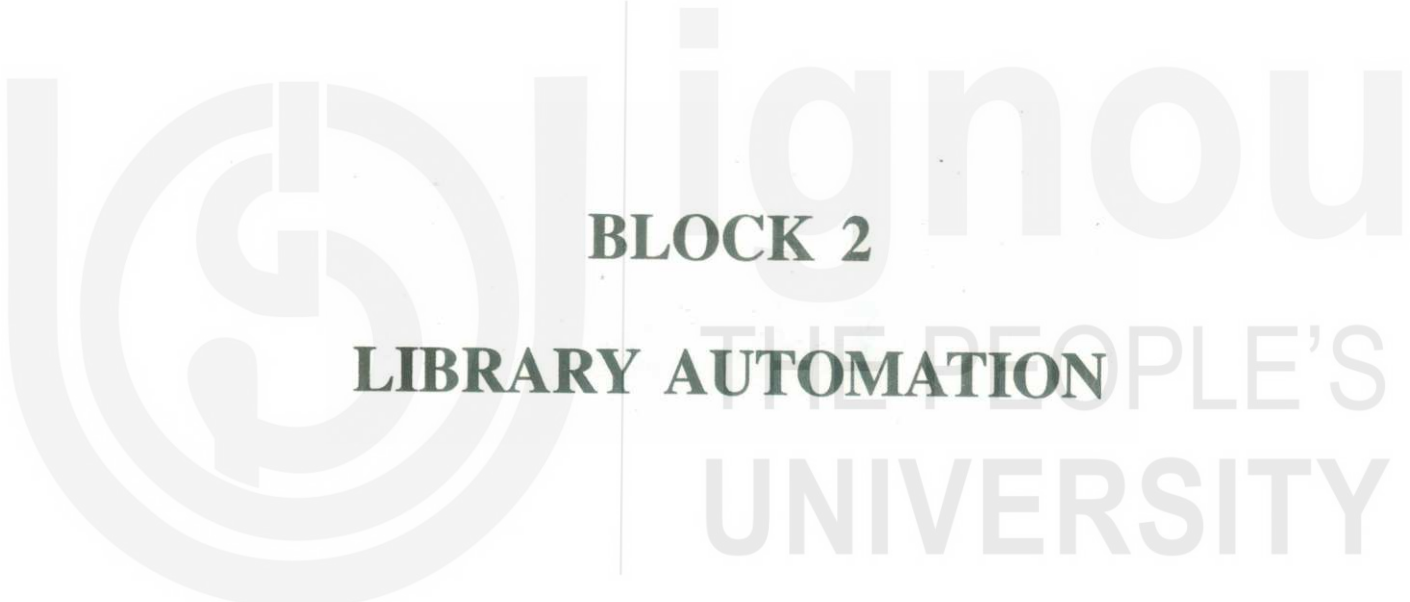


The background of the page features a large, light gray watermark of the Open University logo. The logo consists of a stylized circular emblem on the left, followed by the text 'Open University' in a large, sans-serif font, and 'THE PEOPLE'S UNIVERSITY' in a smaller, all-caps font below it.

BLOCK 2

LIBRARY AUTOMATION

UNIT 5 HOUSEKEEPING OPERATIONS

Structure

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Overview of Library Housekeeping Operations
 - 5.2.1 Acquisition
 - 5.2.2 Processing
 - 5.2.3 Circulation
 - 5.2.4 Serials Control
 - 5.2.5 Maintenance
- 5.3 Historical Perspective of Library Automation
- 5.4 Procedural Model of Library Housekeeping Operations
 - 5.4.1 Procedures and Activities
 - 5.4.2 Analysis of Tasks
- 5.5 Computerised Acquisition Subsystem
 - 5.5.1 Outputs/Reports of Computerised Acquisition Jobs
- 5.6 Computerised Cataloguing Subsystem
 - 5.6.1 Basic Requirements for Catalogue Module
 - 5.6.2 Computerised Cataloguing Jobs
 - 5.6.3 Basic Jobs in Cataloguing
- 5.7 Computerised Serials Control Subsystem
 - 5.7.1 Creation and Maintenance of Master Database
 - 5.7.2 Subscription and Acquisition
 - 5.7.3 Circulation and Binding
 - 5.7.4 Cataloguing and Article Indexing
 - 5.7.5 Information Products
 - 5.7.6 Advantages of Computerised Serials Control Subsystem
- 5.8 Computerised Circulation Subsystem
 - 5.8.1 Functions of a Circulation Subsystem
 - 5.8.2 Activities of Circulation Management
 - 5.8.3 Outputs of Computerised Circulation Jobs
 - 5.8.4 Advantages of Computerised Circulation Subsystem
- 5.9 Summary
- 5.10 Answers to Self Check Exercises
- 5.11 Keywords
- 5.12 References and Further Reading

