place in the library.

3.3 Off-Line Storage

Off-line storage, also known as disconnected storage, is computer data storage on a medium or a device that is not under the control of a processing unit. The medium is recorded, usually in a secondary or tertiary storage device, and then physically removed or disconnected. It must be inserted or connected by a human operator before a computer can access it again. Unlike tertiary storage, it cannot be accessed without human interaction.

Off-line storage is used to transfer information, since the detached medium can be easily physically transported. Additionally, in case a disaster, for example, a fire, destroys the original data, a medium in a remote location will be probably unaffected, enabling disaster recovery. Off-line storage increases a general information security, since it is physically inaccessible from a computer, and data confidentiality or integrity cannot be affected by computer-based attack techniques. Also, if the information stored for archival purposes is accessed seldom or never, off-line storage is less expensive than tertiary storage.

In modern personal computers, most secondary and tertiary storage media are also used for off-line storage. Optical discs and flash memory devices are most popular, and to much lesser extent removable hard disk drives. In enterprise uses, magnetic tape is predominant. Older examples are floppy disks, Zip disks, or punched cards.

SELF ASSESSMENT EXERCISE

What is the meaning of the following acronyms?

RAM, ROM and MMU?

3.4 Characteristics of Storage

Storage technologies at all levels of the storage hierarchy can be differentiated by evaluating certain core characteristics as well as measuring characteristics specific to a particular implementation. These core characteristics are volatility, mutability, accessibility, and addressibility. For any particular implementation of any storage technology, the characteristics worth measuring are capacity and performance.

Volatility

Non-volatile memory

Will retain the stored information even if it is not constantly supplied with electric power. It is suitable for long-term storage of information. Nowadays, it is used for most of secondary, tertiary, and off-line storage. In 1950s and 1960s, it was also used for primary storage, in the form of magnetic core memory.

Volatile memory

Requires constant power to maintain the stored information. The fastest memory technologies of today are volatile ones (not a universal rule). Since primary storage is required to be very fast, it predominantly uses volatile memory.

Differentiation

Dynamic memory

A form of volatile memory which also requires the stored information to be periodically re-read and re-written, or refreshed, otherwise it would vanish.

Static memory

A form of volatile memory similar to DRAM (Dynamic Random Access Memory); with the exception that it does not refresh on occasion.

Mutability

Read/write storage or mutable storage

Allows information to be overwritten at any time. A computer without some amount of read/write storage for primary storage purposes would be useless for many tasks. Modern computers typically use read/write storage also for secondary storage.

Read only storage

Retains the information stored at the time of manufacture, and write **once storage (WORM: Write Once Read Many) allows the information** to be written only once at some point after manufacture. These are called **immutable storage. Immutable storage is used for tertiary and off-line** storage. Examples include CD-ROM and CD-R.

Slow write, fast read storage

Read/write storage which allows information to be overwritten multiple times, but with the write operation being much slower than the read operation. Examples include CD-RW.

Accessibility

Random access

Any location in storage can be accessed at any moment in approximately the same amount of time. Such characteristic is well suited for primary and secondary storage.

Sequential access

The accessing of pieces of information will be in a serial order, one after the other; therefore the time to access a particular piece of information depends upon which piece of information was last accessed. Such characteristic is typical of off-line storage.

Addressability

Location-addressable

Each individually accessible unit of information in storage is selected with its numerical memory address. In modern computers, location-addressable storage usually limits to primary storage, accessed internally by computer programs, since location-addressability is very efficient, but burdensome for humans.

File addressable

Information is divided into files of variable length, and a particular file is selected with human-readable directory and file names. The underlying device is still location-addressable, but the operating system of a computer provides the file system abstraction to make the operation more understandable. In modern computers, secondary, tertiary and off-line storage use file systems.

Content-addressable

Each individually accessible unit of information is selected with a hash value, or a short identifier with a number pertaining to the memory address the information is stored on. Content-addressable storage can be implemented using software (computer program) or hardware (computer device), with hardware being faster but it is a more expensive option.

Capacity

Raw capacity

The total amount of stored information that a storage device or medium can hold. It is expressed as a quantity of bits or bytes (e.g. 10.4 megabytes).

Density

The compactness of stored information. It is the storage capacity of a medium divided with a unit of length, area or volume (e.g. 1.2 megabytes per square inch).

Performance

Latency

The time it takes to access a particular location in storage. The relevant unit of measurement is typically nanosecond for primary storage, millisecond for secondary storage, and second for tertiary storage. It may make sense to separate read latency and write latency, and in case of sequential access storage, minimum, maximum and average latency.

Throughput

The rate at which information can be read from or written in the storage.
In computer data storage, throughput is usually expressed in terms of megabytes per second or MB/s, though bit rate may also be used. As with latency, read rate and write rate may need to be differentiated. Also accessing media sequentially, as opposed to randomly, typically yields maximum throughput.

3.5 Fundamental Storage Technologies

As of 2008, the most commonly used data storage technologies are semiconductor, magnetic, and optical, while paper still sees some limited usage. Some other fundamental storage technologies have also been used in the past or are proposed for development.

Semiconductor

Semiconductor memory uses semiconductor-based integrated circuits to store information. A semiconductor memory chip may contain millions of tiny transistors or capacitors. Both volatile and non-volatile forms of semiconductor memory exist. In modern computers, primary storage almost exclusively consists of dynamic volatile semiconductor memory or dynamic random access memory. Since the turn of the century, a type of non-volatile semiconductor memory known as flash memory has steadily gained share as off-line storage for home computers. Non-volatile semiconductor memory is also used for secondary storage in various advanced electronic devices and specialised computers.

Magnetic

Magnetic storage uses different patterns of magnetisation on a magnetically coated surface to store information. Magnetic storage is *non-volatile. The information is accessed using one or more read/write* heads which may contain one or more recording transducers. A read/write head only covers a part of the surface so that the head or medium or both must be moved relative to another in order to access data. In modern computers, magnetic storage will take these forms:

Magnetic disk

- Floppy disk, used for off-line storage
- Hard disk, used for secondary storage
- Magnetic tape data storage, used for tertiary and off-line storage

In early computers, magnetic storage was also used for primary storage in a form of magnetic drum, or core memory, core rope memory, thin film memory, twistor memory or bubble memory. Also unlike today, magnetic tape was often used for secondary storage.

Optical storage, the typical optical disc, stores information in deformities on the surface of a circular disc and reads this information by illuminating the surface with a laser diode and observing the reflection. Optical disc storage is non-volatile. The deformities may be permanent (read only media), formed once (write once media) or reversible (recordable or read/write media). The following forms are currently in common use:

- CD, CD-ROM, DVD, BD-ROM: Read only storage, used for mass distribution of digital information (music, video, computer programs)
- CD-R, DVD-R, DVD+R BD-R: Write once storage, used for tertiary and off-line storage
- CD-RW, DVD-RW, DVD+RW, DVD-RAM, BD-RE: Slow write, fast read storage, used for tertiary and off-line storage
- Ultra Density Optical or UDO is similar in capacity
- BD-R or BD-RE is slow write, fast read storage used for tertiary and off-line storage.

Magneto-optical disc storage is optical disc storage where the magnetic state on a ferromagnetic surface stores information. The information is read optically and written by combining magnetic and optical methods. Magneto-optical disc storage is non-volatile, sequential *access, slow write, fast read storage used for tertiary and off-line* storage.

3D optical data storage has also been proposed

Paper

Paper data storage, typically in the form of paper tape, or punch cards, has long been used to store information for automatic processing, particularly before general-purpose computers existed. Information was recorded by punching holes into the paper or cardboard medium and was read mechanically (or later optically) to determine whether a particular location on the medium was solid or contained a hole.

Uncommon

Vacuum tube memory

A Williams's tube used a cathode ray tube, and a Selectron tube used a large vacuum tube to store information. These primary storage devices were short-lived in the market, since Williams tube was unreliable and Selectron tube was expensive.

Electro-acoustic memory

Delay line memory used sound waves in a substance such as mercury to store information. Delay line memory was dynamic volatile, cycle sequential read/write storage, and was used for primary storage.

Phase-change memory

Uses different mechanical phases of phase change material to store information in an X-Y addressable matrix, and reads the information by observing the varying electric resistance of the material. Phase-change memory would be non-volatile, random access read/write storage, and might be used for primary, secondary and off-line storage. Most rewritable and many write once optical disks already use phase change material to store information.

Holographic storage

This stores information optically inside crystals or photopolymers. Holographic storage can utilise the whole volume of the storage medium, unlike optical disc storage which is limited to a small number of surface layers. Holographic storage would be non-volatile, sequential access, and either write once or read/write storage. It might be used for secondary and off-line storage. See Holographic Versatile Disc (HVD).

Molecular memory

Molecular memory stores information in polymers which can retain electric charge. Molecular memory is especially suited for primary storage.

4.0 CONCLUSION

Storage is an important aspect of a data processing system. Several storage devices are being developed by the day to match the pace of development in other parts of data processing. Improvements are seen in the capacity, durability and sizes of storage media.

5.0 SUMMARY

- Computer data storage, often called storage or memory, refers to computer components, devices, and recording media that retain digital data used for computing for some interval of time.
- Various forms of storage, based on various natural phenomena, have been invented. So far, no practical universal storage medium exists, and all forms of storage have some drawbacks.
- Various forms of storage are divided according to their distance from the central processing unit. The fundamental components of a general-purpose computer are arithmetic and logic unit, control circuitry, storage space, and input/output devices technology and capacity as in common home computers around 2005.
- Off-line storage, also known as disconnected storage, is a computer data storage on a medium or a device that is not under the control of a processing unit. The medium is recorded, usually in a secondary or tertiary storage device, and then physically removed or disconnected.
- Storage technologies at all levels of the storage hierarchy can be differentiated by evaluating certain core characteristics as well as measuring characteristics specific to a particular implementation. These core characteristics are volatility, mutability, accessibility, and addressibility.
- As of 2008, the most commonly used data storage technologies are semiconductor, magnetic, and optical, while paper still sees some limited usage. Some other fundamental storage technologies have also been used in the past or are proposed for development.

6.0 TUTOR-MARKED ASSIGNMENT

1. Briefly discuss Off-Line Storage.
2. Mention five characteristics of electronic storage system.

7.0 REFERENCES/FURTHER READING

Wikipedia (2008). Primary Storage and Storage Hardware.

A Thesis on Tertiary Storage

National Communications System, Federal Standard 1037C-Telecommunications: Glossary of Terms(1996).

The DVD FAQ as a Comprehensive Reference of DVD Technologies.

UNIT 5 SUCCESS FACTORS IN RECORDS MANAGEMENT SYSTEMS

CONTENTS

1.0 INTRODUCTION

Historically, records management was the responsibility of a small number of specialist staff members within an organisation.

With the shift towards electronic records (including documents and e-mails), efforts have instead, turned to rolling out an electronic records management system (RMS) across the entire organisation.

The challenge then becomes ensuring that staff throughout the organisation adopt the use of the system, and take on the added responsibilities involved. Up to this point, many organisations have attempted this 'enterprise' implementation of records management, but few (if any) have succeeded. This unit explores, from a new perspective, the challenges involved in rolling out a records management system, identifying three critical success factors for widespread adoption.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- identify the challenges facing e-records management systems
- identify the critical factors to overcome the challenges
- describe the platforms for users can use in saving records
- identify the goals of classification scheme within the records management system
- discuss some issues to consider in achieving success in records management.

3.0 MAIN CONTENT

3.1 The Challenges

For a long period of time, records management was a fully centralised activity, with a small team of specialist records managers handling the needs of the entire organisation.

The discipline of records management grew up around this model, leading to the creation of applications and classification schemes designed for specialist users.

The growth of electronic records has, however, had a dramatic impact on this approach to records management.

As it is no longer practicable to manage via a fully-centralised model, records management has instead been decentralised throughout the entire organisation. This has led to rollout of records management systems to be used by general business staff, and not just specialist records managers.

The fundamental challenge of rolling out a records management system is that, for it to be successful, it requires the active involvement of all staff. If staff choose not to file documents in the system, or don't use it to search for existing documents, the system (and the project as a whole) has failed. This introduces a tremendous change management challenge, one that involves a widespread transformation of business processes and practices. It is this challenge that is explored in this article: how to get staff to use the records management system.

3.2 Three Critical Success Factors

From observations of many different records management system rollouts through organisations (most of which haven't been successful), we have identified three areas that must be addressed:

1. Software

The design and usability of the records management software, with a particular focus on ease of use for general business users.

2. Classification scheme

Ensuring the classification scheme is designed to ensure that the general business users, store records in the correct location, and to help them find records again at a later date.

3. Message

The 'message' that drives the internal communications and change management efforts is simply that 'you must use the software'.
Each of these critical success factors is explored in the following sections.

3.3 Software

Records management systems were designed to be used, and used very efficiently, by experts. With a centralised user base of specialist records managers, these systems were developed to provide powerful tools to speed the day-in, day-out work of these staff.

As a result, most records management systems are hugely complex. In most systems, simply filing a document presents the user with a huge dialog box with dozens of drop-down lists, buttons and options.

While the design of these systems has served the records community well for some time, they are now out of step with the needs of the new users of records management systems: general staff throughout organisations.

For these staff, using records management systems, as they are currently designed, can be a terrifying prospect.

The need for simplicity

Users are overwhelmed by the complexity of most records management systems. In fact, it is fair to say that the very poor usability of these systems is one of the single greatest barriers to project success.

Beyond this, it can be argued that unless the usability issues can be addressed, enterprise-wide records management projects will never be *successful.*

What is needed is a simple interface that is at least as easy as other mechanisms staff use for filing records (formally or informally).

A number of approaches have already been taken in records management projects:

- centralising the creation of new files
- limiting the areas of the software that most users have to interact with
- providing training, 'cheat sheets' and other support tools

While these have been valuable, none has tackled the fundamental design problems of most records management systems.

Instead, organisations should plan to test the usability of any records management system being implemented, with the expectation of having to address a range of issues and problems.

While in the short-term, this will end up being the responsibility of organisations implementing records solutions, pressure will inevitably be applied on the vendors themselves to improve product usability.

Competing platforms

There are many different competing platforms that user can make use of when saving records:

- e-mail inbox (or other e-mail folders)
- outlook public folders
- local drive
- network drive
- intranet
- collaboration tools and team workspaces
- lotus notes
- paper (personal files, notes pinned to walls, etc)

Some organisations have attempted to eliminate these other mechanisms for storing information, with the aim of forcing staff to use the records management system.

For example, access to local drives or network drives is restricted, thereby removing one of the most common locations for saving documents.

While this may have some positive impact on staff usage, the frustration generated will invariably increase the resistance to change within the organisation.

Unless there is a strong culture of record keeping (such as in a legal firm), staff will find new (and creative) ways of 'passively resisting' the rollout of the system.

In many cases, they will continue to use their existing filing methods (however ad-hoc), or find new (and simpler) alternatives to the records management system.

Fundamentally, staff will (and should) use the solution that is easiest and most effective for them. The challenge therefore is to ensure that the records management system is very easy to use, on par with other options available to staff.

Invisible operation

While ensuring that the usability of records management systems is vital, some have argued that a necessary step is to make the systems entirely invisible.

In this model, records management systems are seamlessly integrated with core business systems, whether e-mail, customer management systems or front-line applications.

When records are filed from these systems, the context of the users current activities are used to pre-fill most (or all) of the details required by the records management system.

In this way, the system becomes little more than a menu item and a simple dialogue box. There are already systems in the market (particularly in the field of document management) that are exploring this concept.

Whatever the approach, effort must be taken to reduce the barriers that are currently making it too difficult for staff to file records, even when they are self-motivated to do so.

3.4 Classification Scheme

The classification scheme within the records management system serves three main goals:

- helping staff to determine where to file records
- assisting them to find (retrieve) records
- meeting statutory requirements

At present, the strongest emphasis within most projects is on the last of these goals.

Particularly within the public sector, the adoption of records management systems is being driven by the demands of the Archives Act. This typically leads to the adoption of a 'functional classification scheme' (such as Keyword AAA or a variation of it).

These schemes work from general concepts down to the most specific, and are primarily designed to assist in meeting the legislated requirements relating to archiving and disposal.

Finding and storing records

While meeting statutory requirements is important, widespread usage of the records management system primarily rests upon meeting the first two goals (filing and retrieving records).

In this era of enterprise-wide rollouts of records management systems, the users of the classification scheme have changed dramatically. While historically the classification was only really used by the centralised records management staff, it will now be used by staff throughout the organisation.

In order to meet these needs, the classification scheme must be understandable by general business staff.

It is clear that:

- If staff members are unable to easily determine where to file records, the records management system will become littered with mis-filed documents.
- If staff members cannot easily find documents at a later date, then they will simply abandon the use of the records management system entirely.

For these reasons, the effective design of the classification scheme becomes imperative.

Caloundra City Council case study

In this project, the Council planned to implement the Keywords for Councils classification scheme (a derivative of Keyword AAA), but had identified that the success of the project as a whole would rest on the appropriateness of this scheme.

A small-scale usability test was therefore conducted of the classification scheme. This identified a number of important issues:

- success rates in using the classification scheme were highly variable, across both users and tasks
- staff had considerable difficulty understanding the top level of the classification scheme (functions)
- most staff think at the most specific level
- experience with the records management system or classification scheme did not improve levels of success

Design for the greatest audience

In the light of these findings, it is apparent that the classification schemes typically used may not be appropriate for general staff within an organisation.

While training and support may assist to some degree, they will not eliminate the problems generated by an inappropriate classification scheme. Reliance on such approaches will also burden organisations with a never-ending requirement to continue this training and support.

The alternative is to implement a classification scheme that is designed for the greatest audience: general staff throughout the organisation. Taking this path not only reduces the need for training, but also directly addresses one of the greatest causes of frustration with current records management projects.

Successful approaches

There have already been a number of organisations that have moved away from the functional classification scheme for some (or all) of their staff.

Instead, records are classified in line with the core business processes in the organisation. For example:

- legal firms file documents according to client and matter
- councils file according to customer or area of land
- project managers file according to project
- customer service staff file according to client (for example, in CRM systems)

While these examples do not provide a complete solution, they do demonstrate the value of exploring other approaches to the classification scheme.

Practical testing techniques

Thankfully, it is not necessary to rely on a philosophical approach to determine the most effective classification scheme.

Instead, organisations can use a range of practical testing techniques to assess (or design) a classification scheme. Taking this approach will ensure that the end result will be useful for staff throughout the organisation.

The Caloundra City Council case study presents one simple and cost-effective approach to usability testing a records classification scheme. This was completed in only two days, and gave sufficient information to suggest an overall approach (more research would have been required to rework the classification scheme).

Card-based classification evaluation provides another mechanism to directly assess how well staff can understand and use a classification scheme (whether for a records system or for an intranet).

Implement two schemes?

While we have focused on techniques and approaches for implementing a scheme designed for general users, the statutory requirements must still be met.

While it is beyond the scope of this article to propose a total solution to these two potentially conflicting needs, one approach is to implement two classification schemes.

The first scheme is staff-facing, structured along business lines. The second scheme is a functional classification scheme (such as Keyword AAA, expanded to include core business functions).

A mapping is then maintained between the two schemes. General business staff members only ever see the first scheme, while records managers use the second scheme to manage retention and disposal.

A number of document management systems in the marketplace had already implemented this approach, demonstrating its feasibility.
Of course, considerable work is required to maintain these mappings, but at least this is restricted to the staff best able to manage it (record keeping professionals).

SELF ASSESSMENT EXERCISE

Identify the three main goals of classification schemes in records management system.

3.5 Message

Too many records management implementations are driven by a single message to staff: “you must use it!” This is generally supported by highlighting the importance of records management, and the organisation's legislative responsibilities.

The problem is that most staff are fundamentally disinterested in the concept of records management, and there is no easy way of generating the required enthusiasm.

While records management is extremely important for the organisation as a whole, and for those responsible for information management, simply highlighting this to staff will have little effect.

In most organisations, the net result of this approach is to generate small pockets of effective record-keeping (where staff can see for themselves the relevance and value of records management), while the rest of staff essentially ignore the rollout.

Cannot be enforced

The biggest challenge for records management staff in rolling out a new system is that staff can't be forced to use it.

While the establishment of suitable policies and guidelines is certainly required, these alone will have little impact on uptake throughout the organisation.

Furthermore, few (if any) staff directly report to the records team. Even when the message is driven down from senior management, staff members have many ways of 'passively resisting' the rollout.

While a 'carrot and stick' approach is often recommended for these types of projects, in practice, the 'stick' is generally not practical or possible. Instead, the records management project must encourage uptake amongst staff, by identifying one or more messages that will engage staff.

What's in it for me factor

What is generally missing from change management and internal communications activities conducted for records management a project is the 'what's in it for me' factor.

While the benefits to the organisation as a whole are clearly articulated, the more immediate benefits for individual staff are not highlighted.

For records management projects to be successful, these individual benefits must be determined and then clearly communicated.

Aligning with business activities

One approach for determining the 'what's in it for me' factor is to strongly align the records management system with core business activities. For example, project teams have a clear need for an effective way of storing, tracking and communicating project documents. By promoting the solution as a ‘project management solution’ (instead of a records management solution), interest and usage can be generated. Those involved in customer service or client relations have a similarly clear need, as do staff in legal or contracts sections.

In all of these cases, the needs of the individual teams are determined, and the records management system is designed and promoted accordingly. T his model of records management rollout is an incremental one, with uptake being ensured on a team-by-team basis. While it is much slower than a single enterprise-wide rollout, it is much more likely to generate sustained usage.

A more general message

While the approach outlined above can be very effective, it is best suited to those sections of the business that have clearly defined processes and needs. For the rest of the organisation, a more general message will be needed. This message can be determined from the observation that most staff members are clearly aware of the inadequacies of their current 'personal information management' practices. Staff members are flooded with information every day that they have a desire to keep, and they struggle with a variety of methods, none very effective:

- retaining messages in the e-mail system, until forced to delete them by IT
- storing documents on a shared drive, where the lack of consistent structure rapidly makes it hard to find documents, and generates many duplicates
- printing documents, and stacking them up in piles or pinning them to partition walls

Staff members struggle with these approaches because no better solution has been provided to them. If a better solution, in the form of the records management solution is made available, this will naturally lead to adoption.

The message then becomes:

- "Storing your documents in the records management system is the single easiest way of ensuring you can quickly find them again when you need them in six months or a year's time."

This is the 'what's in it for me' factor: a simpler way of meeting personal needs for storing and retrieving information.

Of course, this message will only succeed if it is actually true:

- the records management system must be quick and easy to use (the first critical success factor)
- the classification scheme must be designed so that staff can actually find documents again at a later date (the second critical success factor).

If the records management system is genuinely designed to meet the needs of staff, then communicating this fact will inevitably lead to sustained use.

3.6 More Issues

While this unit has focused on three specific issues relating to records management adoption (system, classification scheme and message), these are obviously not the only aspects that need to be addressed during the project.

Other issues to consider include:

- implementation of hardware and other necessary infrastructure
- deployment of the software
- migration of records
- records management policies and guidelines
- management of both paper and electronic records
- appropriate security measures
- ongoing resource levels
- integration with other systems (including customer management systems and the intranet) etc.

While this unit has not covered any of these issues, the focus on the three specific areas is deliberate, as they will have the greatest impact upon project success.

While all aspects of the project must be explored, failure to address the three critical factors will almost certainly lead to project failure.

As a final note, it is worth highlighting that while this article has focused on records management systems, it is equally applicable to document management systems.

In many cases, the two systems are now considered aspects of the one solution, referred to as an Electronic Documents and Records Management System (EDRMS).

Even when being deployed separately, documents management systems face the same challenges as their records management counterparts, and focusing on the same three aspects will have the greatest impact upon project success.

4.0 CONCLUSION

Projects planning to roll out records management systems across an entire organisation face considerable challenges, not least of which is that they require the active participation of all staff to be successful. To achieve this level of cultural change, three critical success factors have been identified. They are:

- ensuring that records management systems are sufficiently usable for general staff throughout the organisation.
- implementing classification schemes that are matched to the needs and working practices of all staff.
- identifying a clear message that will resonate with users and drive real adoption of the records management system.

While there are many other aspects to be managed within a records management project, these three factors will have the greatest impact upon the success of the project as a whole.

5.0 SUMMARY

- Historically, records management was the responsibility of a small number of specialist staff members within an organisation.
- For a long period of time, records management was a fully centralised activity, with a small team of specialist records managers handling the needs of the entire organisation. The discipline of records management grew up around this model, leading to the creation of applications and classification schemes designed for specialist users.

- From observations of many different records management system rollouts through organisations (most of which haven't been successful), we have identified three areas that must be addressed:
- Records management systems were designed to be used, and used very efficiently, by experts. With a centralised user base of specialist records managers, these systems were developed to provide powerful tools to speed the day-in, day-out work of these staff.
- While meeting statutory requirements is important, widespread usage of the records management system primarily rests upon meeting the first two goals (filing and retrieving records).
- Too many records management implementations are driven by a single message to staff: “you must use it!”. This is generally supported by highlighting the importance of records management, and the organisation’s legislative responsibilities.
- While this unit has focused on three specific issues relating to records management adoption (system, classification scheme and message), these are obviously not the only aspects that need to be addressed during the project.

6.0 TUTOR-MARKED ASSIGNMENT

1. Identify 5 competing platforms for user in saving records
2. Briefly discuss the issue of finding and storing records

7.0 REFERENCE/FURTHER READING

Robertson, J. (undated). Success Factors in Records Management Systems, Step II Designs, Sydney, Australia.

MODULE 2

UNIT 1 DATA PROCESSING AND MANAGEMENT

CONTENTS

Process of Data Processing

1.0 INTRODUCTION

Data processing is any computer process that converts data into information or knowledge. The processing is usually assumed to be automated and running on a computer. Because data are most useful when well-presented and actually informative, data-processing systems are often referred to as information systems to emphasis their practicality. Nevertheless, both terms are roughly synonymous, performing similar conversions; data-processing systems typically manipulate raw data into information, and likewise information systems typically take raw data as input to produce information as output.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define data processing from different perspectives
- identify the elements that constitutes data processing
- explain the concept of information repository
- describe Automated Data Management policies
- explain the concept of data backup
- describe the concept of data recovery
- explain the selection, extraction and manipulation of data.

3.0 MAIN CONTENT

3.1 Definitions

Data processing is defined as numbers or characters that represent measurements from observable phenomena. A single datum is a single measurement from observable phenomena. Measured information is then algorithmically derived and/or logically deduced and/or statistically calculated from multiple data. Information is defined as either a meaningful answer to a query or a meaningful stimulus that can cascade into further queries.

For example, gathering seismic data leads to alteration of seismic data to suppress noise, enhance signal and migrate seismic events to the appropriate location in space. Processing steps typically include analysis of velocities and frequencies, static corrections, deconvolution, normal move out, dip move out, stacking, and migration, which can be performed before or after stacking. Seismic processing facilitates better interpretation because subsurface structures and reflection geometries are more apparent.

General

More generally, the term data processing can apply to any process that converts data from one format to another, although data conversion would be the more logical and correct term. From this perspective, data processing becomes the process of converting information into data and also the converting of data back into information. The distinction is that conversion doesn't require a question (query) to be answered. For example, information in the form of a string of characters forming a sentence in English is converted or encoded meaningless hardware-oriented data to evermore-meaningful information as the processing proceeds toward the human being.

Embedded system

Conversely, that simple example for pedagogical purposes here is usually described as an embedded system (for the software resident in the keyboard itself) or as (operating) systems programming, because the information is derived from a hardware interface and may involve overt control of the hardware through that interface by an operating system. Typically, control of hardware by a device driver manipulating Application Specific Interface Circuit (ASIC) or (FPGA) Field Programmable Gate Array registers is not viewed as part of data processing proper or information systems proper, but rather as the domain of embedded systems or (operating) systems programming. Instead, perhaps a more conventional example of the established practice of using the term data processing is that a business has collected numerous data concerning an aspect of its operations and that this multitude of data must be presented in meaningful, easy-to-access presentations for the managers who must then use that information to increase revenue or to decrease cost. That conversion and presentation of data as information is typically performed by a data-processing application.

Data analysis

When the domain from which the data are harvested is a science or engineering, data processing and information systems are considered too broad of terms and the more specialised term data analysis is typically used, focusing on the highly-specialised and highly-accurate algorithmic derivations and statistical calculations that are less often observed in the typical general business environment. In these contexts, Data Analysis Packages like Download Accelerator Plus (DAP) or Proposed System Package Plan (PSPP) are often used. This divergence of culture is exhibited in the typical numerical representations used in data processing versus numerical; data processing's measurements are typically represented by integers or by fixed-point or binary-coded decimal representations of numbers whereas the majority of data analysis measurements are often represented by floating-point representation of rational numbers.

Processing

Practically, all naturally occurring processes can be viewed as examples of data processing systems where "observable" information in the form of pressure, light, etc. are converted by human observers into electrical signals in the nervous system as the senses we recognise as touch, sound, and vision. Even the interaction of non-living systems may be viewed in this way as rudimentary information processing systems. Conventional usage of the terms data processing and information *systems restricts their use to refer to the algorithmic derivations, logical* deductions, and statistical calculations that recur perennially in general business environments, rather than in the more expansive sense of all conversions of real-world measurements into real-world information in, say, an organic biological system or even a scientific or engineering system.

3.2 Elements of Data Processing

In order to be processed by a computer, the data needs first to be converted into a machine readable format. Once data is in digital format, various procedures can be applied on the data to get useful information. Data Processing includes all the processes from Data Entry up to Data Mining:

- Data Entry
- Data Cleaning
- Data Coding
- Data Translation
- Data Summarisation
- Data Aggregation
- Data Validation
- Data Tabulation
- Statistical Analysis
- Computer graphics
- Data Warehousing
- Data Mining

3.3 Data Processor

In data processing or information processing, a Data Processor or Data **Processing Unit or Data Processing System is a system which** processes data which has been captured and encoded in a format recognisable by the data processing system or has been created and stored by another unit of an information processing system.

A data entry is a specialised component or form of information processing (sub) system. Its chief difference is that it tends to perform a dedicated function (i.e., its program is not readily changeable). Its dedicated function is normally to perform some (intermediate) step of converting input ('raw' or unprocessed) data, or semi-processed information, in one form into a further or final form of information through a process called decoding / encoding or formatting or re-formatting or translation or data conversion before the information can be output from the data processor to a further step in the information processing system.

For the hardware data processing system, this information may be used to change the sequential states of a (hardware) machine called a computer. In all essential aspects, the hardware data processing unit is indistinguishable from a computer's Central Processing Unit (CPU), i.e. the hardware data processing unit is just a dedicated computer. However, the hardware data processing unit is normally dedicated to the specific computer application of format translation.

A software code compiler (e.g., for FORTRAN or ALGOL) is an example of a software data processing system. The software data processing system makes use of a (general purpose) computer in order to complete its functions. A software data processing system is normally a standalone unit of software, in that its output can be directed to any number of other (not necessarily as yet identified) information processing (sub) systems.

3.4 Information Repository

An information repository is an easy to deploy secondary tier of data storage that can comprise multiple, networked data storage technologies running on diverse operating systems, where data, that no longer needs to be in primary storage is protected, classified according to captured metadata, processed, de-duplicated, and then purged, automatically, based on data service level objectives and requirements. In information repositories, data storage resources are virtualised as composite storage sets and operate as a federated environment.

Information repositories were developed to mitigate problems arising from data proliferation and eliminate the need for separately deployed data storage solutions because of the concurrent deployment of diverse storage technologies running diverse operating systems. They feature centralised management for all deployed data storage resources. They are self-contained, support heterogeneous storage resources, support resource management to add, maintain, recycle, and terminate media, track of off-line media, and operate autonomously.

3.5 Automated Data Management

Since one of the main reasons for the implementation of an information repository is to reduce the maintenance workload placed on IT staff by traditional data storage systems, information repositories are automated. Automation is accomplished via polices that can process data based on time, events, data age, and data content. Policies manage the following:

- File system space management
- Irrelevant data elimination (mp3, games, etc.)
- Secondary storage resource management

Data is processed according to media type, storage pool, and storage technology.

Because information repositories are intended to reduce IT staff workload, they are designed to be easy to deploy and offer configuration flexibility, virtually limitless extensibility, redundancy, and reliable failover.

3.6 Data Recovery

Information repositories feature robust, client based data search and recovery capabilities that, based on permissions, enable end users to search the information repository, view information repository contents, including data on off-line media, and recover individual files or multiple files to either their original network computer or another network computer.

3.7 Backup

In information technology, backup refers to making copies of data so that these additional copies may be used to restore the original after a data loss event. These additional copies are typically called "backups." Backups are useful primarily for two purposes. The first is to restore a state following a disaster (called disaster recovery). The second is to restore small numbers of files after they have been accidentally deleted or corrupted.

Since a backup system contains at least one copy of all data worth saving, the data storage requirements are considerable. Organising this storage space and managing the backup process is a complicated undertaking. A data repository model can be used to provide structure to the storage. In the modern era of computing, there are many different types of data storage devices that are useful for making backups. There are also many different ways in which these devices can be arranged to provide geographic redundancy, data security, and portability.

Before data is sent to its storage location, it is selected, extracted, and manipulated. Many different techniques have been developed to optimise the backup procedure. These include optimisations for dealing with open files and live data sources as well as compression, encryption, and de-duplication, among others. Many organisations and individuals try to have confidence that the process is working as expected and work to define measurements and validation techniques. It is also important to recognise the limitations and human factors involved in any backup scheme.

Due to a considerable overlap in technology, backups and backup systems are frequently confused with archives and fault-tolerant systems. Backups differ from archives in the sense that archives are the *primary copy of data and backups are a secondary copy of data.* Archives are the primary copy of the item, usually put away for future use, while backups are the secondary copy, kept on hand to replace the original item. Backup systems differ from fault-tolerant systems in the sense that backup systems assume that a fault will cause a data loss event and fault-tolerant systems assume a fault will not.

3.8 Storage, the Base of a Backup System

Data Repository Models

Any backup strategy starts with a concept of a data repository. The backup data needs to be stored somehow and probably should be organised to a degree. It can be as simple as a sheet of paper with a list of all backup tapes and the dates they were written or a more sophisticated setup with a computerised index, catalogue, or relational database. Different repository models have different advantages. This is closely related to choosing a backup rotation scheme.

Unstructured

An unstructured repository may simply be a stack of floppy disks or CD-R/DVD-R media with minimal information about what was backed up and when. This is the easiest to implement, but probably the least likely to achieve a high level of recoverability.

Full + Incremental

A Full + Incremental repository aims to make storing several copies of the source data more feasible. At first, a full backup (of all files) is taken. After that an incremental backup (of only the files that have changed since the previous full or incremental backup) can be taken. Restoring whole systems to a certain point in time would require locating the full backup taken previous to that time and all the incremental backups taken between that full backup and the particular point in time to which the system is supposed to be restored. This model offers a high level of security that something can be restored and can be used with removable media such as tapes and optical disks. The downside is dealing with a long series of incremental and the high storage requirements.

Full + Differential

A full + differential backup differs from a full + incremental in that after the full backup is taken, each partial backup captures all files created or changed since the full backup, even though some may have been included in a previous partial backup. Its advantage is that a restore involves recovering only the last full backup and then overlaying it with the last differential backup. The downside would be that it takes more storage than the Full + Incremental.

Variable Dump Level

With variable dump levels (0-9), a 0 level dump is a full backup. Other levels backup all files that have changed since the last backup of a lower level. This allows planning strategies that achieve a compromise between the advantages of incrementals and differentials.

Mirror + Reverse Incrementals

A Mirror + Reverse Incrementals repository is similar to a Full + Incrementals repository. The difference is instead of an ageing full backup followed by a series of incrementals, this model offers a mirror that reflects the system state as of the last backup and a history of reverse incrementals. One benefit of this is that it only requires an initial full backup. Each incremental backup is immediately applied to the mirror and the files they replace are moved to a reverse incremental. This model is not suited to use removable media since every backup must be done in comparison to the mirror.

Continuous data protection

This model takes it a step further and instead of scheduling periodic backups, the system immediately logs every change on the host system. This is generally done by saving byte or block-level differences rather than file-level differences. It differs from simple disk mirroring in that it enables a roll-back of the log and thus restoring the old image of the data.

Storage media

Regardless of the repository model that is used, the data has to be stored on some data storage medium somewhere.

Magnetic tape

Magnetic tape has long been the most commonly used medium for bulk data storage, backup, archiving, and interchange. Tape has typically had an order of magnitude better capacity/price ratio when compared to hard disk, but recently the ratios for tape and hard disk have become a lot closer. There are myriad formats, many of which are proprietary or specific to certain markets like mainframes or a particular brand of personal computer. Tape is a sequential access medium, so, even though access times may be poor, the rate of continuously writing or reading data can actually be very fast. Some new tape drives are even faster than modern hard disks.

Hard disk

The capacity/price ratio of hard disk has been rapidly improving for many years. This is making it more competitive with magnetic tape as a bulk storage medium. The main advantages of hard disk storage are low access times, availability, capacity and ease of use. External disks can be connected via local interfaces like Small Computer System Interface (SCSI), USB, FireWire, or eSATA, or via longer distance technologies like Ethernet, iSCSI, or Fibre Channel. Some disk-based backup systems, such as Virtual Tape Libraries, support data de-duplication which can dramatically reduce the amount of disk storage capacity consumed by daily and weekly backup data.

Optical disc

A recordable CD can be used as a backup device. One advantage of CDs is that they can be restored on any machine with a CD-ROM drive. In addition, recordable CD's are relatively cheap. Another common format is recordable DVD. Many optical disk formats are Write Once Read Many (WORM) type, which makes them useful for archival purposes since the data can't be changed. Other rewritable formats can also be utilised such as CD-RW or DVD-RAM. The newer HD-DVDs and Blu-ray Discs dramatically increase the amount of data possible on a single optical storage disk, though, as yet, the hardware may be cost prohibitive for many people. Additionally, the physical lifetime of the optical disk has become a concern as it is possible for some optical disks to degrade and lose data within a couple years.

Floppy disk

During the 1980s and early 1990s, many personal/home computer users associated backup mostly with copying floppy disks. The low data capacity of a floppy disk makes it an unpopular and obsolete choice today.

Solid state storage

Also known as flash memory, thumb drives, USB flash drives, CompactFlash, SmartMedia, Memory Stick, Secure Digital cards, etc., these devices are relatively costly for their low capacity, but offer excellent portability and ease-of-use.

Remote backup service

As broadband internet access becomes more widespread, remote backup services are gaining in popularity. Backing up via the internet to a remote location can protect against some worst-case scenarios such as fires, floods, or earthquakes which would destroy any backups in the immediate vicinity along with everything else. A drawback to a remote backup service is that an internet connection is usually substantially slower than the speed of local data storage devices, so this can be a problem for people with large amounts of data. It also has the risk associated with putting control of personal or sensitive data in the hands of a third party.

Managing the data repository

Regardless of the data repository model or data storage media used for backups, a balance needs to be struck between accessibility, security and cost.

On-line

On-line backup storage is typically the most accessible type of data storage, which can begin restoring in milliseconds time. A good example would be an internal hard disk or a disk array (maybe connected to SAN: Secured Area Network). This type of storage is very convenient and speedy, but it is relatively expensive. On-line storage is vulnerable to being deleted or overwritten, either by accident, or in the wake of a data-deleting virus payload.

Near-line

Near-line storage is typically less accessible and less expensive than on-line storage, but still useful for backup data storage. A good example would be a tape library with restore times ranging from seconds to a few minutes. A mechanical device is usually involved in moving media units from storage into a drive where the data can be read or written.

Off-line

Off-line storage is similar to near-line, except it requires human interaction to make storage media available. This can be as simple as storing backup tapes in a file cabinet. Media access time can be anywhere from a few seconds to more than an hour.