5.0 OBJECTIVES

After going through this Unit you will be able to:

- understand operational subsystems of a library;
- explain activities and tasks involved in various housekeeping operations;

- streamline library housekeeping operations for more efficient functioning;
- make yourself aware about the application of ICT in housekeeping operations; and
- perform automated housekeeping jobs related to acquisition, cataloguing, serials control and circulation.

5.1 INTRODUCTION

A library of any type or size implies the close collaboration of the trio, i.e., the documents, the users and the library personnel. Library personnel are the matchmakers between the user and his or her document of interest. The services provided by library staff comprise a series of jobs namely, acquiring, preparing and preserving the documents. The activities related to acquisition of documents and making them available to the users, technical processing of acquired documents; circulation and maintenance of processed documents are known as housekeeping operations. These activities are highly labour-intensive and basically routine clerical chores performed slowly and expensively by human beings, until recently. With the advent of Information and Communication Technologies (ICT), libraries are increasingly attempting to automate some of these activities for minimising human clerical routines and thereby making library staff more productive and enabling them to provide fast and comprehensive information through services.

5.2 OVERVIEW OF LIBRARY HOUSEKEEPING OPERATIONS

The performance of a library largely depends on the organisation of its housekeeping operations. Most of the activities related to library housekeeping follow some definite routines and obviously amenable to computerisation. It means a computer or a group of computers can perform routine clerical chores quickly and cheaply. The basic functions related to housekeeping operations in a library irrespective of type or size may be grouped as follows:

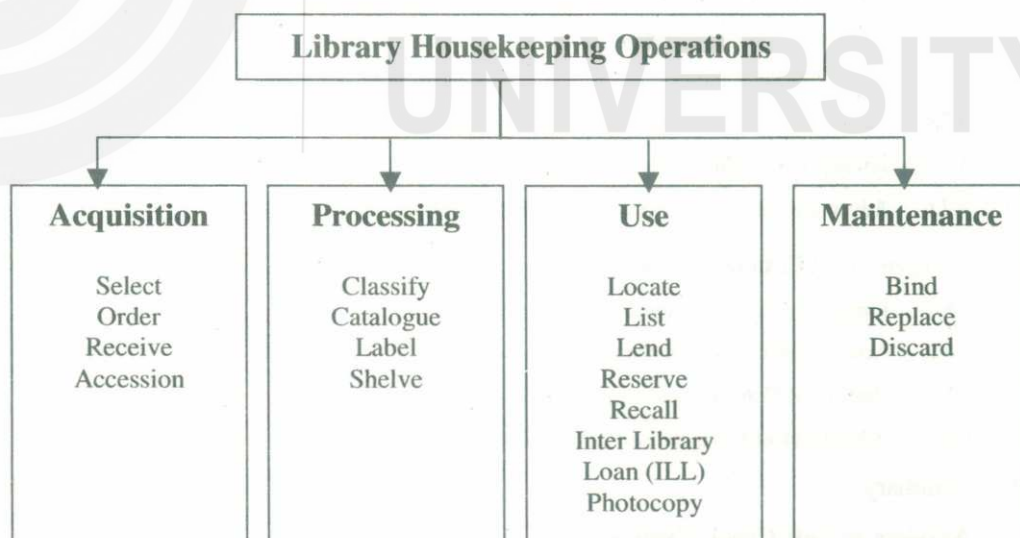


Fig. 5.1: Divisions of Library Housekeeping Operations

The activities of these functional groups are entirely different from other divisions but they are closely connected in a sequence and the combined efforts lead to better library services. It can be seen that libraries are complex systems that include subsystems and components. The main two subsystems are the operational subsystem and the administrative subsystem. Library housekeeping operations are part of the operational subsystem. According to an analytical study of ASLIB (Association for Information Management, UK), the operational subsystem may be divided into four further subdivisions,

namely, Acquisition, Processing, Use and Maintenance. Within each of these divisions there are a number of procedures and within each procedure there are one or more of six possible activities. Such subsystems procedures and activities related is monogroups in a Library are listed the Table-5.1.

Table 5.1: Procedures and Activities of Library System

System	Subsystems	Operational Subsystems	Procedures	Activities (Common to all Procedures)
Library System	Operational Subsystem	Acquisition	Select Order Receive Accession	<i>Initiate</i> (To commence a procedure)
		Processing	Classify Catalogue Label Shelve	<i>Authorize</i> (To approve a procedure)
		Use	Locate List Lend/Issue Reserve Recall/Return ILL (Inter Library Loan) Photocopy	<i>Activate</i> (To implement a procedure through appropriate action)
		Maintenance	Bind Replace Discard	<i>Record</i> (To record what action has been taken)
	Administrative Subsystem			<i>Report</i> (To notify staff or user about the action taken) <i>Cancel</i> (To stop a procedure or undoing an action)

5.2.1 Acquisition

Acquisition of documents is a basic function of libraries. A library should acquire and provide all the relevant documents to its users so that the basic functions of the library are fulfilled. An acquisition subsystem shall perform four basic procedures – Select, Order, Receive and Accession. These procedures are explained below:

- Select:** 'Selection of documents for a library is a very responsible job and should be based on definite principles.' It is done with the help of selection tools (such as bibliographies, publishers' catalogues, book references, etc.) and requests/suggestions from library users received. Purchase orders are sent after formal sanction of the competent authority/library is obtained.
- Order:** Before placing purchase orders with the suppliers each title to be ordered is checked with the catalogue and supply orders sent earlier to avoid duplication in ordering them. In the next stage purchase orders are generated and placed either directly to the respective publishers or to the approved vendors/book sellers. Additionally, generation of reminders for overdue supplies and cancellation of orders is also done as necessary.
- Receive:** When the documents ordered are received the bills covering the supplies are checked with the purchase order list before processing for payment. Care is to be taken to ensure the details (the author, title, edition; imprints etc.) of the books supplied are the same as per the purchase order.
- Accession:** A stock register is maintained by libraries in which all the documents purchased or received in exchange or as gift are entered. Each document is provided

with a consecutive serial number. The register is called Accession Register and the serial number of the document is referred as Accession Number.

All these procedures and related activities of the acquisition subsystem can be mechanised through suitable library management software. In such a library automation system these basic activities are linked with the files of publishers, suppliers, budget and fund accounting, currency etc. to achieve the benefit of an integrated library system.

5.2.2 Processing

The processing procedure is the pivot around which all the housekeeping operations revolve in a library. It helps in the transformation of a library collection into serviceable resources. The procedures under this subdivision are classification, cataloguing, labelling and shelving.

- a) **Classify:** Classification is grouping of similar objects together. It forms the foundation of librarianship. The following are the important classification schemes, which are used in various libraries across the world - Dewey Decimal Classification Scheme (DDC), Universal Decimal Classification Scheme (UDC), Library of Congress Classification (LC), Colon Classification (CC) and Subject Classification (SC) etc. There are two purposes for classification of documents:

- To help the user to find a document whose call number he/she knows
- To find out all the documents on a given subject though collocation

Classification of documents involves analysis of the subject content and representation of the content by class numbers. Thus, it is an intellectual process. So, to automate the classification work, suitable artificial intelligence (AI) techniques are to be used. In India, some research work has been carried out at DRTC, Bangalore for building class numbers (based on Colon Classification) automatically through a software (called Vasya), written in PROLOG (PROgramming in LOGic) – a non-procedural programming language. The present edition of DDC is also available in CDROM and known as Electronic Dewey.