Off-site vault

To protect against a disaster or other site-specific problem, many people choose to send backup media to an off-site vault. The vault can be as simple as the System Administrator's home office or as sophisticated as a disaster hardened, temperature controlled, high security bunker that has facilities for backup media storage.

Backup site, Disaster Recovery Centre or DR Centre

In the event of a disaster, the data on backup media will not be sufficient to recover. Computer systems onto which the data can be restored and properly configured networks are necessary too. Some organisations have their own data recovery centers that are equipped for this scenario. Other organisations contract this out to a third-party recovery center. Note that because DR site is itself a huge investment, backup is very rarely the considered or preferred method of moving data to DR site. More typical way would be remote disk mirroring, which keeps the DR data as up-to-date as possible.

SELF ASSESSMENT EXERCISE

Mention four other terms associated with solid state storage.

3.9 Selection, Extraction and Manipulation of Data

Selection and extraction of file data

Deciding what to back up at any given time is a harder process than it seems. By backing up too much redundant data, the data repository will fill up too quickly. Backing up an insufficient amount of data can eventually lead to the loss of critical information.

Copying files

Making copies of files is the simplest and most common way to perform a backup. A means to perform this basic function is included in all backup software and all operating systems.

Partial file copying

Instead of copying whole files, one can limit the backup to only the blocks or bytes within a file that have changed in a given period of time. This technique can use substantially less storage space on the backup medium, but requires a high level of sophistication to reconstruct files in a restore situation. Some implementations require integration with the source filesystem.

Filesystem dump

Instead of copying files within a filesystem, a copy of the whole filesystem itself can be made. This is also known as a raw partition *backup and is related to disk imaging. The process usually involves* unmounting the filesystem and running a program like dump. This type of backup has the possibility of running faster than a backup that simply copies files. A feature of some dump software is the ability to restore specific files from the dump image.

Identification of changes

Some filesystems have an archive bit for each file that says it was recently changed. Some backup software looks at the date of the file and compares it with the last backup, to determine whether the file was changed.

Versioning File System

A versioning file system keeps track of all changes to a file and makes those changes accessible to the user. Generally, this gives access to any previous version, all the way back to the file's creation time. An example of this is the Wayback versioning file system for Linux.

4.0 CONCLUSION

The processing of data to produce information is very vital in all endeavours. Businesses, institutions, governments, and homes are driven by information. Several developments are emerging especially in the information technology terrain that enhances data processing efforts.

5.0 SUMMARY

- Data processing is any computer process that converts data into information or knowledge.
- Data processing is defined as numbers or characters that represent measurements from observable phenomena.
- In order to be processed by a computer, the data needs first to be converted into a machine readable format
- In data processing or information processing, a Data Processor or Data Processing Unit or Data Processing System is a system which processes data which has been captured and encoded in a format recognisable by the data processing system or has been created and stored by another unit of an information processing system.
- An information repository is an easy to deploy secondary tier of data storage that can comprise multiple, networked data storage technologies running on diverse operating systems, where data that no longer needs to be in primary storage is protected, classified according to captured metadata, processed, de-duplicated, and then purged, automatically, based on data service level objectives and requirements
- Since one of the main reasons for the implementation of an Information repository is to reduce the maintenance workload placed on IT staff by traditional data storage systems, information repositories are automated
- In information technology, backup refers to making copies of data so that these additional copies may be used to restore the original after a data loss event. These additional copies are typically called “backups.”
- Any backup strategy starts with a concept of a data repository. The backup data needs to be stored somehow and probably should be organised to a degree.
- Deciding what to back up at any given time is a harder process than it seems. By backing up too much redundant data, the data repository will fill up too quickly. Backing up an insufficient amount of data can eventually lead to the loss of critical information.

6.0 TUTOR-MARKED ASSIGNMENT

1. Mention 10 elements of data processing.
2. Discuss briefly about optic disc.

7.0 REFERENCES/FURTHER READING

Linda B.; Bourgue Virginia, A. C. (2006). Processing Data: The Survey *Example (Quantitative Applications in the Social Sciences).Sage* Publications, Inc.

NGDC Conference: Understand advanced IT infrastructures, Protecting Information: Benefits of a Federated Information Repository as a Secondary Storage Tier.

SNIA Enterprise Information World Conference: (2007). Benefits of a Federated Information Repository as a Secondary Storage Tier.

UNIT 2 DATA ARCHIVING

CONTENTS

1.0 INTRODUCTION

An archive refers to a collection of historical records, and also refers to the location in which these records are kept.

Archives are made up of records (a.k.a. primary source documents) which have been accumulated over the course of an individual or organisation's lifetime. For example, the archives of an individual may contain letters, papers, photographs, computer files, scrapbooks, financial records, diaries or any other kind of documentary materials created or collected by the individual--regardless of media or format. The archives of an organisation (such as a corporation or government), on the other hand, tend to contain different types of records, such as administrative files, business records, memos, official correspondence, meetings' minutes, and so on.

In general, archives of any individual or organisation consist of records which have been especially selected for permanent or long-term preservation, due to their enduring research value. Archival records are normally unpublished and almost always unique, unlike books or magazines, in which many identical copies exist. This means that archives (the places) are quite distinct from libraries with regard to their functions and organisation, although archival collections can often be found within library buildings.

Archives are sometimes described as information generated as the "by-product" of normal human activities, while libraries hold specifically authored information "products".

A person who works in archives is called an archivist. The study and practice of organising, preserving, and providing access to information and materials in archives is called archival science.

Archivists tend to prefer the term 'archives' (with an S) as the correct terminology to serve as both the singular and plural, since 'archive,' as a noun or a verb, has meanings related to computer science.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- explain the concept of data archiving
- outline the history and development of archiving
- differentiate archives from library
- identify the main users of archiving systems
- describe the concept of digital archiving
- identify the limits and difficulties faced in web archiving.

3.0 MAIN CONTENT

3.1 Digital Archives

Archives differ from libraries in several ways. Traditionally, archives were defined as:

- Containing primary sources of information (typically letters and papers directly produced by an individual or organisation) rather than the secondary sources found in a library (books, etc)
- Having their contents organised in groups rather than individual items. Whereas books in a library are catalogued individually, items in an archive are typically grouped by provenance (the individual or organisation that created them) and original order (the order in which the materials were kept by the creator).
- Having unique contents. Whereas a book may be found at many different libraries, depending on its rarity, the records in an archive are usually one-of-a-kind, and cannot be found or consulted at any other location except at the archive that holds them.

The technology used to create digital libraries has been even more revolutionary for archives since it breaks down the second and third of these general rules. The use of search engines, Optical Character Recognition and metadata allow digital copies of individual items (i.e. letters) to be catalogued, and the ability to remotely access digital copies has removed the necessity of physically going to a particular archive to find a particular set of records. The Oxford Text Archive is generally considered to be the oldest digital archive of academic primary source materials.

Project Gutenberg, Google Book Search, Windows Live Search Books, Internet Archive, Cornell University, The Library of Congress World Digital Library, The Digital Library at the University of Michigan, and CMU's Universal library are considered leaders in the field of digital archive creation and management. There are hundreds of regionals such as the Wisconsin Historical Society. The Vatican maintains an extensive digital library inventory and associated technology. The Packard Foundation maintains digitisation facilities near the Acropolis in Athens, Greece, as examples.

3.2 Archive Users and Institutions

Historians, genealogists, lawyers, demographers, and others conduct research at archives. The research process at each archive is unique, and depends upon the institution in which the archive is housed. While there are many different kinds of archives, the most recent census of archivists in the United States identified five major types: academic, for profit (business), government, nonprofit, and other.

Academic Archives

Archives existing in colleges, universities, or other educational facilities are usually grouped as academic archives. Academic archives typically exist within a library, and duties may be carried out by an archivist or a librarian. Occasionally, history professors may run a smaller academic archive. Academic archives exist to celebrate and preserve the history of their school and academic community. The inventory of an academic archive may contain items such as papers of former professors and presidents, memorabilia related to school organisations and activities, and items the academic library wishes to remain in a closed-stack setting, such as rare books or thesis copies. It is always a good idea to contact an academic archive before visiting, as the majority of these institutions are available by appointment only. Users of academic archives are often graduate students and those wishing to view rare or historical documents for research purposes. Many academic archives work closely with alumni relations to help raise funds for their library or school. Because of their library setting, a degree certified by the American Library Association is preferred for employment in an academic archive in the USA.

Business (For Profit) Archives

Archives located in for-profit institutions are usually those owned by a private business. Examples of prominent business archives in the United States include Coca-Cola (which also owns the separate museum World of Coca-Cola), Proctor and Gamble, Motorola Heritage Services and Archives, and Levi Strauss & Co... These corporate archives maintain historic documents and items related to the history of their companies. Business archives serve the purpose of helping their corporations maintain control over their brand by retaining memories of the company's past, especially in business archives; records management is separate from the historic aspect of archives. Workers in these types of archives may have any combination of training and degrees, from either a history or library background. These archives are typically not open to the public and only used by workers of the owner company, although some will allow approved visitors by appointment. Business archives are concerned with maintaining the integrity of their parent company, and therefore selective of how their materials may be used.

Government Archives

The category of government archives includes those institutions run on a local and state level as well as those run by the national (federal) government. Anyone may use a government archive, and frequent users include reporters, genealogists, writers, historians, students, and anyone wanting more information on the history of their home or region. While it is a good idea to make an appointment before visiting government archives, at many government archives no appointment is required, as they are open to the public.

In the United States, NARA archives exist not only in the District of Columbia, but regionally as well. Some city or local governments may have repositories, but their organisation and accessibility varies widely. State or Province archives typically require at least a bachelor's degree in history for employment, although some ask for certification by test (government or association) as well.

In the UK, the National Archives, formerly known as the Public Record Office, is the government archive for England and Wales. The National Monuments Record is the public archive of English Heritage. The National Archives of Scotland, located in Edinburgh, serve that country while the Public Record Office of Northern Ireland in Belfast is the government archive for Northern Ireland.

A network of local authority-run record offices and archives exists throughout England, Wales and Scotland and holds many important collections, including local government, landed estates, church and business records. Many archives have contributed catalogues to the national Access to Archives program and online searching across collections is possible.

In France, the Directorate of the Archives of France (Direction des *Archives de France) in the Ministry of Culture manages the National* Archives (Archives nationales) which possess 364 km. (226 miles) of archives as of 2004 (the total length of occupied shelves put next to each other), with original records going as far back as A.D. 625, as well as the departmental archives (archives départementales), located in the *préfectures,of each of the 100 départements of France, which possess* 1,901 km (1,181 miles) of archives (as of 2004), and also the local city archives, about 600 in total, which possess 449 km (279 miles) of archives (as of 2004). Put together, the total volume of archives under the supervision of the Directorate of the Archives of France is the largest in the world, a testimony to the very ancient nature of the French state which has been in existence for more than eleven centuries already.

- In India, the National Archives are located in New Delhi.
- In Taiwan, the National Archives Administration is located in Taipei.

Most intergovernmental organisations keep their own historical archives. However, a number of European organisations, including the European Commission, choose to deposit their archives with the European University Institute in Florence.

Church's Archives

Include Archdioceses, dioceses and parishes archives in the Roman Catholic and Episcopalian Churches and other kinds of archives in the different denominations of the Christian Churches. The most famous Church's Archive is the Vatican Secret Archive.

Non-Profit Archives

Non-profit archives include those in historical societies, not-for-profit businesses such as hospitals, and the repositories within foundations. Non-profit archives are typically set up with private funds from donors to preserve the papers and history of specific persons or places. Often, these institutions rely on grant funding from the government as well. Depending on the funds available, non-profit archives may be as small as the historical society in a rural town to as big as a state historical society that rivals a government archives. Users of this type of archive may vary as much as the institutions that hold them. Employees of non-profit archives may be professional archivists, para-professionals, or volunteers, as the education required for a position at a non-profit archive varies with the demands of the collection's user base.

Special (Other) Archives

Some archives defy categorisation. There are tribal archives within the Native American nations in North America, and there are archives that exist within the papers of private individuals. Many museums keep archives in order to prove the provenance of their pieces. Any institution or persons wishing to keep their significant papers in an organised fashion that employs the most basic principles of archival science may have an archive. In the 2004 census of archivists taken in the United States, 2.7% of archivists were employed in institutions that defied categorisation. This was a separate figure from the 1.3% that identified themselves as self-employed.

3.3 Etymology

The word archive is derived from the Greek (arkh) meaning government or order (compare an-archy, mon-archy). The word originally developed from the Greek (arkheion) which refers to the home or dwelling of the Archon, in which important official state documents were filed and interpreted under the authority of the Archon.

3.4 Archives in History

The word “archives” can refer to any organised body of records fixed on media. The management of archives is essential for effective day-to-day organisational decision making, and even for the survival of organisations. Archives were well developed by the ancient Chinese, the ancient Greeks, and ancient Romans. Modern archival thinking has many roots in the French Revolution. The French National Archives, which possess perhaps the largest archival collection in the world, with records going as far back as A.D. 625, were created in 1790 during the French Revolution from various government, religious, and private archives seized by the revolutionaries.

SELF ASSESSMENT EXERCISE

Briefly define business archives and mention some examples.

3.5 Web Archiving

3.5.1 Introduction

Web archiving is the process of collecting portions of the World Wide Web and ensuring the collection is preserved in an archive, such as an archive site, for future researchers, historians, and the public. Due to the massive size of the Web, web archivists typically employ web crawlers for automated collection. The largest web archiving organisation based on a crawling approach is the Internet Archive which strives to maintain an archive of the entire Web. National libraries, national archives and various consortia of organisations are also involved in archiving culturally important Web content. Commercial web archiving software and services are also available to organisations which need to archive their own web content for legal or regulatory purposes.

3.5.2 Collecting the Web

Web archivists generally archive all types of web content including Hyper Text Markup Language (HTML) web pages, style sheets, JavaScript, images, and video. They also archive metadata about the collected resources such as access time, MIME type, and content length. This metadata is useful in establishing authenticity and provenance of the archived collection.

3.5.3 Methods of Collection

Remote harvesting

The most common web archiving technique uses web crawlers to automate the process of collecting web pages. Web crawlers typically view web pages in the same manner that users with a browser see the Web, and therefore provide a comparatively simple method of remotely harvesting web content. Examples of web crawlers frequently used for web archiving include:

- Heritrix
- HTTrack
- Offline Explorer
- Web Curator

On-demand

There are numerous services that may be used to archive web resources "on-demand", using web crawling techniques:

- WebCite, a service specifically for scholarly authors, journal editors and publishers to permanently archive and retrieve cited Internet references (Eysenbach and Trudel, 2005).
- Archive-It, a subscription service, allows institutions to build, manage and search their own web archive.
- Hanzo Archives offer commercial web archiving tools and services, implementing an archive policy for web content and enabling electronic discovery, litigation support or regulatory compliance.

Database archiving

Database archiving refers to methods for archiving the underlying content of database-driven websites. It typically requires the extraction of the database content into a standard schema, often using XML. Once stored in that standard format, the archived content of multiple databases can then be made available using a single access system. This approach is exemplified by the DeepArc and Xinq tools developed by the Bibliothèque nationale de France and the National Library of Australia respectively. DeepArc enables the structure of a relational database to be mapped to an XML schema, and the content exported into an XML document. Xinq then allows that content to be delivered online. Although the original layout and behavior of the website cannot be preserved exactly, Xinq does allow the basic querying and retrieval functionality to be replicated.

Transactional archiving

Transactional archiving is an event-driven approach, which collects the actual transactions which take place between a web server and a web browser. It is primarily used as a means of preserving evidence of the content which was actually viewed on a particular website, on a given date. This may be particularly important for organisations which need to comply with legal or regulatory requirements for disclosing and retaining information.

A transactional archiving system typically operates by intercepting every HTTP request to, and response from, the web server, filtering each response to eliminate duplicate content, and permanently storing the responses as bitstreams. A transactional archiving system requires the installation of software on the web server, and cannot therefore be used to collect content from a remote website.

Examples of commercial transactional archiving software include:

- PageVault
- Vignette WebCapture

3.5.4 Difficulties and Limitations

Crawlers

Web archives, which rely on web crawling as their primary means of collecting the Web are influenced by the difficulties of web crawling:

- the robots exclusion protocol may request crawlers not access portions of a website. Some web archivists may ignore the request and crawl those portions anyway.
- large portions of a web site may be hidden in the Deep Web. For example, the results page behind a web form lies in the deep web because a crawler cannot follow a link to the results page.
- some web servers may return a different page for a web crawler than it would for a regular browser request. This is typically done to fool search engines into sending more traffic to a website.
- crawler traps (e.g., calendars) may cause a crawler to download an infinite number of pages, so crawlers are usually configured to limit the number of dynamic pages they crawl.

The Web is so large that crawling a significant portion of it takes a large amount of technical resources. The Web is changing so fast that portions of a website may change before a crawler has even finished crawling it.

General limitations

Not only must web archivists deal with the technical challenges of web archiving, they must also contend with intellectual property laws. Peter Lyman (2002) states that "although the Web is popularly regarded as a public domain resource, it is copyrighted; thus, archivists have no legal right to copy the Web". However national libraries in many countries do have a legal right to copy portions of the web under an extension of a legal deposit.

Some private non-profit web archives that are made publicly accessible like WebCite or the Internet Archive allow content owners to hide or remove archived content that they do not want the public to have access to. Other web archives are only accessible from certain locations or have regulated usage. WebCite also cites on its FAQ a recent lawsuit against the caching mechanism, which Google won.

3.5.5 Aspects of Web Curation

Web curation, like any digital curation, entails:

- collecting verifiable Web assets

- providing Web asset search and retrieval
- certification of the trustworthiness and integrity of the collection content
- semantic and ontological continuity and comparability of the collection content

Thus, besides the discussion on methods of collecting the web, those of providing access, certification, and organising must be included. There are a set of popular tools that addresses these curation steps:

A suit of tools for Web Curation by International Internet Preservation Consortium:

- Heritrix - official website - collecting Web asset
- NutchWAX - search Web archive collections
- Wayback (Open source Wayback Machine) - search and navigate Web archive collections using NutchWax
- Web Curator Tool - Selection and Management of Web Collection

Other open source tools for manipulating web archives:

- WARC Tools - for creating, reading, parsing and manipulating, web archives programmatically
- Search Tools - for indexing and searching full-text and metadata within web archives

4.0 CONCLUSION

Archiving, as a long- standing method of preserving historical materials, has helped in no small measure to preserve national and institutional heritages. This art is undergoing modernisation as we now have electronic form of archiving. It gives a document and materials much longer life span and also makes such document and material to be easily accessed.

5.0 SUMMARY

- An archive refers to a collection of historical records, and also refers to the location in which these records are kept.
- Archives differ from libraries in several ways.
- Historians, genealogists, lawyers, demographers, and others conduct research at archives. The research process at each archive is unique, and depends upon the institution in which the archive is housed.

- Archives located in for-profit institutions are usually those owned by a private business, Examples of prominent business archives in the United States include Coca-Cola (which also owns the separate museum World of Coca-Cola), Proctor and Gamble, Motorola Heritage Services and Archives, and Levi Strauss & Co.
- The word "archives" can refer to any organised body of records fixed on media. The management of archives is essential for effective day-to-day organisational decision making, and even for the survival of organisations.
- Web archiving is the process of collecting portions of the World Wide Web and ensuring the collection is preserved in an archive, such as an archive site, for future researchers, historians, and the public.

6.0 TUTOR-MARKED ASSIGNMENT

1. Briefly discuss non-profit archives.
2. Give four examples of web crawlers frequently used in web archiving

7.0 REFERENCES/FURTHER READING

"A Glossary of Archival and Records Terminology". Society of American Archivists.

"Business Archives Council". Business Archives Council.

"Business Archives in North America - Invest in your future: Understand your past". Society of American Archivists.

"Definition of archive". Wikitionary.

"Directions for Change". Libraries and Archives Canada.

"Directory of Corporate Archives". Hunter Information Management.

"Glossary of Library and Internet Terms". University of South Dakota Library.