- b) **Catalogue:** A basic task in a library is to prepare a catalogue of its collection and maintain it upto date for meeting the information needs of its users. Cataloguing is the prime method of providing access to the collection. The current practices of cataloguing of resources in Indian libraries are listed in Table 5.2:

Table 5.2: Catalogue Products

Group	Cataloguing Procedure	Product
1.	Manual cataloguing	• Card catalogue
2.	Computerised cataloguing	• Machine readable catalogue • OPAC (Online Public Access Catalogue)
3.	Hybrid Model (Use of computer to produce catalogue on cards and manual filing and also catalogue in machine – readable form)	• Printed catalogue card • Machine readable catalogue • OPAC (Online Public Access Catalogue)

In case of manual cataloguing, the cataloguer makes separate cards for author, title, subject, cross-references and analytical entries by following any standard catalogue code (such as AACR II, CCC, etc.) and file them as per the rules laid down by the library. Computerised cataloguing begins with entering bibliographical data in a pre-designed worksheet. The worksheet or datasheet is very similar to data entry form and is based on any standard content designators scheme (such as MARC 21, CCF, UNIMARC, etc.). Finally bibliographical data recorded in the worksheets are entered into the computer to

produce machine-readable catalogue file and OPAC. Computer-based cataloguing supports importing of bibliographical datasets for the library resources either from centralised cataloguing services or from other libraries and exporting of bibliographical data from its own collection to other library systems. This facility reduces unit cost of cataloguing and ensures standardisation in cataloguing. The recent trend of cataloguing is to utilize Z39.50 protocol to download bibliographical data from other libraries and to provide global access to its own collection through Web-OPAC.

c) **Label:** Labelling involves pasting various labels on different parts of a document. The following labels are generally pasted in books:

- **Spine label:** This is done to make call number (a combination of class number and book number) properly visible to the users when the book is shelved. The size of the label is in the range of 1.25" × 1.25".
- **Ownership slip/mark:** These are generally pasted on the inner side of the front cover at left hand top most corner. Ownership marks are also put at various parts of a document by rubber stamps. The size of slip/mark is 3" × 2.5".
- **Date slip:** It is pasted on the top most portion of the front or back flyleaf of each book. The size of date slip is 5" × 3".
- **Book pocket:** On the bottom of the inner right side of the front cover or back cardboard cover a book pocket is pasted.
- **Book card:** One printed/hand-written book card of size 5" × 3" is put in the book pocket of each book.

In a computerised environment, various labels are printed by using library management software. In case of barcode-based computerised circulation, accession numbers of documents are converted into barcodes and printouts of barcodes are pasted on the inner back cover of documents.

d) **Shelve:** Shelving is the arrangement of documents on the shelves to fulfill the fourth law of library science – ‘Save time of the reader’. Generally, books are arranged on the shelves in a classified order as per the call number. Bound periodicals are generally shelved alphabetically by title and then by volume numbers.

5.2.3 Circulation

Most libraries lend books and other library materials to be read elsewhere by users. This function requires some sort of record keeping arrangement of what has been lent and to whom.

A rich variety of systems of record keeping of loans have arisen out of such needs and these are known as circulation systems. These include some common jobs for successful operations such as enrolment of members, issue and return of library documents, reservation of documents, renewal of documents, maintenance of documents and records, maintenance of statistics, interlibrary loan, issuing of gate pass, etc. In a computer-based circulation system, the machine-readable file consists of records for all items on loan from the library and it is updated periodically with new records. This file is called “transaction file” and it takes required data from other two files – “document file” and “borrower file”. Modern library management softwares support barcode-based circulation system. In such a system a barcode reader scans bar-coded accession number of a document and the barcode in turn acts as a pointer to the document file. It helps to minimise labour and error in data entry operation. The concept of RFID (Radio Frequency IDentification) based circulation system is emerging rapidly in developed countries. It comprises three components: a tag, a reader and an antenna. The tag contains important bibliographical data. The reader decodes the information stored on the chip after receiving it through the antenna and sends data to the central server to communicate to the library automation system. RFID

technology supports user self-checkout machines and has the ability to conduct inventory counts without removing a single book from shelves. As a whole, RFID improves library workflow, staff productivity and customer service with these attributes.