"Guidelines for College and University Archives". Society of American Archivists.

"The National Archives". United States National Archives and Records Administration.

"U.S. - State Level Records Repositories: State Libraries, Archives, Genealogical & Historical Societies". Cyndi's List of Genealogy Sites on the Internet.

"Welcome to University Archives and Records Management". Kennesaw State University Archives

"What Are Archives?" National Museum of American History

(French) "Les archives en France". Quid – 2007

Brown, A. (2006). Archiving Websites: A Practical Guide for Information Management Professionals. Facet Publishing.

Brügger, N. (2005). Archiving Websites. General Considerations and Strategies. The Centre for Internet Research.

Creigh, by Dorothy Weyer (1995). A Primer for Local Historical Societies: Revised and Expanded from the First Edition. AltaMira *Press, 122.*

Day, M. (2003). "Preserving the Fabric of Our Lives: A Survey of Web Preservation Initiatives". Research and Advanced Technology for *Digital Libraries: Proceedings of the 7th European Conference (ECDL): 461–472.*

Eysenbach, G. and Trudel, M. (2005). "Going, going, still there: using the WebCite Service to Permanently archive Cited web Pages". *Journal of Medical Internet Research 7 (5). doi:10.2196/jmir.7.5.e60.*

Fitch, Kent (2003). "Web site Archiving - an Approach to Recording every Materially different Response Produced by a Website". *Ausweb 03.*

International Organisation for Standardisation "ISO/DIS 11506 Document management applications -- Archiving of electronic data -- Computer output microform (COM) / Computer output laser disc (COLD

International Organisation for Standardisation. "ISO/NP TS 21547-1 Health informatics -- Secure Archiving of Electronic Health Records -- Part 1: Principles and Requirements".

Lyman, P. (2002). "Archiving the World Wide Web". Building a National Strategy for Preservation: Issues in Digital Media Archiving.

Maher, William J. (1992). The Management of College and University Archives. Metuchen, New Jersey: Society of American Archivists *& The Scarecrow Press, Inc.*

Masanès, J. (ed.) (2006). Web Archiving. Springer-Verlag. ISB N 3-540-23338-5.

*Walch, Victoria Irons (2006). "A*Census: A Closer Look". The* American Archivist 69 (2): 327–348.

Walch, Victoria Irons (2006). "Archival Census and Education Needs Survey in the United States: Part 1: Introduction". The American Archivist 69 (2): 294–309.

Whitehill, Walter Muir (1962). "Introduction", Independent Historical Societies: An Enquiry into Their Research and Publication Functions and Their Financial Future. Boston, Massachusetts: *The Boston Athenaeum, 311.*

UNIT 3 DIGITAL ASSET MANAGEMENT

CONTENTS

1.0 INTRODUCTION

Digital preservation is the management of digital information over time. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media. This constant input of effort, time, and money to handle rapid technological and organisational advance is considered the main stumbling block for preserving digital information. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost, creating a digital Dark Age.

Digital preservation can therefore be seen as the set of processes and activities that ensure continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. This includes the preservation of materials resulting from digital reformatting, but particularly information that is born-digital and has no analogue counterpart. In the language of digital imaging and electronic resources, preservation is no longer just the product of a program but an ongoing process. In this regard, the way digital information is stored is important in ensuring their longevity. The long-term storage of digital information is assisted by the inclusion of preservation metadata.

Digital preservation is defined as: long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for. Long-term is defined as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long-Term may extend indefinitely". "Retrieval" means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. "Interpretation" means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded and transformed into usable representations. This is often interpreted as "rendering", i.e. making it available for a human to access. However, in many cases, it will mean able to be processed by computational means.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define digital preservation
- explain why it is necessary to engage in digital preservation
- identify the strategies to embark on in executing digital preservation
- describe how to sustain digital preservation
- list the standards that guide digital preservation.

3.0 MAIN CONTENT

3.1 Why Active Preservation is Necessary

Society's heritage has been presented on many different materials, including stone, vellum, bamboo, silk, paper and etc. Now, a large quantity of information exists in digital forms, including emails, blogs, social networking websites, national elections websites, web photo albums, and sites which change their content over time. According to a report by the US Library of Congress, 44% of the sites available on the internet in 1998 had vanished one year later.

The unique characteristic of digital forms makes it easy to create content and keep it up-to-date, but at the same time brings many difficulties in the preservation of this content. Margaret Hedstrom points out that "digital preservation raises challenges of a fundamentally different nature which are added to the problems of preserving traditional format materials."

Physical deterioration

The first challenge digital preservation faces is that the media on which digital contents stand are more vulnerable to deterioration and catastrophic loss. While acid paper are prone to deterioration in terms of brittleness and yellowness, the deterioration does not become apparent in at least six decades; and when the deterioration really happens, it happens over decades too. It is also highly possible to retrieve all information without loss after deterioration is spotted. The recording media for digital data deteriorate at a much more rapid pace, and once the deterioration starts, in most cases there is already data loss. This characteristic of digital forms leaves a very short time frame for preservation decisions and actions.

Digital obsolescence

Another challenge, perhaps a more serious and important one, is the problem of long-term access. Digital technology is developing extremely fast, and one retrieval and playback technology can become obsolete in a matter of years. When faster, more capable and cheaper storage and processing devices are developed, the older version gets replaced almost immediately. When a software or decoding technology is abandoned, or a hardware device is no longer in production, records created under the environment of such technologies are at great risk of loss, simply because they are not tangible any more.
This process is known as digital obsolescence.

This challenge is exacerbated by the lack of established standards, protocols, and proven methods for preserving digital information. We used to save copies of data on tapes, but media standards for tapes have changed considerably over the last five to ten years, and there is no guarantee that tapes will be readable in the future. Hedstrom further explained that almost all digital library researches have been focused on "architectures and systems for information organisation and retrieval, presentation and visualisation, and administration of intellectual property rights" and that "digital preservation remains largely experimental and replete with the risks associated with untested methods". While the rapid advance of technology threats access of digital contents in length, the lack of digitising standards affects the issue in width.

3.2 Strategies

In 2006, the Online Computer Library Centre (OCLC) developed a four-point strategy for the long-term preservation of digital objects that consisted of:

- assessing the risks for loss of content posed by technology variables such as commonly used proprietary file formats and software applications.

- evaluating the digital content objects to determine what type and degree of format conversion or other preservation actions should be applied.
- determining the appropriate metadata needed for each object type and how it is associated with the objects.
- providing access to the content.

There are several additional strategies that individuals and organisations may use to actively combat the loss of digital information.

Refreshing

Refreshing is the transfer of data between two types of the same storage medium so there are no bitrate changes or alteration of data. For example, consider transferring census data from a gold preservation CD to a new one. This strategy may need to be combined with migration when the software or hardware required to read the data is no longer available or is unable to understand the format of the data. Refreshing will likely always be necessary due to the deterioration of physical media.

Migration

Migration is the transferring of data to newer system environments (Garrett et al., 1996). This may include conversion of resources from one format to another (e.g., conversion of Microsoft Word to PDF or Open Document), from one operating system to another (e.g., Solaris to Linux) or from one programming language to another (e.g., C to Java) so the resource remains fully accessible and functional. Resources that are migrated run the risk of losing some type of functionality since newer formats may be incapable of capturing all the functionality of the original format, or the converter itself may be unable to interpret all the nuances of the original format. The latter is often a concern with proprietary data formats.

The National Archives Electronic Records Archives and Lockheed Martin are jointly developing a migration system that will preserve any type of document, created on any application or platform, and delivered to the archives on any type of digital media (Reagan, 2006). In the system, files are translated into flexible formats, such as XML; they will therefore be accessible by technologies in the future (Reagan, 2006). Lockheed Martin argues that it would be impossible to develop an emulation system for the National Archives ERA because the volume of records and cost would be prohibitive (Reagan, 2006).

Replication

Creating duplicate copies of data on one or more systems is called *replication. Data that exists as a single copy in only one location is* highly vulnerable to software or hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places.

Emulation

Emulation is the replicating of functionality of an obsolete system (Rothenberg, 1998). For example, emulating an Atari 2600 on a Windows system or emulating WordPerfect 1.0 on a Macintosh. Emulators may be built for applications, operating systems, or hardware platforms. Emulation has been a popular strategy for retaining the functionality of old video game systems, such as with the MAME project. The feasibility of emulation as a catch-all solution has been debated in the academic community (Granger, 2000).

Raymond A. Lorie has suggested that a Universal Virtual Computer (UVC) could be used to run any software in the future on a yet unknown platform (Lorie, 2001). The UVC strategy uses a combination of emulation and migration. The UVC strategy has not yet been widely adopted by the digital preservation community.

Jeff Rothenberg, a major proponent of Emulation for digital preservation in libraries, working in partnership with Koninklijke Bibliotheek and National Archief of the Netherlands, has recently helped launch Dioscuri, a modular emulator that succeeds in running MS-DOS, WordPerfect 5.1, DOS games, and more (Hoeven, 2007).

Metadata attachment

Metadata is data on a digital file that includes information on creation, access rights, restrictions, preservation history, and rights management. Metadata attached to digital files may be affected by file format obsolescence. ASCII is considered to be the most durable format for metadata because it is widespread, backwards compatible when used with Unicode, and utilises human-readable characters, not numeric codes. It retains information, but not the structure information is presented in. For higher functionality, SGML or XML should be used. Both markup languages are stored in ASCII format, but contain tags that denote structure and format.

Trustworthy digital objects

Digital objects that can speak to their own authenticity are called *trustworthy digital objects (TDOs). TDOs were proposed by Henry M.* Gladney to enable digital objects to maintain a record of their change history so future users can know with certainty that the contents of the object are authentic (Gladney, 2004). Other preservation strategies like replication and migration are necessary for the long-term preservation of TDOs.

3.3 Digital Sustainability

Digital sustainability encompasses a range of issues and concerns that contribute to the longevity of digital information. Unlike traditional, temporary strategies and more permanent solutions, digital sustainability implies a more active and continuous process. Digital sustainability concentrates less on the solution and technology and more on building an infrastructure and approach that is flexible, with an emphasis on interoperability, continued maintenance and continuous development. Digital sustainability incorporates activities in the present that will facilitate access and availability in the future.

3.4 Digital Preservation Standards

To standardise digital preservation practice and provide a set of recommendations for preservation program implementation, the Reference Model for an Open Archival Information System (OAIS) was developed. The reference model (ISO 14721:2003) includes the following responsibilities that an OAIS archive must abide by.

- Negotiate for and accept appropriate information from information producers.
- Obtain sufficient control of the information provided to the level needed to ensure long-term preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the designated community and, therefore, should be able to understand the information provided.
- Ensure that the information to be preserved is independently understandable to the designated community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as

authenticated copies of the original, or as traceable to the original.

- Make the preserved information available to the designated community.

OAIS is concerned with all technical aspects of a digital object's life cycle: ingest into and storage in a preservation infrastructure, data management, accessibility, and distribution. The model also addresses metadata issues and recommends that five types of metadata be attached to a digital object: reference (identification) information, provenance (including preservation history), context, fixity (authenticity indicators), and representation (formatting, file structure, and what "imparts meaning to an object's bit stream". Prior to Gladney's proposal of TDOs was the Research Library Group's (RLG) development of "attributes and responsibilities" that denote the practices of a "Trusted Digital Repository" (TDR). The seven attributes of a TDR are: "compliance with the Reference Model for an Open Archival Information System (OAIS), Administrative responsibility, Organisational viability, Financial sustainability, Technological and Procedural suitability, System security, Procedural accountability." Among RLG's attributes and responsibilities were recommendations calling for the collaborative development of digital repository certifications, models for cooperative networks, and sharing of research and information on digital preservation with regards to intellectual property rights.

3.5 Digital Sound Preservation Standards

In January 2004, the Council on Library and Information Resources (CLIR) hosted a roundtable meeting of audio experts discussing best practices, which culminated in a report delivered in March 2006. This report investigated procedures for reformatting sound from analogue to digital, summarising discussions and recommendations for best practices for digital preservation. Participants made a series of 15 recommendations for improving the practice of analogue audio transfer for archiving:

- develop core competencies in audio preservation engineering. Participants noted with concern that the number of experts qualified to transfer older recordings is shrinking and emphasised the need to find a way to ensure that the technical knowledge of these experts can be passed on.
- develop arrangements among smaller institutions that allow for cooperative buying of esoteric materials and supplies.
- pursue a research agenda for magnetic-tape problems that focuses on a less destructive solution for hydrolysis than baking, relubrication of acetate tapes, and curing of cupping.

- develop guidelines for the use of automated transfer of analogue audio to digital preservation copies.
- develop a web-based clearinghouse for sharing information on how archives can develop digital preservation transfer programs.
- carry out further research into nondestructive playback of broken audio discs.
- develop a flowchart for identifying the composition of various types of audio discs and tapes.
- Develop a reference chart of problematic media issues.
- collate relevant audio engineering standards from organisations.
- research safe and effective methods for cleaning analogue tapes and discs.
- develop a list of music experts who could be consulted for advice on transfer of specific types of musical content (e.g., determining the proper key so that correct playback speed can be established).
- research the life expectancy of various audio formats.
- establish regional digital audio repositories.
- cooperate to develop a common vocabulary within the field of audio preservation.
- investigate the transfer of technology from such fields as chemistry and materials science to various problems in audio preservation.

3.6 Large-Scale Digital Preservation Initiatives (LSDIS)

Many research libraries and archives have begun or are about to begin Large-Scale digital preservation initiatives (LSDI's). The main players in LSDIs are cultural institutions, commercial companies such as Google and Microsoft, and non-profit groups including the Open Content Alliance (OCA) and the Million Book Project (MBP). The primary motivation of these groups is to expand access to scholarly resources.

LSDIs: library perspective

Approximately 30 cultural entities, including the 12-member Committee on Institutional Cooperation (CIC), have signed digitisation agreements with either Google or Microsoft. Several of these cultural entities are participating in the Open Content Alliance (OCA) and the Million Book Project (MBP). Some libraries are involved in only one initiative and others have diversified their digitisation strategies through participation in multiple initiatives. The three main reasons for library participation in LSDIs are: Access, Preservation and Research and Development. It is hoped that digital preservation will ensure that library materials remain accessible for future generations. Libraries have a perpetual responsibility for their materials and a commitment to archive their digital materials. Libraries plan to use digitised copies as backups for works in case they go out of print, deteriorate, or are lost and damaged.

4.0 CONCLUSION

Digital management of data, that is, digital preservation is the trend in document preservation. It gives a document a much longer life span and also makes such document to be easily assessed. Lots of standards established are all aimed at making the art of digital preservation effective and efficient.

5.0 SUMMARY

- Digital preservation is the management of digital information over time. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media.
- Society's heritage has been presented on many different materials, including stone, vellum, bamboo, silk, paper and etc. Now, a large quantity of information exists in digital forms, including emails, blogs, social networking websites, national elections websites, web photo albums, and sites which change their content over time.
- in 2006, the Online Computer Library Center (OCLC) developed a four-point strategy for the long-term preservation of digital objects
- Digital sustainability encompasses a range of issues and concerns that contribute to the longevity of digital information. Unlike traditional, temporary strategies and more permanent solutions, digital sustainability implies a more active and continuous process.
- To standardise digital preservation practice and provide a set of recommendations for preservation program implementation, the Reference Model for an Open Archival Information System (OAIS) was developed.
- In January 2004, the Council on Library and Information Resources (CLIR) hosted a roundtable meeting of audio experts discussing best practices, which culminated in a report delivered March 2006.
- Many research libraries and archives have begun or are about to begin Large-Scale digital preservation initiatives (LSDI's).

6.0 TUTOR-MARKED ASSIGNMENT

Mention 10 recommendations to improve the practice of analogue audio transfer for archiving.

7.0 REFERENCES/FURTHER READING

"Lifecycle Information for E-literature". LIFE.

"U.S. Congress Approves Library of Congress Plan for Preservation of Digital Materials", Library of Congress, 2003-02-07.

BBC (2008). Writing the history of virtual worlds.

Bradley, K. (Summer 2007). Defining Digital Sustainability. Library *Trends v. 56 no 1 p. 148-163.*

Consultative Committee for Space Data Systems. (2002). Reference *Model for an Open Archival Information System (OAIS).* Washington, DC: CCSDS Secretariat, p. 1-1.

Consultative Committee for Space Data Systems. (2002). Reference *Model for an Open Archival Information System (OAIS).* Washington, DC: CCSDS Secretariat, p. 3-1.

Cornell University Library. (2005). Digital Preservation Management: *Implementing Short-term Strategies for Long-term Problems.*

Cornell University Library. (2005). Digital Preservation Management: *Implementing Short-term Strategies for Long-term Problems*

Council on Library and Information Resources. Publication 137: *Capturing Analogue Sound for Digital Preservation: Report of a Roundtable Discussion of Best Practices for Transferring Analogue Discs and Tapes March 2006.*

Flugstad, M. (2007). Website Archiving: The Long-Term Preservation of *Local Born Digital Resources. Arkansas Libraries v. 64 no. 3* (Fall 2007) p. 5-7.

Gladney, H. M. (2004). "Trustworthy 100-Year Digital Objects: Evidence after every Witness is dead". ACM Transactions on Information Systems 22 (3): 406–436.

Hedstrom, M. (1997). Digital Preservation: A Time Bomb for Digital *Libraries.*

Levy, D. M. & Marshall, C. C. (1995). Going Digital: A Look at *Assumptions Underlying Digital Libraries,” Communications of* the ACM, 58, No. 4: 77-84.

National Initiative for a Networked Cultural Heritage (2002). NINCH *Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials.*

NISO Framework Advisory Group (2007). A Framework of Guidance *for Building Good Digital Collections, 3rd edition, p. 57.*

Online Computer Library Center, Inc. (2006). OCLC Digital Archive *Preservation Policy and Supporting Documentation, p. 5*

Research Libraries Group (2002). [Trusted Digital Repositories: *Attributes and Responsibilities]*

Sustainability of Digital Resources (2008). TASI: Technical Advisory *Service for Images.*

UNIT 4 DIGITAL ASSET MANAGEMENT

CONTENTS

1.0 INTRODUCTION

Digital Asset Management consists of tasks and decisions surrounding ingesting, annotating, cataloguing, storage and retrieval of digital assets, such as digital photographs, animations, videos and music. Digital asset management systems are computer software and/or hardware systems that aid in the process of digital asset management.

The term "Digital Asset Management" (DAM) also refers to the protocol for downloading, renaming, backing up, rating, grouping, archiving, optimising, maintaining, thinning, and exporting files. The term "Media Asset Management" (MAM) is sometimes used as a sub-category of "Digital Asset Management", mainly for audio or video content. The more recent concept of Enterprise Content Management (ECM) often describes solutions which address similar features but in a wider range of industries or applications.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define digital asset management
- list the uses of digital asset management in organisation
- differentiate the various types of digital asset management
- describe complementary concepts of digital asset and document imaging.

3.0 MAIN CONTENT

3.1 Uses

Many businesses and organisations are adopting Digital Asset Management as a business strategy because managing image, video and other media assets presents unique challenges and requires solutions designed specifically to streamline the acquisition, storage and retrieval of digital media. Effective implementation of a DAM system should reduce the time and cost of content production, maximise the return on investment (ROI) from media assets, bring new products and services to market faster and streamline compliance. This system should be designed in such a way that enables cost-effective optimisation of media asset management across an organisation.

3.2 Technical Context

Generally, the "asset" being managed is collected and stored in a digital format. There is usually a target version of that referred to as "essence" and is generally the highest resolution and fidelity representation. The asset is detailed by its "metadata". Metadata is the description of the asset and the description depth can vary depending on the needs of the system, designer, or user. Metadata can describe, but is not limited to, the description of: asset content (what is in the package?); the means of encoding/decoding (e.g. JPEG, tar, MPEG 2); provenance (history to point of capture); ownership; rights of access; as well as many others. There exist some predefined standards and template for metadata such as Dublin Core and PB Core. In cases of systems that contain large size asset essences, such as MPEG 2 and JPEG2000 for the cases of images and video, there are usually related "proxy" copies of the essence. A proxy copy is a lower resolution representation of the essence that can be used as a reference in order to reduce the overall bandwidth requirements of the DAM system infrastructure. It can be generated and retained at the time of ingestion of the asset simultaneous or subsequent to the essence, or it can be generated on the fly using transcoders.