5.2.4 Serials Control

Serials in general and periodicals in particular are essential for research and development (R & D) activities. These are the primary means of communication for the exchange of information and ideas. The periodicals or journals subscribed by libraries can be grouped into these categories:

- Indexing/Abstracting periodicals
- Periodicals containing news items
- Periodicals containing full-text research articles and technical papers.

Acquisition of serials/periodicals in a library is different from book ordering system. In contrast to books, the libraries regularly subscribe to periodicals against advance payment. The modes of subscription of periodicals in a library are as follows:

- Through local vendors/subscription agents
- Through foreign vendors/subscription agents
- Direct from the publishers
- As gift or Complementary
- Through membership
- In exchange.

The fundamental tasks of any serials control system manual or mechanised can be listed as below:

- 1) Selection of serials
- 2) Selection of subscription mode
- 3) Formulation of terms of procurement
- 4) Selection of vendors
- 5) Placing supply orders
- 6) Making advance payment
- 7) Receiving and registration of serials issues in a kardex system.
- 8) Sending reminders for issues not received.
- 9) Making adjustments in advance payments for missing issues
- 10) Preparation of lists of subscribed journals, new arrivals and serials holdings for consultation by users
- 11) Binding and accessioning of back volumes of serials In an automated system all these tasks are performed by library management software efficiently. It reduces workload of library staff. Computer based serials control systems may be predictive or non-predictive. Predictive systems predict the arrival of individual journal issues and can generate reminders for of non-receipts of issues. Prediction means the ability to inform that a particular issue of a particular journal will arrive in the library within a stated time interval. Modern library management software supports predictive mode of serials control with the facilities of online acquisition and access of journals through World Wide Web (WWW).

5.2.5 Maintenance

If we don't take proper care to organise and administer the library documents regularly, these documents would become unserviceable resources soon. The workflow of the maintenance division/section includes following tasks:

- Shelf Rectification : Shelving misplaced documents in proper locations
- Bind : Preserving library resources for posterior and present use
- Replace : Replacing a lost document by the library
- Discard/Withdrawn : Weeding out out-dated, torn and unserviceable documents from the library.

The integrated library automation environment requires information on lost, damaged, missing and withdrawn documents as well as documents sent for binding. These datasets are to be entered to generate and display appropriate messages for the library users and staff. This is also required to generate reports on lost books, missing books, books sent for binding, and so on, for the library administration.

5.3 HISTORICAL PERSPECTIVE OF LIBRARY AUTOMATION

The whole automation process in our society began with a librarian – Dr. John Shaw Billing. Herman Hollerith, a Census Bureau of USA employee, who invented punched card machinery, attributes the idea to a suggestion by Dr. Billing, the then librarian of Surgeon-General's Library (now the National Library of Medicine). Hollerith formed the Tabulating Machine Company in 1896, which later became the International Business Machines (IBM) Corporation. Libraries are an important component of the modern social apparatus for the management of public knowledge. The existing library practices and procedures began to break down under economic, social and bibliographical pressure emerging in the late 1950s and early 1960s. Librarians then began seeking relief in automation. Rapid developments in computer and communication technologies have been catalysing the development of automated library systems. A decade wise analysis of developments related to library automation reveals following facts:

- First there was the pre computer era of unit record equipment
- Then came the offline computerisation in the 1960s and the early 1970s
- This was followed by the online systems of the 1970s.
- The 1980s saw the advent of microcomputers and the emergence of CD-ROM technology.
- Finally, we came to the Internet revolution of the 1990s.

The effects of ICT on libraries and information centers are characterised by:

- Mechanisation – doing what we were already doing, but more efficiently
- Innovation – experimenting with new capabilities, that the ICT makes possible
- Transformation – fundamentally altering the nature of the library operations and services through the capabilities extended by ICT

1936-60: These developments can be further detailed as follows: Punched card for circulation control; Use of IBM 402, 403 or 407, sorting and retrieval of data; Introduction of the concept of 'Memex' by Vannevar Bush in 1945.