Smaller DAM systems are easier to categorise as to content and usage since they would normally operate in a particular operational context. This would hold true for systems attached to audio or video production systems. The key differentiators here are the type of decoders and I/O (input/output) used for the asset ingest, use and outgest. Since the essence (and proxy copy) is described by metadata, the metadata can be used as a guide to the playout decoders, transcoders, and channels as well as an input to access control rules. This means that essentially, the essence can be treated as a non-described storage object except when being accessed for viewing or editing. There is relevance to this when considering the overall design and use of larger implementations. The closer the asset is to the ingest/edit/play out tool, the greater the technical architecture needs to accommodate delivery requirements such as bandwidth, latency, capacity, access control, availability of resources, etc. The further the asset moves into a general storage architecture (e.g. Hierarchical Storage Management [HSM]) the more it can be treated as a general blob (binary large object) that is typically held in the file system, not the database. The impact of this set of needs means that it is possible and reasonable to design larger systems using smaller, more expensive performance systems at the edge of the network where the essence is being used in its intended form and less expensive systems further back for storage and archival. This type of design is an example of Infrastructure Convergence Architecture where the line of business operations technology and IT technologies are dependent on one another for functional and performance (non-functional requirements) requirements.

3.3 Types of Digital Asset Management systems

The following broad categories of digital asset management systems may be distinguished.

- Brand asset management systems, with a focus on facilitation of content re-use within large organisations. Here the content is largely marketing- or sales-related, for example, product imagery, logos, marketing collateral or fonts, to give a few examples.
- Library asset management systems, with a focus on storage and retrieval of large amounts of infrequently changing media assets, for example in video or photo archiving.
- Production asset management systems, with a focus on storage, organisation and revision control of frequently changing digital assets, for example in digital media production.
- Digital supply chain services, pushing digital content out to digital retailers (e.g. music, videos and games).

3.4 Providers

Enterprise-level solutions often involve scalable, reliable, configurable products that can handle vast numbers of assets (files) as well as large numbers of simultaneous users, workflows, or use cases (multiple applications simultaneously operating against the system). Enterprise systems may, but do not necessarily, include customised products or features added on to the base system or custom developed to match an organisation's workflow. Enterprise class systems are also applicable to Small and Medium Businesses (SMBs), or departments or work groups within an organisation. In many cases these systems enter a company in one department and eventually expand to others or the entire enterprise as its utility becomes proven, understood and valued. Enterprise systems are offered as installed software or as Software as a Service (SaaS) hosted, web-based offers that are managed and maintained externally.

For individuals, either proprietary or open source applications can be adequate for digital asset management. Some image viewers provide management functionality, including backing up, organising, and reading/writing metadata and keywords.

3.5 Digital Asset

A digital asset is any form of content and/or media that have been formatted into a binary source which include the right to use it. A digital file without the right to use it is not an asset. Digital assets are categorised in three major groups which may be defined as textual content (digital assets), images (media assets) and multimedia (media assets); (van Niekerk, A. J. 2006).

In order to have a clear understanding of digital asset management the definitions of the different types of digital assets need to be defined and the difference must be specified. There are a number of management systems related to digital asset management (Austerberry, 2004) which are:

a) Digital Asset Management (DAM).
b) Digital Content Management (DCM).
c) Enterprise Content Management (ECM).
d) Digital Media Management (DMM).
e) Media Asset Management (MAM).
f) WEB Content Management (WCM).

An art asset, in computer graphics and related fields (particularly video game and visual effects production) is an individual piece of digital media used in the creation of a larger production. Art assets include synthetic and photographic bitmaps (often used for texture mapping, 3D models consisting of polygon meshes or curved surfaces), shaders, motion captured or hand-animated animation data, video and audio samples.

The term “art” is used to distinguish the creative (or real-world) elements of a production from the software or hardware used to create it: there is no requirement that the data represents anything artistic.

Digital asset management is expected to be a multi-billion dollar industry as corporations and individuals migrate traditional graphic, broadcast and print assets to the digital format. Companies including Apple, Oracle, Microsoft, Getty Images and others are aggressively expanding their enterprises to provide third-party digital asset management via web-based repositories. This trend will continue as business and consumers evolve from traditional analogue materials.

SELF ASSESSMENT EXERCISE

Define digital asset.

3.6 Document Imaging

Document Imaging is an information technology category for systems capable of replicating documents commonly used in business. Document Imaging Systems can take many forms including microfilm, on demand printers, facsimile machines, copiers, multifunction printers, document scanners, Computer Output Microfilm (COM) and archive writers. In the last 15 years, Document Imaging has been used to describe software-based computer systems that capture, store and reprint images.

Document Imaging is part of the set of technologies within the Enterprise Content Management category.

In the early days of content management technologies, the term “Document Imaging” was used interchangeably with “Document Image Management” as the industry tried to separate itself from the micrographic and reprographic technologies. Organisations like National Micrographic Association (NMA) and American Records Manager Association (ARMA) found themselves inventing new ways to describe these new archive and library tools. The ‘NMA’ became the Association of Image and Information Management. Currently, document imaging industry information can be found at the Association for Information and Imaging Management (AIIM). A certification for document imaging can be obtained from CompTIA called the CDIA+.

In the late 80's and early 90's a new document management technology emerged: Electronic Document Management. This new technology was built around the need to manage and secure the volume of electronic documents (spreadsheets, word processing documents) created in organisations. Electronic documents can change constantly and those changes require security authorisations and tracking, which are the core functionality of an EDMS (Electronic Document Management System).

EDMS is not limited to native word processing and spreadsheet files, scanned images also have a life being redacted by users as you would a paper document.

4.0 CONCLUSION

Digital asset management is the trend in document preservation. It gives a document a much longer life span and also makes such document to be easily assessed. Through digital asset management system, a document or image and the likes can be made available over networks such as the Internet. By so doing, it is available to more users per time.

5.0 SUMMARY

- Digital Asset Management consists of tasks and decisions surrounding ingesting, annotating, cataloguing, storage and retrieval of digital assets, such as digital photographs, animations, videos and music.
- Many businesses and organisations are adopting Digital Asset Management as a business strategy because managing image, video and other media assets presents unique challenges and requires solutions designed specifically to streamline the acquisition, storage and retrieval of digital media.
- There are broad categories of digital asset management systems that may be distinguished.
- Enterprise-level solutions often involve scalable, reliable, configurable products that can handle vast numbers of assets (files) as well as large numbers of simultaneous users, workflows, or use cases (multiple applications simultaneously operating against the system).
- Digital Asset: A digital asset is any form of content and/or media that have been formatted into a binary source which include the right to use it.
- Document Imaging is an information technology category for systems capable of replicating documents commonly used in business.

6.0 TUTOR-MARKED ASSIGNMENT

1. Briefly discuss the types of digital asset management.
2. Mention 5 types of digital assets.

7.0 REFERENCES/FURTHER READING

Arthur, M. (undated) Introduction to Digital Asset Management: Just What is a DAM?

Austerberry, D. (2006). Digital Asset Management. (2nd ed.). Focal Press.

CompTIA CDIA+™ Certification Description.

Jacobsen, Jens., Schlenker, Tilman & Edwards, Lisa (2005). *Implementing a Digital Asset Management System: For Animation, Computer Games, and Web Development. Focal* Press.

Krogh, P. (2005). The DAM Book: Digital Asset Management for *Photographers. O'Reilly Media.*

Mauthe, A. & Thomas, P. (2004). Professional Content Management *Systems: Handling Digital Media Assets. Wiley*.

UNIT 5 DOCUMENT MANAGEMENT: RETURN ON INVESTMENT (ROI) ANALYSIS

CONTENTS

1.0 INTRODUCTION

This unit will outline how a document management system can generate a significant return on investment for small to medium sized businesses. The changes in technology pricing over the last 10 years has finally made it possible for small to medium sized businesses to enjoy the same efficiencies provided by information technology systems that large enterprises have enjoyed for decades. Think about where we would be today if banks did not have computer systems, or clinics could not provide access to patient folders at remote locations. These advances in technology can now be applied with ease at the small to medium enterprise level.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define and explain the concept of return on investment (ROI) for document management systems
- explain the current state of ROI in document management
- identify the benefits of document management system
- explain how the benefits of document management systems reflect in ROI
- demonstrate how to calculate document management systems in ROI
- prepare some statistics of records management system.

3.0 MAIN CONTENT

3.1 Current Conditions

In today's business world, the ROI of any project is important as competition increases, it is imperative that a company make sure investments generate a large enough return. Increasing business efficiency is the most compelling reason for investing money in any project. Obviously, technology has helped many businesses over the last 10 to15 years become more efficient. By just replacing typewriters with word processors and calculators with spreadsheets, had initiated a huge increase in productivity per employees. Taking the next step is more challenging than simply replacing one tool for another as these programs were basically point solutions. It was predicted as early as 1975, that the "paperless office had arrived". Obviously progress has been made toward this goal, but the amount of paperwork necessary to run a business (particularly a regulated one such as financial services and medical clinics) has increased a great deal since 1975. Just to maintain the status quo requires businesses to move to the next level in office productivity and implement a complete document management solution.

Almost every business in today's workplace uses some type of document management system (including paper based systems) to maintain the documents used on a daily basis. Accounting departments maintain A/P and A/R files for customers and vendors. The sales department maintains orders for each of the customers. Customer service representatives maintain records on service calls and the service needed. A comprehensive Electronic Document Management System (EDMS) can provide a method for storing all relevant documents about a particular customer or vendor, allowing office staff to gain a total view of the customer or vendor. If the EDMS can be integrated into the point software solutions that each department uses, significant gains in productivity and efficiency can be gained. For instance, the accounting department could access documents (orders, invoices, contracts,) online from their accounting application. Sales could access documents from their CRM (customer relationship management) software. The goal of an EDMS should be the software glue that ties the different software packages together in such a manner that all departments in a business gain efficiencies.

3.2 What is ROI?

ROI is simply Return On Investment. In layman's terms this is the amount of time it takes to get the value of the dollars spent on a project or item back (returned to the business). For instance, if a large lathe is purchased for a machine shop and it cost N50,000 and each month a profit of N10,000 can be tied to the lathe, then your ROI (in simplistic terms) is five months (N50,000/N10,000). This type of formula can be applied to virtually anything purchased, although it can be harder to determine for some purchases. Typically, the effort to calculate an ROI is only done on medium to large investments, as the time spent can be extensive and requires gathering appropriate data.

3.3 What is a Document Management Solution?

Simply put, a document management solution is a system that can replace, not only the paper files and documents in an office, but enables the user to send an electronic document through the same steps a paper document or file would follow. For instance, if a financial services firm has a new account form that is filled out for each client account and it must proceed from the agent to the trader and then to accounting, then a document management system must provide the same path. So, a true management system not only provides storage, but must provide the same workflow capabilities. Along the way, it must also protect documents so that only the necessary people can view them. An example, might be a medical clinic which must comply with HIPAA regulations. Patient records must be kept confidential and an audit trail kept for all record access. An EDMS is a step up from the point solutions like spreadsheets and word processors as the entire office must be capable of using them. Therefore, they must be easy and straightforward to use and ideally would integrate with other applications the office is already using with a minimum of effort.

So, to sum up, the attributes of a document management solution are:

- maintain a repository of electronic documents.
- provide a mechanism of securing the documents.
- integrate with other software solutions being used.
- provide a method for putting the documents into a defined workflow.
- supply audit data providing the four W's (Who, What, Where, When).
- capture form data in a database for usage in monitoring and managing the business.
- comply with any necessary regulatory requirements.

4.0 CONCLUSION

Projects planning to roll out records management systems across an entire organisation face considerable challenges, not least of which is that they require the active participation of all staff to be successful. To achieve this level of cultural change, three critical success factors have been identified. They are:

5.0 SUMMARY

- Historically, records management was the responsibility of a small number of specialist staff members within an organisation.

6.0 TUTOR-MARKED ASSIGNMENT

3. Identify 5 competing platforms for user in saving records.
4. Briefly discuss the issue of finding and storing records.

7.0 REFERENCES/FURTHER READING

Robertson, J. (undated). Success Factors in Records Management Systems, Step II Designs, Sydney, Australia.

MODULE 3

UNIT 1 DOCUMENT MANAGEMENT SYSTEM

CONTENTS

1.0 INTRODUCTION

A document management system (DMS) is a computer system (or set of computer programs) used to track and store electronic documents and/or images of paper documents. The term has some overlap with the concepts of Content Management Systems and is often viewed as a component of Enterprise Content Management Systems and related to Digital Asset Management, Document imaging, Workflow systems and Records Management systems. Contract Management and Contract Lifecycle Management (CLM) can be viewed as either components or implementations of ECM.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define document management system
- outline the history and development process of document management system

- compare and contrast document management system and content management systems
- describe the basic components of document management systems
- specify the issues addressed by document management systems.
- list the types of document management systems available off the shelf.

3.0 MAIN CONTENT

3.1 History

Beginning in the 1980s, a number of vendors began developing systems to manage paper-based documents. These systems managed paper documents, which included not only printed and published documents, but also photos, prints, etc.

Later, a second system was developed, to manage electronic documents, i.e., all those documents, or files, created on computers, and often stored on local user file systems. The earliest Electronic Document Management (EDM) systems were either developed to manage proprietary file types, or a limited number of file formats. Many of these systems were later referred to as document imaging systems, because the main capabilities were capture, storage, indexing and retrieval of image file formats. These systems enabled an organisation to capture faxes and forms, save copies of the documents as images, and store the image files in the repository for security and quick retrieval (retrieval was possible because the system handled the extraction of the text from the document as it was captured, and the text indexer provided text retrieval capabilities).

EDM systems evolved to where the system was able to manage any type of file format that could be stored on the network. The applications grew to encompass electronic documents, collaboration tools, security, and auditing capabilities.

3.2 Document Management and Content Management

There is considerable confusion in the market between Document Management Systems (DMS) and Content Management Systems (CMS). This has not been helped by the vendors, who are keen to market their products as widely as possible.

These two types of systems are very different, and serve complementary needs. While there is an ongoing move to merge the two together (a positive step), it is important to understand when each system is appropriate.

Document Management Systems (DMS)

Document management is certainly the older discipline, born out of the need to manage huge numbers of documents in organisations.

Mature and well-tested, document management systems can be characterised as follows:

- focused on managing documents, in the traditional sense (like Word files)
- each unit of information (document) is fairly large, and self-contained
- there are few (if any) links between documents
- provides limited integration with repository (check-in, check-out, etc)
- focused primarily on storage and archiving
- includes powerful workflow
- targeted at storing and presenting documents in their native format
- limited web publishing engine typically produces one page for each document

Note that this is just a generalised description of a DMS, with most systems offering a range of unique features and capabilities. Nonetheless, this does provide a representative outline of common DMS functionality.

A typical document management scenario:

- A large legal firm purchases a DMS to track the huge number of *advice documents, contracts and briefs. It allows lawyers to easily retrieve earlier advice, and to use 'precedent' templates to quickly create new documents.*

You can’t build a website with just a DM system

Content Management Systems (CMS)

Content management is more recent, and is primarily designed to meet the growing needs of the website and intranet markets.

A content management system can be summarised as follows:

- manages small, interconnected units of information (e.g. web pages)

- each unit (page) is defined by its location on the site
- extensive cross-linking between pages
- focused primarily on page creation and editing
- provides tight integration between authoring and the repository (metadata, etc)
- provides a very powerful publishing engine (templates, scripting, etc)

A typical content management scenario:

- A CMS is purchased to manage the 3000 page corporate website. *Template-based authoring allows business groups to easily create content, while the publishing system dynamically generates richly formatted pages.*

Content management and document management are complementary, not competing technologies. You must choose an appropriate system if business needs are to be met.

3.3 Components

Document management systems commonly provide storage, versioning, metadata, security, as well as indexing and retrieval capabilities. Here is a description of these components:

Metadata

Metadata is typically stored for each document. Metadata may, for example, include the date the document was stored and the identity of the user storing it. The DMS may also extract metadata from the document automatically or prompt the user to add metadata. Some systems also use optical character recognition on scanned images, or perform text extraction on electronic documents. The resulting extracted text can be used to assist users in locating documents by identifying probable keywords or providing for full text search capability, or can be used on its own. Extracted text can also be stored as a component of metadata, stored with the image, or separately as a source for searching document collections.

Integration

Many document management systems attempt to integrate document management directly into other applications, so that users may retrieve existing documents directly from the document management system repository, make changes, and save the changed document back to the repository as a new version, all without leaving the application. Such integration is commonly available for office suites and e-mail or collaboration/groupware software.

Capture

Images of paper documents using scanners or multifunction printers can be captured. Optical Character Recognition (OCR) software is often used, whether integrated into the hardware or as stand-alone software, in order to convert digital images into machine readable text.

Indexing

Track electronic documents. Indexing may be as simple as keeping track of unique document identifiers; but often it takes a more complex form, providing classification through the documents' metadata or even through word indexes extracted from the documents' contents. Indexing exists mainly to support retrieval. One area of critical importance for rapid retrieval is the creation of an index topology.

Storage

Stores electronic document: storage of the documents often includes management of those same documents; where they are stored, for how long, migration of the documents from one storage media to another (Hierarchical storage management) and eventual document destruction.

Retrieval

Retrieve the electronic documents from the storage. Although, the notion of retrieving a particular document is simple, retrieval in the electronic context can be quite complex and powerful. Simple retrieval of individual documents can be supported by allowing the user to specify the unique document identifier, and having the system use the basic index (or a non-indexed query on its data store) to retrieve the document. More flexible retrieval allows the user to specify partial search terms involving the document identifier and/or parts of the expected metadata. This would typically return a list of documents which match the user's search terms. Some systems provide the capability to specify a Boolean expression containing multiple keywords or example of phrases expected to exist within the documents' contents. The retrieval for this kind of query may be supported by previously-built indexes, or may perform more time-consuming searches through the documents' contents to return a list of the potentially relevant documents. See also Document retrieval.

Distribution Security

Document security is vital in many document management applications. Compliance requirements for certain documents can be quite complex depending on the type of documents. For instance, the Health Insurance Portability and Accountability Act (HIPAA) requirements dictate that medical documents have certain security requirements. Some document management systems have a right management module that allows an administrator to give access to documents based on the type to only certain people or groups of people.

Workflow

Workflow is a complex problem and some document management systems have a built in workflow module. There are different types of workflows. Usage depends on the environment the EDMS is applied to. Manual workflow requires a user to view the document and decide who to send it to. Rules-based workflow allows an administrator to create a rule that dictates the flow of the document through an organisation: for instance, an invoice passes through an approval process and then is routed to the accounts payable department. Dynamic rules allow for branches to be created in a workflow process. A simple example would be to enter an invoice amount and if the amount is lower than a certain set amount, it follows different routes through the organisation.

Collaboration

Collaboration should be inherent in an EDMS. Documents should be capable of being retrieved by an authorised user and worked on. Access should be blocked to other users while work is being performed on the document.

Versioning

Versioning is a process by which documents are checked in or out of the document management system, allowing users to retrieve previous versions and to continue work from a selected point. Versioning is useful for documents that change over the time and require updating, but it may be necessary to go back to a previous copy.

SELF ASSESSMENT EXERCISE

Explain the term workflow.

3.4 Issues Addressed in Document Management

There are several common issues that are involved in managing documents, whether the system is an informal, ad-hoc, paper-based method for one person or if it is a formal, structured, computer enhanced system for many people across multiple offices. Most methods for managing documents address the following areas:

Location	Where will documents be stored? Where will people need to go to access documents? Physical journeys to filing cabinets and file rooms are analogous to the onscreen navigation required to use a document management system.
Filing	How will documents be filed? What methods will be used to organise or index the documents to assist in later retrieval? Document management systems will typically use a database to store filing information.
Retrieval	How will documents be found? Typically, retrieval encompasses both browsing through documents and searching for specific information.
Security	How will documents be kept secure? How will unauthorised personnel be prevented from reading, modifying or destroying documents?
Disaster recovery	How can documents be recovered in case of destruction from fires, floods or natural disasters?
Retention period	How long should documents be kept, i.e. retained? As organisations grow and regulations increase, informal guidelines for keeping various types of documents give way to more formal records management practices.
Archiving	How can documents be preserved for future readability?
Distribution	How can documents be available to the people that need them?
Workflow	If documents need to pass from one person to another, what are the rules for how their work should flow?
Creation	How are documents created? This question becomes important when multiple people need to collaborate, and the logistics of version control and authoring arise.
Authentication	Is there a way to vouch for the authenticity of a document?

3.5 Using XML in Document and Information Management

The attention paid to XML (Extensible Markup Language), whose 1.0 standard was published February 10, 1998, is impressive. XML has been heralded as the next important internet technology, the next step following HTML, and the natural and worthy companion to the Java programming language itself. Enterprises of all stripes have rapturously embraced XML. An important role for XML is in managing not only documents but also the information components on which documents are based.

Document Management: Organising Files

Document management as a technology and a discipline has traditionally augmented the capabilities of a computer's file system. By enabling users to characterise their documents, which are usually stored in files, document management systems enable users to store, retrieve, and use their documents more easily and powerfully than they can do within the file system itself.

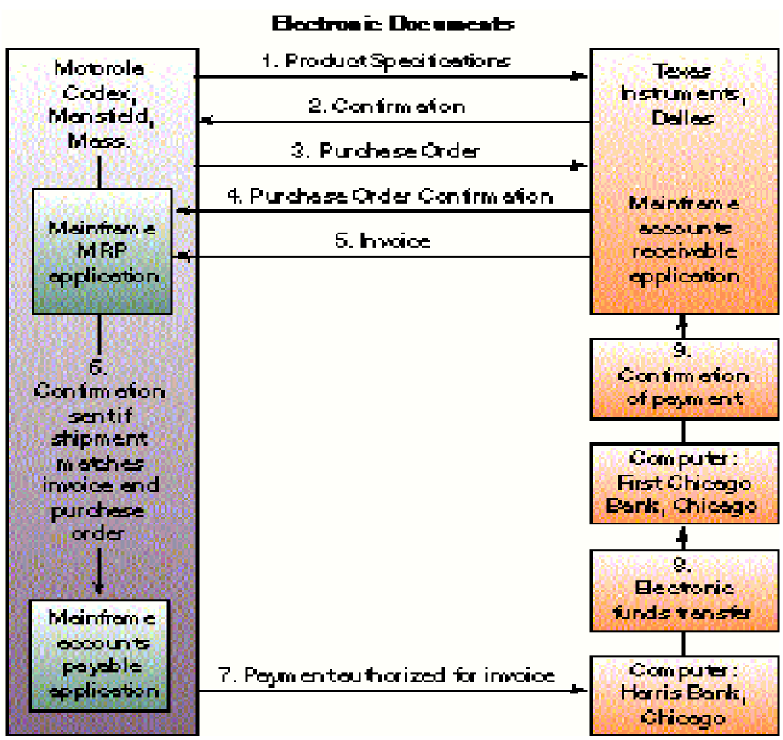


Fig. 1: An example of EDI. Motorola Codex uses EDI links to its supplier, Texas Instruments, for the exchange of business documents. Codex also makes electronic funds transfers to its banks to pay its suppliers.

Long before anyone thought of XML, document management systems were originally developed to help law offices maintain better control over and access to the many documents that legal professionals generate.
The basic mechanisms of the first document management systems performed, among others, these simple but powerful tasks:

- add information about a document to the file that contains the document
- organise the user-supplied information in a database
- create information about the relationships between different documents

In essence, document management systems created libraries of documents in a computer system or a network. The document library contained a "card catalogue" where the user-supplied information was stored and through which users could find out about the documents and access them. The card catalogue was a database that captured information about a document, such as these:

- Author: who wrote or contributed to the document
- Main topics: what subjects are covered in the document
- Origination date: when was it started
- Completion date: when was it finished
- Related documents: what other documents are relevant to this document
- Associated applications: what programs are used to process the document
- Case: to which legal case (or other business process) is the document related

Armed with a database of such information about documents, users could find information in more sensible and intuitive ways than scanning different directories' lists of contents, hoping that a file's name might reveal what the file contained. Many people consider document management systems' first achievement to have created "a file system within the file system."

Soon, document management systems began to provide additional and valuable functionality. By enriching the databases of information about the documents (the metadata), these systems provided these capabilities:

- Version tracking: sees how a document evolves over time
- Document sharing: sees in what business processes the document is used and re-used

- Electronic review: enables users to add their comments to a document without actually changing the document itself
- Document security: refines the different types of access that different users need to the document
- Publishing management: controls the delivery of documents to different publishing process queues
- Workflow integration: associates the different stages of a document's life-cycle with people and projects with schedules

These critical capabilities (among others) of document management systems have proven enormously successful, fueling a multi-billion dollar business.

XML: Managing Document Components

XML and its parent technology, SGML (Standard Generalized Markup Language), provide the foundation for managing not only documents but also the information components of which the documents are composed. This is due to some notable characteristics of XML data.

Documents versus Files

In XML, documents can be seen independently of files. One document can comprise many files, or one file can contain many documents. This is the distinction between the physical and logical structure of information. XML data is primarily described by its logical structure. In a logical structure, principal interest is placed on what the pieces of information are and how they relate to one another and secondary interest is placed on the physical items that constitute the information.

Rather than relying on file headers and other system, specific characteristics of a file as the primary means for understanding and managing information, XML relies on the markup in the data itself. A chapter in a document is not a chapter because it resides in a file called chapter1.doc but because the chapter's content is contained in the <chapter> and </chapter> element tags.

Because elements in XML can have attributes, the components of a document can be extensively self-descriptive. For example, in XML you can learn a lot about the chapter without actually reading it if the chapter's markup is rich in attributes, as in <chapter language=”English” subject=”colonial economics” revision_date=”19980623” author=”Joan X. Pringle” thesis_advisor=”Ramona Winkelhoff”>. When the elements carry self-describing metadata with them, systems that understand XML syntax can operate on those elements in useful ways, just like a traditional document management system can. But there is a major difference.

Information versus Documents

XML markup provides metadata for all components of a document, not merely the object that contains the document itself. This makes the pieces of information that constitute a document just as manageable as the fields of a record in a database. Because XML data follows syntactic rules for well-formedness and proper containment of elements, document management systems that can correctly read and parse XML data can apply the functions of document management system, such as those mentioned above, to any and all information components inside the document.

The focus on information rather than documents from XML offers some important capabilities such as the following.

- Reuse of information

While standard document management systems do offer some measure of information reuse through file sharing, information management systems based on XML or SGML enable people to share pieces of common information without storing the piece of information in multiple places.

- Information harvesting

By enabling people to focus on information components that make up documents rather than on the documents themselves, these systems can identify and capture useful information components that have ongoing value “buried” inside documents whose value as documents is limited. That is, a particular document may be useful only for a short time, but chunks of information inside that document may be reusable and valuable for a longer period.

- Fine-granularity text-management applications

Because the information components in XML documents are identifiable, manipulatable, and manageable, XML information management technology can support real economies in applications such as translation of technical manuals.

- Evaluating Product Offerings

While the general world of document management and information management is moving toward adoption of structured information and use of XML and SGML, some product offerings distinguish themselves by using underlying database management products with native support for object-oriented data. Object-oriented data matches the structure of XML data quite well and database systems that comprehend object-oriented data adapt well to the tasks of managing XML information.

By contrast, other information management products that comprehend XML or SGML data use relational database systems and provide their own object-oriented extensions to those database systems in order to comprehend object-oriented data such as XML or SGML data, and relying on such implementations have also garnered success and respect in the document management marketplace.

4.0 CONCLUSION

Document management systems have added variety to the pool of options available in database management in corporations. Many products are on the shelf for end users to choose from. T he use of document management systems has encouraged the concept and drive for paperless office and transactions. It is a concept that truly makes the future bright as man tends toward greater efficiency by eliminating use of papers and hard copies of data and information.

5.0 SUMMARY

- A document management system (DMS) is a computer system (or set of computer programs) used to track and store electronic documents and/or images of paper documents
- Beginning in the 1980s, a number of vendors began developing systems to manage paper-based documents. These systems managed paper documents, which included not only printed and published documents, but also photos, prints, etc.
- There is considerable confusion in the market between document management systems (DMS) and content management systems (CMS).
- Document management systems commonly provide storage, versioning, metadata, security, as well as indexing and retrieval capabilities.
- There are several common issues that are involved in managing documents, whether the system is an informal, ad-hoc, paper-based method for one person or if it is a formal, structured,

computer enhanced system for many people across multiple offices

- The attention paid to XML (Extensible Markup Language), whose 1.0 standard was published February 10, 1998, is impressive. XML has been heralded as the next important internet technology, the next step following HTML, and the natural and worthy companion to the Java programming language itself. Enterprises of all stripes have rapturously embraced XML.

6.0 TUTOR-MARKED ASSIGNMENT

1. List 5 characteristics of a document management system.
2. Briefly discuss the evaluation of product offerings.

7.0 REFERENCES/FURTHER READING

BBC -h2g2 guide Shoebox Storage.

Robertson, J. Published on 14 February 2003.

Craine, K. “Excerpts from Designing a Document Strategy” (HTML). Craine Communications Group.

Miles, L. M. & Ernest, A. C. (2002). The Paperless Office: Accepting *Digitized Data (PDF). Troy State University.*

UNIT 2 DATA WAREHOUSE

CONTENTS

1.0 INTRODUCTION

A data warehouse is a repository of an organisation's electronically stored data. Data warehouses are designed to facilitate reporting and analysis.

This classic definition of the data warehouse focuses on data storage. However, the means to retrieve and analyse data, to extract, transform and load data, and to manage the dictionary data are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata.

In contrast to data warehouses are operational systems which perform day-to-day transaction processing.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- define data warehouse
- outline the history and development process of data warehouse
- list various benefits of data warehouse
- describe the architecture of a data warehouse
- compare and contrast Data Warehouses and Operational Systems
- explain what is a data warehouse appliance, and the disadvantages of data warehouse
- predict what the future holds for data warehouse concept.

3.0 MAIN CONTENT

3.1 History

The concept of data warehousing dates back to the late-1980s when IBM researchers Barry Devlin and Paul Murphy developed the "business data warehouse". In essence, the data warehousing concept was intended to provide an architectural model for the flow of data from operational systems to decision support environments. The concept attempted to address the various problems associated with this flow - mainly, the high costs associated with it. In the absence of a data warehousing architecture, an enormous amount of redundancy of information was required to support the multiple decision support environment that usually existed. In larger corporations, it was typical for multiple decision support environments to operate independently. Each environment served different users but often required much of the same data. The process of gathering, cleaning and integrating data from various sources, usually long existing operational systems (usually referred to as legacy systems), was typically in part replicated for each environment. Moreover, the operational systems were frequently re-examined as new decision support requirements emerged. Often, new requirements necessitated gathering, cleaning and integrating new data from the operational systems that were logically related to prior gathered data.

Based on analogies with real-life warehouses, data warehouses were intended as large-scale collection/storage/staging areas for corporate data. Data could be retrieved from one central point or data could be distributed to "retail stores" or "data marts" which were tailored for ready access by users.

3.2 Benefits of Data Warehousing

Some of the benefits that a data warehouse provides are as follows.

- A data warehouse provides a common data model for all data of interest regardless of the data's source. This makes it easier to report and analyse information than it would be if multiple data models were used to retrieve information such as sales invoices, order receipts, general ledger charges, etc.
- Prior to loading data into the data warehouse, inconsistencies are identified and resolved. This greatly simplifies reporting and analysis.
- Information in the data warehouse is under the control of data warehouse users so that, even if the source system data is purged over time, the information in the warehouse can be stored safely for extended periods of time.
- Because they are separate from operational systems, data warehouses provide retrieval of data without slowing down operational systems.
- Data warehouses facilitate decision support system applications such as trend reports (e.g., the items with the most sales in a particular area within the last two years), exception reports, and reports that show actual performance versus goals.
- Data warehouses can work in conjunction with and, hence, enhance the value of operational business applications, notably customer relationship management (CRM) systems.

3.3 Data Warehouse Architecture

Architecture, in the context of an organisation’s data warehousing efforts, is a conceptualisation of how the data warehouse is built. There is no right or wrong architecture. The worthiness of the architecture can be judged in how the conceptualisation aids in the building, maintenance, and usage of the data warehouse.

One possible simple conceptualisation of a data warehouse architecture consists of the following interconnected layers:

Operational database layer

The source data for the data warehouse - An organisation’s ERP systems fall into this layer.

Informational access layer

The data accessed for reporting and analysing and the tools for reporting and analysing data - business intelligence tools fall into this layer. And the Inmon-Kimball differences about design methodology, discussed later in this article, have to do with this layer.

Data access layer

The interface between the operational and informational access layer - tools to extract, transform, load data into the warehouse fall into this layer.

Metadata layer

The data directory - This is often usually more detailed than an operational system data directory. There are dictionaries for the entire warehouse and sometimes dictionaries for the data that can be accessed by a particular reporting and analysis tool.

3.4 Normalised versus Dimensional Approach to Storage of Data

There are two leading approaches to storing data in a data warehouse - the dimensional approach and the normalised approach.

In the dimensional approach, transaction data are partitioned into either "facts", which are generally numeric transaction data, or "dimensions", which are the reference information that gives context to the facts. For example, a sales transaction can be broken up into facts such as the number of products ordered and the price paid for the products, and into dimensions such as order date, customer name, product number, order ship-to and bill-to locations, and salesperson responsible for receiving the order. A key advantage of a dimensional approach is that the data warehouse is easier for the user to understand and to use. Also, the retrieval of data from the data warehouse tends to operate very quickly.
The main disadvantages of the dimensional approach are: 1) in order to maintain the integrity of facts and dimensions, loading the data warehouse with data from different operational systems is complicated, and 2) it is difficult to modify the data warehouse structure if the organisation adopting the dimensional approach changes the way in which it does business.

In the normalised approach, the data in the data warehouse are stored following, to a degree, the Codd normalisation rule. Tables are grouped together by subject areas that reflect general data categories (e.g., data on customers, products, finance, etc.) The main advantage of this approach is that it is straightforward to add information into the database. A disadvantage of this approach is that, because of the number of tables involved, it can be difficult for users both to 1) join data from different sources into meaningful information and then 2) access the information without a precise understanding of the sources of data and of the data structure of the data warehouse.

These approaches are not exact opposites of each other. Dimensional approaches can involve normalising data to a degree.

3.5 Conforming Information

Another important decision in designing a data warehouse is which data to conform and how to conform the data. For example, one operational system feeding data into the data warehouse may use “M” and “F” to denote sex of an employee while another operational system may use “Male” and “Female”. Though, this is a simple example, much of the work in implementing a data warehouse is devoted to making similar meaning data consistent when they are stored in the data warehouse. Typically, extract, transform, load tools are used in this work.

3.6 Top-Down versus Bottom-Up Design Methodologies

Bottom-up design

Ralph Kimball, a well-known author on data warehousing, is a proponent of the bottom-up approach to data warehouse design. In the bottom-up approach data marts are first created to provide reporting and analytical capabilities for specific business processes. Data marts contain atomic data and, if necessary, summarised data. These data marts can eventually be unified together to create a comprehensive data warehouse. The combination of data marts is managed through the implementation of what Kimball calls “data warehouse bus architecture”.

Business value can be returned as quickly as the first data marts can be created. Maintaining tight management over the data warehouse bus architecture is fundamental to maintaining the integrity of the data warehouse. The most important management task is making sure dimensions among data marts are consistent. In Kimball words, this means that the dimensions “conform”.

Top-down design

Bill Inmon, one of the first authors on the subject of data warehousing, has defined a data warehouse as a centralised repository for the entire enterprise. Inmon is one of the leading proponents of the top-down approach to data warehouse design, in which the data warehouse is designed using a normalised enterprise data model. “Atomic” data, that is, data at the lowest level of detail, are stored in the data warehouse. Dimensional data marts containing data needed for specific business processes or specific departments are created from the data warehouse.
In the Inmon vision, the data warehouse is at the center of the “Corporate Information Factory” (CIF), which provides a logical framework for delivering business intelligence (BI) and business management capabilities. The CIF is driven by data provided from business operations.

Inmon states that the data warehouse is:

Subject-oriented

The data in the data warehouse is organised so that all the data elements relating to the same real-world event or object are linked together.

Time-variant

The changes to the data in the data warehouse are tracked and recorded so that reports can be produced showing changes over time.

Non-volatile

Data in the data warehouse is never over-written or deleted - once committed, the data is static, read-only, and retained for future reporting.

Integrated

The data warehouse contains data from most or all of an organisation's operational systems and this data is made consistent.

The top-down design methodology generates highly consistent dimensional views of data across data marts since all data marts are loaded from the centralised repository. Top-down design has also proven to be robust against business changes. Generating new dimensional data marts against the data stored in the data warehouse is a relatively simple task. The main disadvantage to the top-down methodology is that it represents a very large project with a very broad scope. The up-front cost for implementing a data warehouse using the top-down methodology is significant, and the duration of time from the start of project to the point that end users experience initial benefits can be substantial. In addition, the top-down methodology can be inflexible and unresponsive to changing departmental needs during the implementation phases.

Hybrid design

Over the time, it has become apparent to proponents of bottom-up and top-down data warehouse design that both methodologies have benefits and risks. Hybrid methodologies have evolved to take advantage of the fast turn-around time of bottom-up design and the enterprise-wide data consistency of top-down design

SELF ASSESSMENT EXERCISE

What are the meanings of the acronyms BI and CIF.

3.7 Data Warehouses versus Operational Systems

Operational systems are optimised for preservation of data integrity and speed of recording of business transactions through the use of database normalisation and an entity-relationship model. Operational system designers generally follow the Codd rules of data normalisation in order to ensure data integrity. Codd defined five increasingly stringent rules of normalisation. Fully normalised database designs (that is, those satisfying all five Codd rules) often result in information from a business transaction being stored in dozens to hundreds of tables. Relational databases are efficient at managing the relationships between these tables. The databases have very fast insert/update performance because only a small amount of data in those tables is affected each time a transaction is processed. Finally, in order to improve performance, older data are usually periodically purged from operational systems.

Data warehouses are optimised for speed of data retrieval. Frequently, data in data warehouses are denormalised via a dimension-based model. Also, to speed data retrieval, data warehouse data are often stored multiple times - in their most granular form and in summarised forms called aggregates. Data warehouse data are gathered from the operational systems and held in the data warehouse even after the data has been purged from the operational systems.

3.8 Evolution in Organisation Use of Data Warehouses

Organisations generally start off with relatively simple use of data warehousing. Over the time, more sophisticated use of data warehousing evolves. The following general stages of use of the data warehouse can be distinguished:

Off line Operational Databases

Data warehouses in this initial stage are developed by simply copying the data of an operational system to another server where the processing load of reporting against the copied data does not impact the operational system's performance.

Off line Data Warehouse

Data warehouses at this stage are updated from data in the operational systems on a regular basis and the data warehouse data is stored in a data structure designed to facilitate reporting.

Real Time Data Warehouse

Data warehouses at this stage are updated every time an operational system performs a transaction (e.g., an order or a delivery or a booking.)
Integrated Data Warehouse

Data warehouses at this stage are updated every time an operational system performs a transaction. The data warehouses then generate transactions that are passed back into the operational systems.

3.9 Disadvantages of Data Warehouses

There are also disadvantages to using a data warehouse. Some of them are:

- Over their life, data warehouses can have high costs. The data warehouse is usually not static. Maintenance costs are high.
- Data warehouses can get outdated relatively quickly. There is a cost of delivering suboptimal information to the organisation.
- There is often a fine line between data warehouses and operational systems. Duplicate, expensive functionality may be developed. Or, functionality may be developed in the data warehouse that, in retrospect, should have been developed in the operational systems and vice versa.