1960-70: Use of general-purpose computers that became widely available in the 1960s; H.P. Luhn, in 1961, used a computer to produce the "Keyword in Context" or KWIC index for articles appearing in Chemical Abstracts; Project "MEDLARS" started in 1961. Computerised circulation system first appeared in 1962. Project 'Intrex' which, it was hoped, would provide a design for evolution of a large university library as a new information transfer system, started in 1965. Project

MARC, initiative by Library of Congress to provide a format for machine readable cataloguing data, started in 1965; Introduction of on-line interactive computer system in place of offline batch processing systems began in mid 1960s; Initiation of projects like BALLOTS by Stanford University and MAC by M.I.T. These developments subsequently led to the automation of most of the library operations and services.

1970-80: Minicomputers were introduced to automate circulation and books were bar-coded; Computer based ordering systems were introduced to pass orders to book and serial jobbers. ISBDs started appearing from 1971; OCLC was established in 1971 to facilitate library cooperation and to reduce costs of processing work. ISO-2709 was developed in 1973 as the standard for data exchange format. OCLC started development of Worldcat in 1975 (Worldcat now contains 46 million cataloguing records). Library networks begin appearing all over the world.

1980-90: Shared copy-cataloguing systems by using computer and communication technologies were established as a norm in 1980s. Remote access to online databases became a reality. Appearance of CD-ROM databases on indexing and abstracting journals started in early 1980s; Library automation packages initiated shifting towards relational architecture. Integrated automation packages began appearing in mid 1980s along with bar-coded circulation system. OPAC became very popular in this decade and made available on campus wide LANs for accessing.

1990: Library automation packages started upgrading from client server architecture to web architecture. Large scale developments took place in the areas of resource sharing, union catalogues and computerised inter-library loans. Release of Z39.50 protocol in 1995 to share bibliographical information and to overcome the problems of database searching with many search languages; Formation of collective purchasing consortia that can negotiate prices for all members of the consortium. Emergence of multimedia databases; Emergence of www and web-based library services. Release of Dublin Core Metadata Standard in 1995; Web-OPACs began appearing in almost all automated libraries. Conversion and digitisation of print contents into electronic format started in a big way; Full text access to information resources over Internet started against IP authentication. Integrated access interface emerged to act as one-stop access interface; IFLA introduced FRBR as a conceptual data model for bibliographical databases in 1999. Eprint archives and digital libraries. RFID are being developed in a big way. Smart card based user access to online library services have come into use.

5.4 PROCEDURAL MODEL OF LIBRARY HOUSE KEEPING OPERATIONS

For planning and designing automation of the operations of a library, it is necessary to analyse them in necessary detail. A close analysis of those operations provides us three hierarchical levels – procedures, activities and tasks.

5.4.1 Procedures and Activities

The eighteen procedures listed in Table 5.2 are common to libraries of different types. The procedures under each and every operational subsystem have been analysed by P.A. Thomas in terms of six possible activities – initiate, authorise, activate, record, report and cancel. All of these activities may not be involved in every procedure. There are one or more six possible activities against each procedure. The six common activities are defined as:

- **Initiate** – That which makes it apparent that a procedure should be commenced.
- **Authorise** – In some cases, the decision to carry out a certain procedure must be approved before any further action is taken.

- **Activate** – When a procedure is known to be necessary and in some cases approved, it is usually implemented by taking appropriate actions.
- **Record** – The function that states or records what action has been taken.
- **Report** – To notify library staff or user that an action has been taken.
- **Cancel** – To stop a procedure, in particular the aspect of revoking or undoing an action.