3.10 Data Warehouse Appliance

A data warehouse appliance is an integrated set of servers, storage, OS, DBMS and software specifically pre-installed and pre-optimised for data warehousing. Alternatively, the term is also used for similar software-only systems that purportedly are very easy to install on specific recommended hardware configurations. DW appliances provide solutions for the mid-to-large volume data warehouse market, offering low-cost performance most commonly on data volumes in the terabyte to petabyte range.

Technology Primer

Most DW appliance vendors use Massively Parallel Processing (MPP) architectures to provide high query performance and platform scalability. MPP architectures consist of independent processors or servers executing in parallel. Most MPP architectures implement a "shared nothing architecture" where each server is self-sufficient and controls its own memory and disk. Shared nothing architectures have a proven record for high scalability and little contention. DW appliances distribute data onto dedicated disk storage units connected to each server in the appliance. This distribution allows DW appliances to resolve a relational query by scanning data on each server in parallel. The divide-and-conquer approach delivers high performance and scales linearly as new servers are added into the architecture.

MPP database architectures are not new. Teradata, Tandem, Britton Lee, and Sequent offered MPP SQL-based architectures in the 1980s. The re-emergence of MPP data warehouses has been aided by open source and commodity components. Advances in technology have reduced costs and improved performance in storage devices, multi-core CPUs and networking components. Open source RDBMS products, such as Ingres and PostgreSQL, reduce software license costs and allow DW appliance vendors to focus on optimisation rather than providing basic database functionality. Open source Linux provides a stable, well-implemented OS for DW appliances.

History

Many consider Teradata's initial product as the first DW appliance (or Britton-Lee's, but Britton Lee—renamed ShareBase—was acquired by Teradata in June, 1990). Some regard Teradata's current offerings as still being other appliances, while others argue that they fall short in ease of installation or administration. Interest in the data warehouse appliance category is generally dated to the emergence of Netezza in the early 2000s.

More recently, a second generation of modern DW appliances has emerged, marking the move to mainstream vendor integration. IBM integrated its InfoSphere Warehouse (formerly DB2 Warehouse) with its own servers and storage to create the IBM InfoSphere Balanced Warehouse. Other DW appliance vendors have partnered with major hardware vendors to help bring their appliances to market. DATAllegro partners with EMC and Dell and implements open source Ingres on Linux. Greenplum has a partnership with Sun Microsystems and implements Bizgres (a form of PostgreSQL) on Solaris using the ZFS file system. HP Neoview has a wholly-owned solution and uses HP NonStop SQL.

Recently, the market has seen the emergence of data warehouse bundles where vendors combine their hardware and database software together as a data warehouse platform. The Oracle Optimised Warehouse Initiative combines the Oracle Database with the industry's leading computer manufacturers Dell, EMC, HP, IBM, SGI and Sun Microsystems. Oracle's Optimised Warehouses are pre-validated configurations and the database software comes pre-installed, though some analysts differ as to whether these should be regarded as appliances.

Benefits

Reduction in Costs

The total cost of ownership (TCO) of a data warehouse consists of initial entry costs, on-going maintenance costs and the cost of increasing capacity as the data warehouse grows. DW appliances offer low entry and maintenance costs. Initial costs range from $10,000 to $150,000 per terabyte, depending on the size of the DW appliance installed.

The resource cost for monitoring and tuning the data warehouse makes up a large part of the TCO, often as much as 80%. DW appliances reduce administration for day-to-day operations, setup and integration. Many also offer low costs for expanding processing power and capacity.
With the increased focus on controlling costs combined with tight IT Budgets, data warehouse managers need to reduce and manage expenses while leveraging their technology as much as possible making DW appliances a natural solution.

Parallel Performance

DW appliances provide a compelling price/performance ratio. Many support mixed-workloads where a broad range of ad-hoc queries and reports run simultaneously with loading. DW appliance vendors use several distribution and partitioning methods to provide parallel performance. Some DW appliances scan data using partitioning and sequential I/O instead of index usage. Other DW appliances use standard database indexing.

With high performance on highly granular data, DW appliances are able to address analytics that previously could not meet performance requirements.

Reduced Administration

DW appliances provide a single vendor solution and take ownership for optimising the parts and software within the appliance. This eliminates the customer's costs for integration and regression testing of the DBMS, storage and OS on a terabyte scale and avoids some of the compatibility issues that arise from multi-vendor solutions. A single support point also provides a single source for problem resolution and a simplified upgrade path for software and hardware.

The care and feeding of DW appliances is less than many alternate data warehouse solutions. DW appliances reduce administration through automated space allocation, reduced index maintenance and in most cases, reduced tuning and performance analysis.

Built-in High Availability

DW appliance vendors provide built-in high availability through redundancy on components within the appliance. Many offer warm-standby servers, dual networks, dual power supplies, disk mirroring with robust failover and solutions for server failure.

Scalability

DW appliances scale for both capacity and performance. Many DW appliances implement a modular design that database administrators can add to incrementally, eliminating up-front costs for over-provisioning. In contrast, architectures that do not support incremental expansion result in hours of production downtime, during which database administrators export and re-load terabytes of data. In MPP architectures, adding servers increases performance as well as capacity. This is not always the case with alternate solutions.

Rapid Time-to-Value

Companies increasingly expect to use business analytics to improve the current cycle. DW appliances provide fast implementations without the need for regression and integration testing. Rapid prototyping is possible because of reduced tuning and index creation, fast loading and reduced needs for aggregation in some cases.

Application Uses

DW appliances provide solutions for many analytic application uses, including:

- enterprise data warehousing
- super-sized sandboxes isolate power users with resource intensive queries
- pilot projects or projects requiring rapid prototyping and rapid time-to-value
- off-loading projects from the enterprise data warehouse; i.e. large analytical query projects that affect the overall workload of the enterprise data warehouse
- applications with specific performance or loading requirements
- data marts that have outgrown their present environment
- turnkey data warehouses or data marts
- solutions for applications with high data growth and high performance requirements
- applications requiring data warehouse encryption

Trends

The DW appliance market is shifting trends in many areas as they evolve

- Vendors are moving toward using commodity technologies rather than proprietary assembly of commodity components.
- Implemented applications show usage expansion from tactical and data mart solutions to strategic and enterprise data warehouse use.
- Mainstream vendor participation is now apparent.
- With a lower total cost of ownership, reduced maintenance and high performance to address business analytics on growing data volumes, most analysts believe that DW appliances will gain market share.

3.11 The Future of Data Warehousing

Data warehousing, like any technology niche, has a history of innovations that did not receive market acceptance.

A 2007 Gartner Group paper predicted the following technologies could be disruptive to the business intelligence market.

- Service Oriented Architecture
- Search capabilities integrated into reporting and analysis technology
- Software as a Service
- Analytic tools that work in memory
- Visualisation

Another prediction is that data warehouse performance will continue to be improved by use of data warehouse appliances, many of which incorporate the developments in the aforementioned Gartner Group report.

Finally, management consultant Thomas Davenport, among others, predicts that more organisations will seek to differentiate themselves by using analytics enabled by data warehouses.

4.0 CONCLUSION

Data warehouse is now emergimg as very important in database management systems. This is as a result of the growth in the database of large corporations. A data warehouse now makes it easier for the holding of data while in use. However, there are challenges that are constraints in the acceptance and implementation of data warehouse, which is normal in the development of any concept. The future of data warehouse is good as some organisations will opt for it.

5.0 SUMMARY

- A data warehouse is a repository of an organisation’s electronically stored data. Data warehouses are designed to facilitate reporting and analysis.
- The concept of data warehousing dates back to the late-1980s when IBM researchers Barry Devlin and Paul Murphy developed the “business data warehouse”.
- Architecture, in the context of an organisation’s data warehousing efforts, is a conceptualisation of how the data warehouse is built.
- There are two leading approaches to storing data in a data warehouse - the dimensional approach and the normalised approach.
- Another important decision in designing a data warehouse is which data to conform and how to conform the data.

- Ralph Kimball, a well-known author on data warehousing, is a proponent of the bottom-up approach to data warehouse design.
- Operational systems are optimised for preservation of data integrity and speed of recording of business transactions through the use of database normalisation and an entity-relationship model.
- Organisations generally start off with relatively simple use of data warehousing. Over the time, more sophisticated use of data warehousing evolves.
- A data warehouse appliance is an integrated set of servers, storage, OS, DBMS and software specifically pre-installed and pre-optimised for data warehousing
- Data warehousing, like any technology niche, has a history of innovations that did not receive market acceptance.

6.0 TUTOR-MARKED ASSIGNMENT

1. Discuss the benefits associated with the use of data warehouse.
2. Identify five technologies that could be disruptive to business intelligent market.

7.0 REFERENCES/FURTHER READING

Caldeira, C. (2008). “Data Warehousing - Conceitos e Modelos”. Edições Sílabo.

Davenport, T. & Harris, J. (2007). “Competing on Analytics: The New Science of Winning”. Harvard Business School Press. ISBN 1--3S2 — DataBase Management System Services»Blog Archive » Data warehouse appliances – fact and fiction.

Ericsson, R. (2004). “Building Business Intelligence Applications with NET”. (1st ed.). Charles River Media. pp. 28-29.

Inmon, W. H. (1995). Tech Topic: What is a Data Warehouse? Prism *Solutions. Volume 1.*

Kimball, R. & Ross, M. (2002). “The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling”. pp. 310. Wiley. 2nd Ed ISBN 0-471-20024-7.

Pendse, N. & Bange, C. “The Missing Next Big Things”,

Queries from Hell blog » When is an Appliance not an Appliance?

Schlegel, K. (2007). “Emerging Technologies Could Prove Disruptive to the Business Intelligence Market”, Gartner Group.

Todd, W. (1990). “Teradata Corp. Suffers First Quarterly loss in Four years”. Los Angeles Business Journal.

Yang, J. (1998). Warehouse Information Prototype at Stanford (WHIPS). Stanford University.

UNIT 3 DATABASE ADMINISTRATOR AND ADMINISTRATION

CONTENTS

1.0 INTRODUCTION

A Database Administrator (DBA) is a person who is responsible for the environmental aspects of a database. In general, these include:

- recoverability - Creating and testing Backups
- integrity - Verifying or helping to verify data integrity
- security - Defining and/or implementing access controls to the data
- availability - Ensuring maximum uptime
- performance - Ensuring maximum performance
- development and testing support - helping programmers and engineers to efficiently utilise the database.

The role of a database administrator has changed according to the technology of database management systems (DBMSs) as well as the needs of the owners of the databases. For example, although logical and physical database designs are traditionally the duties of a database **analyst or database designer, a DBA may be tasked to perform those** duties.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- explain who a database administrator is
- identify the various functions of database administrator
- state the different types of database administration
- describe the nature of database administration
- specify the tools used in database administration.

3.0 MAIN CONTENT

3.1 Duties of Database Administrator

Society's heritage has been presented on many different materials, including stone, vellum, bamboo, silk, paper and etc. Now, a large quantity of information exists in digital forms, including emails, blogs, social networking websites, national elections websites, web photo albums, and sites which change their content over time. According to a report by the US Library of Congress, 44% of the sites available on the internet in 1998 had vanished one year later.

SELF ASSESSMENT EXERCISE

Explain the role of data base administration in the security of data.

3.2 Typical Work Activities

The work of database administrator (DBA) varies according to the nature of the employing organisation and level of responsibility associated with the post. The work may be pure maintenance or it may also involve specialising in database development.

Typical responsibility includes some or all of the following:

- establishing the needs of the users and monitoring users access and security
- monitoring performance and managing parameters to provide fast query responses to ‘front end’ users
- mapping out the conceptual design for a planned database in outline
- considering both back end organisation of data and front end accessibility for the end user
- refining the logical design so that it can be translated into specific data model

- further refining the physical design to meet systems storage requirements
- installing and testing new versions of the database management system
- maintaining data standards including adherence to the Data Protection Act
- writing database documentation, including data standards, procedures and definitions for the data dictionary (metadata)
- controlling access permissions and privileges
- developing, managing and testing backup recovery plans
- ensuring that storage , archiving, and backup procedures are functioning properly
- capacity planning
- working closely with IT project manager, database programmers, and web developers
- communicating regularly with technical applications and operational staff to ensure database integrity and security
- commissioning and installing new applications

Because of the increasing level of hacking and the sensitive nature of data stored, security and recoverability or disaster recovery has become increasingly important aspects of the work.

3.3 Database Administrations and Automation

Database administration is the function of managing and maintaining database management systems (DBMS) software. Mainstream DBMS software such as Oracle, IBM DB2 and Microsoft SQL Server need ongoing management. As such, corporations that use DBMS software often hire specialised IT (Information Technology) personnel called Database Administrators or DBAs.

3.3.1 Types of Database Administration

There are three types of DBAs:

1. Systems DBAs (sometimes also referred to as Physical DBAs, Operations DBAs or Production Support DBAs)
2. Development DBAs
3. Application DBAs

Depending on the DBA type, their functions usually vary. Below is a brief description of what different types of DBAs do:

Systems DBAs usually focus on the physical aspects of database administration such as DBMS installation, configuration, patching, upgrades, backups, restores, refreshes, performance optimisation, maintenance and disaster recovery.

Development DBAs usually focus on the logical and development aspects of database administration such as data model design and maintenance, DDL (Data Definition Language) generation, SQL writing and tuning, coding stored procedures, collaborating with developers to help choose the most appropriate DBMS feature/functionality and other pre-production activities.

Application DBAs are usually found in organisations that have purchased 3rd party application software such as ERP (Enterprise Resource Planning) and CRM (Customer Relationship Management) systems. Examples of such application software include Oracle Applications, Siebel and PeopleSoft (both now part of Oracle Corp.) and SAP. Application DBAs straddle the fence between the DBMS and the application software and are responsible for ensuring that the application is fully optimised for the database and vice versa. They usually manage all the application components that interact with the database and carry out activities such as application installation and patching, application upgrades, database cloning, building and running data cleanup routines, data load process management, etc.

While individuals usually specialise in one type of database administration, in smaller organisations, it is not uncommon to find a single individual or group performing more than one type of database administration.

3.3.2 Nature of Database Administration

The degree to which the administration of a database is automated dictates the skills and personnel required to manage databases. On one end of the spectrum, a system with minimal automation will require significant experienced resources to manage; perhaps 5-10 databases per DBA. Alternatively, an organisation might choose to automate a significant amount of the work that could be done manually therefore reducing the skills required to perform tasks. As automation increases, the personnel needs of the organisation splits into highly skilled workers to create and manage the automation and a group of lower skilled "line" DBAs who simply execute the automation.

Database administration work is complex, repetitive, time-consuming and requires significant training. Since databases hold valuable and mission-critical data, companies usually look for candidates with multiple years of experience. Database administration often requires DBAs to put in work during off-hours (for example, for planned after hours downtime, in the event of a database-related outage or if performance has been severely degraded). DBAs are commonly well compensated for the long hours.

3.3.3 Database Administration Tools

Often, the DBMS software comes with certain tools to help DBAs manage the DBMS. Such tools are called native tools. For example, Microsoft SQL Server comes with SQL Server Enterprise Manager and Oracle has tools such as SQL*Plus and Oracle Enterprise Manager/Grid Control. In addition, 3rd parties such as BMC, Quest Software, Embarcadero and SQL Maestro Group offer GUI tools to monitor the DBMS and help DBAs carry out certain functions inside the database more easily.

Another kind of database software exists to manage the provisioning of new databases and the management of existing databases and their related resources. The process of creating a new database can consist of hundreds or thousands of unique steps from satisfying prerequisites to configuring backups where each step must be successful before the next can start. A human cannot be expected to complete this procedure in the same exact way time after time - exactly the goal when multiple databases exist. As the number of DBAs grows, without automation the number of unique configurations frequently grows to be costly/difficult to support. All of these complicated procedures can be modeled by the best DBAs into database automation software and executed by the standard DBAs. Software has been created specifically to improve the reliability and repeatability of these procedures such as Stratavia's Data Palette and GridApp Systems Clarity.

3.3.4 The Impact of IT Automation on Database Administration

Recently, automation has begun to impact this area significantly. Newer technologies such as HP/Opsware's SAS (Server Automation System) and Stratavia's Data Palette suite have begun to increase the automation of servers and databases respectively causing the reduction of database related tasks. However, at best this only reduces the amount of mundane, repetitive activities and does not eliminate the need for DBAs. The intention of DBA automation is to enable DBAs to focus on more proactive activities around database architecture and deployment.

3.3.5 Learning Database Administration

There are several educational institutes that offer professional courses, including late-night programs, to allow candidates to learn database administration. Also, DBMS vendors such as Oracle, Microsoft and IBM offer certification programs to help companies to hire qualified DBA practitioners.

4.0 CONCLUSION

Database management system (DBMS) is so important in an organisation that a special manager is often appointed to oversee its activities. The database administrator is responsible for the installation and coordination of DBMS. They are responsible for managing one of the most valuable resources of any organisation - its data. The database administrator must have a sound knowledge of the structure of the database and of the DBMS. The DBA must be thouroughly conversant with the organisation, its system and the information need of managers.

5.0 SUMMARY

- A database administrator (DBA) is a person who is responsible for the environmental aspects of a database
- The duties of a database administrator vary and depend on the job description, corporate and Information Technology (IT) policies and the technical features and capabilities of the DBMS being administered. They nearly always include disaster recovery (backups and testing of backups), performance analysis and tuning, data dictionary maintenance, and some database design.
- Techniques for database performance tuning have changed as DBA's have become more sophisticated in their understanding of what causes performance problems and their ability to diagnose the problem
- The work of database administrator (DBA) varies according to the nature of the employing organisation and level of responsibility associated with post.
- Database administration is the function of managing and maintaining database management systems (DBMS) software.
- The degree to which the administration of a database is automated dictates the skills and personnel required to manage databases

6.0 TUTOR-MARKED ASSIGNMENT

1. Discuss briefly the recoverability duty of a database administrator.
2. Mention the types of database administrations.

7.0 REFERENCES/FURTHER READING

Association for Computing Machinery SIGIR Forum archive Volume 7, Issue 4

How Database Systems Share Storage

The origins of the data base concept, early DBMS systems including IDS and IMS, the Data Base Task Group, and the hierarchical, network and relational data models are discussed in Thomas Haigh, "A Veritable Bucket of Facts:' Origins of the Data Base Management System, (2006). " ACM SIGMOD Record 35:2 .

UNIT 4 CASE STUDY: EVALUATING CALOUNDRA CITY COUNCIL'S ELECTRONIC DATA MANAGEMENT SYSTEM CLASSIFICATION

CONTENTS

1.0 INTRODUCTION

Caloundra City Council recently selected an Electronic Document Management System (EDMS) to meet their document and records management needs. Early on in the project, the Council recognised that the success of the EDMS would largely be determined by the usability of the classification scheme (or file structure) they implemented.

The Council knew that deploying a classification with a high level of usability would enable staff to efficiently and effectively retrieve, as well as file documents. On the other hand, if the classification did not match the way staff think about their jobs and the documents they produce, the result would be inconsistent filing of documents and a high level of frustration when staffs try to search for them. In short, the effectiveness of the EDMS as a business and information management tool would be in jeopardy.

At the time of the EDMS implementation, we were assisting Caloundra City Council with their intranet, this led to us being asked to coach the project team in evaluating the usability of the EDMS classification. The Council had selected Keyword for Councils as a starting point, as it was developed specifically for local government and is used by Queensland State Archives as the basis for the appraisal of records.

This case study sets out the approach taken to evaluate the Keyword for Councils classification for Caloundra City Council, as well as the findings and recommendations made. The case study also provides a brief definition of Electronic Document Management Systems and an overview of Keyword for Councils.

2.0 OBJECTIVES

By the end of this unit, you should be able to:

- describe the Caloundra City Council and records management keeping there
- define e-records management systems
- identify the keywords in the Councils e-records management
- explain why evaluate Keyword for Councils?
- describe the coaching and mentoring roles adopted by Caloundra City Council
- extrapolate the case of Caloundra City Council on how to conduct the usability test for a records system.

3.0 MAIN CONTENT

3.1 Caloundra City Council

Caloundra City Council is responsible for the management of the Caloundra City District, located in the southeast corner of Queensland and the southern gateway to the Sunshine Coast, one of Queensland's leading tourist destinations. Caloundra City has a population of approximately 85,000, and like many local councils in Australia, the Council is responsible for a vast array of activities, everything from water management to eco-tourism and building certification. The Council itself has over 600 staff located within the heart of the city as well as throughout the district. As part of the Council's corporate plan, there are significant initiatives to improve the Council's information management systems and processes. Part of this includes the implementation of an EDMS.

3.2 Records Management at the Council

Caloundra City Council has a well established records management team that centrally manages Council records.

Paper documents are provided to Records Management, which converts the documents to microfiche. The details of the documents are then entered into a records management system, by attaching documents to appropriate files. This process involves classifying each document so that it can be retrieved at a later date.

The records management system has a large subject based classification, running into hundreds of classifications. This is workable when there is a central team classifying documents, but poses problems if some of the document management responsibilities are to be decentralised.

Currently, staff members are able to access the records management system to look up documents they wish to request. Some staff members are frequent users of the system and therefore have a level of familiarity with the current classification. However, there are many staff members who do not use the system and rely on Records Management to do the searching for them.

In addition to the records management system, Caloundra City Council uses other software solutions to manage their human resource functions, properties and main council activities. As these systems are core to their business, they hold documents that are not in the records management system.

Faced with documents held in multiple systems and no way to manage documents electronically, the Council saw the opportunity to integrate an EDMS with their core business systems. For example, if a customer service officer was to enter an application into the front-office system, a record would automatically be created in the EDMS and populated with the appropriate metadata.

This integration would address the reluctance of staff to enter data into multiple systems and the loss of productivity and data integrity that such double-handling presents.

3.3 What is an EDMS?

Electronic Document Management System enables organisations to store documents electronically, including scanned images and documents created on computers, such as word processing files, spreadsheets, and graphics.

An EDMS allows documents to be retrieved, shared, tracked, revised and distributed. It can include workflow encapsulating business rules, and has the facility to capture metadata about documents.

A complete EDMS includes various capabilities, such as document imaging, optical character recognition, and text retrieval.

3.4 What is Keyword for Councils?

Keyword for Councils is a thesaurus for local government, designed for use in classifying, titling and indexing council records independent of the technological environment.

It was developed in 2000 by State Records New South Wales and contains a business classification scheme. The classification describes and groups all of the business activities performed by councils, including general administrative activities.

As the business classification scheme is used as the basis for the disposal and retention of records for local government, it has been adopted by Queensland State Archives.

Keyword for Councils classification is hierarchical in nature and contains three levels. The first level of the classification lists council functions, then council activities related to those functions, and finally the subjects related to the activities. The levels are explained more fully in Table 1 below.

Table 1: Keyword for Councils Classification

Level	Component	Description
1	Keyword	Lists functions of local government, such as council properties.
2	Activity	Lists activities performed within the function. For example, acquisition and disposal are activities related to the function of council properties.
3	Subject	Defines the subject content for activities. For example, under acquisition and disposal, the

subjects are dedications of land, tenements, title deeds and so on.

The subject listings are not designed to include every possible subject descriptor. Rather, it is offered as a guide to councils.

When subject descriptors are insufficient for a particular activity, Keyword for councils recommends that councils add their own. Likewise, subjects tied to an activity can be used with other activities considered appropriate by the council implementing the classification.

Given the three tier classification, a title deed related to the acquisition of a council property would be filed under Council properties/acquisition and disposal/title deeds.

3.5 Why Evaluate Keyword for Councils?

Given that Keyword for Councils was developed specifically for local government, why would a council need to evaluate its suitability for their EDMS? Why not implement the classification as it is?

The answer lies in an appreciation on how the classification will be used and by whom. A function based classification is useful and usable for a team of specialised records management staff, but how will staff with no records management experience respond to functions? This is especially the case when the Council is looking at moving from a subject to a function based classification.

With this understanding, the Council had the foresight to evaluate the classification with real end-users to identify any problem staff would have using the classification. Once armed with this knowledge, the Council would be in a position to make informed decisions about modifying the classification in the way that would best meet their needs.

Modifying the classification could involve:

- changing terminology
- adding levels to the classification
- adding or removing activities and subjects, and to a lesser extent keywords.

The Council would also gain insight into whether the setup or design of the EDMS needed to change to support the use of the classification, as well as the type of metadata needed to support the integration of the EDMS with other corporate systems.

3.6 Coaching and Mentoring Role

Rather than the Council engaging us to plan and conduct the evaluation process for them, we worked in a coaching and mentoring role with the project team.

As our expertise lies in evaluating a range of information systems from the user perspective, rather than records management, we worked closely with the Council's Information Management Unit to come up with the best evaluation approach. We then worked with the Records Manager and a member of the project team to coach them in the approach selected, so that they could conduct the majority of the evaluation internally and have a technique they could apply on future projects.

Finally, we worked with the team to analyse the results of the evaluation and then conducted a final review of their findings and recommendations report.

Using a coaching model was important to the success of the evaluation, as it ensured that Caloundra City Council owned the process and was comfortable with the decisions made at the conclusion of the research.

3.7 Evaluation Objectives

In consultation with the project team, we identified the objectives of the evaluation. The objectives were to:

- validate if Keyword for Councils is both suitable and usable for Caloundra City Council's needs.
- if Keyword for Councils is appropriate, identify areas of the classification that require modification.
- identify whether a fourth level of classification is required beyond the three levels of Keyword for Councils. Although Keyword for Councils only allows for three levels, the EDMS selected by the Council allows for a fourth level.
- identify training and performance support requirements to aid in the implementation of the classification.

It was decided that, during the evaluation, particular attention needed to be paid to the mental models staff employ when retrieving and filing documents. For example, if a staff is looking to retrieve the Employee Code of Conduct, do they first think of the subject (code of conduct), of the activity (employee conditions), or of the council function (personnel)?

Identifying if there is a common pattern in the way staff think that would:

- determine whether Keyword for Councils aligns with staff mental models of filing and retrieving documents.
- determine the best way for staff to use the EDMS classification. For example, are staff more successful at filing and retrieving documents using a search facility, backed by rich document metadata, or drilling down from the top level of the classification.
- provide insight into the type of metadata that needs to be captured at the point of entering records into a corporate system so it can be replicated in the EDMS.

3.8 Approach Taken

To assess the suitability of Keyword for Councils, a usability test was selected as the most appropriate evaluation method. A usability test involves end users attempting scenarios using a system, or in this case a classification, while being observed by a facilitator and other stakeholders. Actions and issues are analysed and changes are made as a result (if required).

In the case of the Council, the usability test involved participants being asked for the types of words they would use to file and retrieve certain documents, as well as where they would file the same documents using level one of Keyword for Councils (functions).

The usability test did not involve staff using the actual EDMS system, as it was not yet available to the project team and we wanted the evaluation to focus on the classification rather than the design of the EDMS.

Detailed below is the approach taken to recruit participants, create scenarios and conduct the usability testing sessions.

3.9 Recruiting Participants

Approximately 20 staff from various areas of the Council were recruited to participate in the usability test. Although the majority of participants worked in an administrative role, around 25% of participants were managers and there was a small representation of records management staff.

Every attempt was made to include staff with varied experience of the current records management system, ranging from never having seen the system to staff who use the system extensively in their job.

Finally, the project team was aware that a number of staff throughout the Council had seen Keyword for Councils. Given this, 50% of the participants recruited had seen the classification, ranging from a quick glance to reviewing the full classification to determine where their documents would be stored.

By recruiting staff with varied experience of the current records management system and Keyword for Councils, we could determine if their prior experience had a positive effect on using the Keyword for Councils classification.

Once the participants were recruited, they were asked to attend a separate 30-minute session, during which they would be asked to complete a number of scenarios based around retrieving and filing documents.

3.10 Creating Scenarios

In preparation for the usability test, six scenarios were created for each participant. They consisted of three scenarios given to all participants and three scenarios related to the participant's job. That way, participants were asked to retrieve and file documents they regularly use as well as documents they are less familiar with.

We also attempted to cover as much of the classification as possible with the scenarios, so that not all the scenarios were concentrated within one or two functions or activity types.

Each scenario was written on a small filing card and allocated a unique letter. The letter was used to identify the scenario during the analysis of the testing results.

Examples of scenarios include:

- an election is coming up and the Electoral Office has sent a letter reminding the Council to submit their boundary figures. What words would you use to find the boundary figures? (Scenario A)
- you have received a letter complaining that the park over the road has not been mowed for six months. What would you file this letter under? (Scenario B)

- someone has lodged a request to be married at Moffat Beach on a certain date. They also want undercover facilities in case of rain. What would you file the request under? (Scenario C)

3.11 Conducting the Usability Testing

As mentioned, each participant was asked to attend a separate 30-minute session. Sessions were broken into two parts:

- Participants were asked for the types of words they would use to file or retrieve documents listed in the scenarios. This was based on the knowledge that the EDMS has a search facility that can be used to retrieve document as well as search for the correct place to file them.
- Using the same scenarios, participants were asked where they would file or retrieve the documents using the first level of Keyword for Councils. This was with a view to evaluating how easy or difficult it is to understand the functional aspect of the classification. See Table 2 below for the list of council functions shown to participants.

To demonstrate how to conduct the sessions, we facilitated the first few, and then handed over to a Council project member to facilitate the remaining sessions. Whoever was not facilitating recorded the responses given by the participants.

The sessions were also attended by the Records Manager, as business owner of the EDMS, who was able to provide background to the project and answer participant questions about the project.

As with all usability testing, participants were told that there is no right or wrong answer and that their responses would help to determine the suitability of Keyword for Councils for the EDMS.

As each scenario was presented to the participant, the words they used to file or retrieve the document were written down. If they mentioned more than one word or phrase, we asked them to state their first choice. Likewise, in the second half of the session, if they could not decide between a number of functions, we asked them for their first choice.

3.12 Analysing the Results

After all the usability testing sessions were completed, we collated the participants' responses and calculated their success rates at retrieving and filing documents based on implementing the Keyword for Councils classification without changes.

Success rates when searching

For the first half of the sessions, where participants were asked what words they would use to file and retrieve documents, success rates were calculated by scoring:

- a full point (1) where the participant's first response is contained in Keyword for Councils, and would be likely to retrieve a match using a keyword based search feature
- half a point (0.5) where a subsequent response is contained in Keyword for Councils
- no points (0) where none of the responses are contained in Keyword for Councils.

During the analysis, we were acutely aware that it is somewhat subjective whether participants would find successful matches, given they were not entering search terms into a live system. However, we felt that the ratings at least provided an indication of success sufficient to assess the suitability of Keyword for Councils.

Once we scored the responses, percentage success rates where calculated for each participant, the average across all participants, and for each of the three common scenarios. For example, the first participant was successful at retrieving or filing documents in five out of the six scenarios, and therefore had a success rate of 83%.

Levels of entry

The results from the first half of the sessions were also analysed to see what level of the classification participants were using as their entry point. That is, were the words they mentioned a function, activity or subject.

Success rates for functions

For the second half of the sessions, where participants were asked where they would file or retrieve documents using the first level of Keyword for Councils (refer to Table 2), success rates were calculated by scoring:

- a full point (1) where the participant's first response matched the Keyword for Councils function
- half a point (0.5) where a subsequent response matched the Keyword for Councils function
- no points (0) where none of the responses matched.

Once again, percentage success rates were calculated for each participant, the average across all participants, and for each of the three common scenarios.

Impact of participant experience

Success rates were then analysed to see if they were impacted by participants' use of the current records management classification or the fact that some participants had viewed Keyword for Councils prior to the usability test.

3.13 Findings

After analysing the usability testing results, the following findings were made.

Usability of functions is low

If staff were required to file and retrieve documents by knowing the first level of Keyword for Councils (functions), their level of successful document retrieval and filing would be low.

In terms of the calculated success rates, in only 52% of cases did participants successfully identify the correct function for a document. Participants:

- had trouble distinguishing between some of the functions, such as Information Management and Information Technology
- did not know what types of documents they would find under certain functions, such as Governance and Commercial Activities
- did not associate certain subjects with a function, for example, very few participants worked out that a marriage permit would be filed under Recreational and Cultural Services.

Staff's Response at subject level

When participants were asked what words or phrases they would use to search for, or file, a document, close to half of all attempts (43%) were at the subject level. This compares to only 8% of responses at the function level and 20% of responses at the activity level.

Searching via word or phrase increases success

When participants searched for a document using a word or phrase, as opposed to navigating down from the function, their success rates generally increased.

Across all participants, the average success rate was 75% compared to a success rate of 52% when identifying the function.

Experience of records management system had no impact on success

Regular and extensive users of the current records management system fared no better in successfully retrieving or filing documents.

Prior viewing of Keyword for Councils had no impact on success

Those participants who had seen Keyword for Councils before the usability test fared no better than staff who had never seen the classification. This demonstrates that providing staff with the complete classification and expecting them to read the document (amid all their other work) is not a feasible approach.

Staff appreciated consultation

A number of staff mentioned that they were pleased to be involved in evaluating Keyword for Councils. In fact, the user-centred approach adopted by Caloundra City Council will help establish an ongoing dialogue with staff, with the view to improving the EDMS classification over time.

3.14 Recommendations

Based on the findings, the following recommendations were made.

Rely on the subject level of Keyword for Councils

Based on the results of the usability test, Keyword for Councils will meet the needs of Caloundra City Council provided that staff do not have to understand level one of the classification (functions).

As the majority of the participants used subject to retrieve or file a document, the EDMS implementation must focus on allowing staff to search by subject.

Modify the subject level

Although participants were reasonably successful when searching using subject descriptors, the average success rate (75%) is still insufficient to guarantee a high level of satisfaction with the EDMS.

To improve the overall success rate, it was recommended that the subject level of Keyword for Councils is modified to more fully support the way the Council does business. This is to be done in close consultation with staff in each of the functional areas, and involves adding new subjects and possibly modifying the terminology of some existing subjects.

Investigate the EDMS searching capabilities

As participants largely searched by subject, the success of the EDMS will depend upon its searching capabilities.

As the EDMS was not available during the planning of the usability test, it is vital that its searching capabilities are assessed. For example, does the EDMS support synonyms? Are the search results easy to understand? What advanced searching capabilities are provided and are they easy to use?

Implement synonym functionality

It was recommended that the EDMS caters for synonyms within its search feature. The need for synonym functionality was highlighted by Scenario C (marriage at Moffat Beach) when only one participant searched on the Keyword for Councils preferred term marriage. Nine other participants used the term wedding.

Do not implement a fourth level

Although the EDMS is capable of implementing a fourth level to the classification, there is no perceived need to do so. Few participants mentioned words that would be candidates for a fourth level, such as the names of banks listed under a credit card subject descriptor.

A fourth level is likely to add complexity to the maintenance of the classification as well as the filing and retrieving of documents.

Usability test once EDMS is set up

Once a working prototype of the EDMS is available, it was recommended that a usability test is conducted prior to launch. This will identify:

- any modification that needs to be made to the EDMS to improve usability
- whether the classification needs to be tweaked further to work effectively within the EDMS
- any issues that need to be included in the training and / or performance support materials.

The usability test will also continue with the user-centred nature of the project to date by involving staff in the next level of the implementation.

Provide training and performance support materials

If the use of the EDMS requires an understanding of the functions, it is critical that this understanding is provided in training and performance support materials, such as cheat sheets.

This understanding will also be important for the small percentage of staff that think at higher levels within the classification.
However, training will not guarantee the success of the EDMS, as a successful implementation is not a training issue. Rather, training is an adjunct to improving the usability of the classification through user-centred design and evaluation techniques.

Given this, possible training and performance support materials include:

- handout providing a high level description of what documents are filed within which functions.
- cheat sheets or laminated handouts showing the path used to file common documents. These cheat sheets need to be tailored to different groups within the Council.

Review the classification on an ongoing basis

An EDMS classification is a living entity. Therefore, it was recommended that the classification is reviewed at regular intervals to improve its match to the way the Council does business and accommodate any changes to the way Council works over the time.
Suggested ways of reviewing the classification include:

- investigate whether search reports can be run from the EDMS, highlighting the most common successful searches and the most common failed searches. Failed searches indicate words or phrases that need to be added as subject descriptors or included in metadata.
- six months after launch review the file structure and look for files paths that are full as well as those that are empty. What does this tell the Council about the classification? Are the empty files that way because the file path (function/activity/subject) is not relevant to the Council or because staff do not understand the terminology? Have the full files become dumping grounds because staff do not understand how to correctly categorise their documents?

The results of the research can be used to continually improve the classification. This will ensure the EDMS does not become stagnant and continues to support the needs of the Council in their day-to-day operations.

3.15 How the Report Was Used

Once the findings and recommendations report was finalised, the Council used the report to:

- hold discussions with the EDMS vendor to ensure that the EDMS could support the recommendations made
- form the basis for an approach to training as well as the development of performance support materials
- plan future sessions with staff to determine how to best modify the subject level of the classification for their needs
- inform the metadata that needs to be captured as part of integrating the EDMS with other corporate systems.

4.0 CONCLUSION

This case study demonstrates how user-centred evaluation techniques can be applied to the world of document and records management.

In the case of Caloundra City Council, a usability test was conducted to assess the suitability of Keyword for Councils for their EDMS classification. The usability test involved asking participants for the words they would use to file and retrieve documents, as well as where they would place certain documents when faced with the list of 32 council functions.

By calculating success rates for document filing and retrieval, we determined that Keyword for Councils would be suitable for the Council on the premise that staff did not have to understand level one of the classification (functions). As staff members largely think at the subject level, the usability of the EDMS search feature was identified as paramount.

The Council has used the results of the usability test to inform many areas of the EDMS implementation, such as the set up of the EDMS itself, modification of the classification, training approach to be adopted, and the integration of the EDMS with other corporate systems.

Finally, the exposure to user-centred techniques has prompted Caloundra City Council to continue consulting with their staff to develop a classification that meets their day-to-day needs. This consultation will not only improve the usability of the classification but will also build buy in to the concept of an EDMS.

5.0 SUMMARY

- Caloundra City Council recently selected an Electronic Document Management System (EDMS) to meet their document and records management needs.
- Caloundra City Council is responsible for the management of the Caloundra City district, located in the southeast corner of Queensland and the southern gateway to the Sunshine Coast, one of Queensland's leading tourist destinations.
- Caloundra City Council has a well established records management team that centrally manages Council records.
- Electronic Document Management Systems enables organisations to store documents electronically, including scanned images and documents created on computers, such as word processing files, spreadsheets, and graphics.
- Keyword for Councils is a thesaurus for local government, designed for use in classifying, titling and indexing council records independent of the technological environment.
- Rather than the Council engaging us to plan and conduct the evaluation process for them, we worked in a coaching and mentoring role with the project team.
- To assess the suitability of Keyword for Councils, a usability test was selected as the most appropriate evaluation method.
- Approximately 20 staff from various areas of the Council were recruited to participate in the usability test
- In preparation for the usability test, six scenarios were created for each participant. They consisted of three scenarios given to all participants and three scenarios related to the participant's job
- As mentioned, each participant was asked to attend a separate 30-minute session. Sessions were broken into two parts.
- After all the usability testing sessions were completed, we collated the participants' responses and calculated their success rates at retrieving and filing documents based on implementing the Keyword for Councils classification without changes.
- Once the findings and recommendations report was finalised, the Council used the report to hold discussions with the EDMS vendor to ensure that the EDMS could support the recommendations made

6.0 TUTOR-MARKED ASSIGNMENT

1. Discuss the evaluation objectives of Caloundra City Council project.
2. Mention 5 recommendations from the Caloundra City Council project.

7.0 REFERENCES/FURTHER READING

Robertson, J. (2004). Evaluating Caloundria City Councils EDMS Classification, Step II Design, Australia.

Tina Calabria, Step II Designs, Sydney Australia.