5.4.2 Analysis of Tasks

The third level in the hierarchy is concerned with tasks within an activity under each procedure. Task means a related group of operations carried out to perform a particular kind of job. The analysis of tasks to perform activities within procedures may be done through a set of five primary questions:

- What information is needed for the activity?
- Where is the information to be obtained?
- When is it required?
- Who requires it?
- How is it used? These five questions should be asked to carry out possible activities under each procedure. It provides depth to the framework provided by the procedural model. An example of this approach is shown in Table 4 in the context of five possible activities of book order procedure in acquisition subsystem.
- Sharing and transferring data between different systems and media
- Availability of Distributed Information System (Internet)
- Increased capacity of data storage and data transmission
- Decreased cost and size of equipment
- Increased reliability of hardware and software to perform repetitive jobs
- Introduction of GUI based user friendly software with online help.

Table 5.3: Analysis of Activities for Book Ordering

SYSTEM SUBSYSTEM PROCEDURE	LIBRARY SYSTEM ACQUISITION SUBSYSTEM ORDER				
	INITIATE	AUTHORIZE	ACTIVATE	RECORD	CANCEL
What information?	Author, Title, Sub-title, Edition, Place, Publishers, Date, ISBN etc	Signature of Approval	Library/Branch Library, Date of Order, Order number, Name of Vendor and Bibliographical details etc.	Administrative data, Bibliographic data	Order Number, and Date Vendor, Book details
Where from?	Bibliographies, Index, Requisition, Suggestions	Competent Authority	Book Selection Tools, MIS	Order form/ Order letter	Order File/ Computer Database
When?	After Select Procedure	Before Activation	After Authorisation	After Activation	After Activation
Who?	Library Asst./ Technical Asst.	Librarian/ Section-In-Charge	Library Asst./ Technical Asst	Library Asst./ Library clerk	Library Asst.
How?	Receiving copy of Bibliographies, Suggestion slip	Enter Signature	Enter data/ information on Order form/ Computer Database and Generate Order	Filing the Copy of Order form/ Saving in Computer	Deletion from Database/ Removal from File

Self Check Exercise

1) Explain procedural model of library automation.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.5 COMPUTERISED ACQUISITION SUBSYSTEM

The basic activities of any computerised ordering or acquisition subsystem can be specified as follows:

- To receive indents/suggestions for the documents to be acquired
- To check whether items requested are already in the library or on order
- To obtain approval of the competent authorities for purchase of the indented/suggested documents
- To print orders or dispatch order electronically to supplier/publishers
- To check when orders are overdue
- To follow up overdue orders
- To maintain a file of records of items on order
- To note the arrival of ordered items
- To process bills for payment
- To maintain book fund statistics and accounts
- To generate printed and electronic listing of various reports
- To control currency conversions
- To maintain vendor performance reports and statistics

The acquisition module of modern Library Management Software(LMS) should also carry out the following jobs.

- Accommodate a variety of materials, including but not limited to – monographs, monograph series, annual and cumulative indexes, loose leaf materials, supplements, reports, musical scores, etc.
- Accommodate and identify items in a variety of formats, including but not limited to – print, microform, film, videotape, audio cassette, CD-ROM, magnetic tape, etc.
- Record, store and display - bibliographic information, acquisition type (order, gift, approval etc.), status (reported, received, etc.), library/branch/copy/fund information, invoice information, vendor information, accounting information, requestor information etc.
- Accommodate different types of orders regular order, membership, approval, standing (continuing) order, deposit account etc.

5.5.1 Outputs/Reports of Computerised Acquisition Jobs

The outputs form a computerised acquisition subsystem usually include the following:

- Lists/Reports of titles requested
- List/Report of item(s) from suppliers/publishers
- Item(s) selected for approval
- Item(s) approved by the authority/library committee
- Item(s) rejected in the approval process
- List of gratis item(s) received by library
- Report on request status
- Printout of notes for obtaining approval of the competent authority for screened titles.
- Printout or softcopy of order letters & query letters
- Printout or softcopy of reminder letters
- Printout or softcopy of order cancellation letters
- Printout or softcopy of reordering letters
- Letters/notes for adjustment of advance payments
- Letters/notes to bank for foreign exchange rate
- Reports on order status
- List/reports of overdue supplying of titles
- List/Reports of overdue items
- Lists/reports of titles actually ordered
- Reports of budget commitment
- Lists/Reports of item ordered against advance payment
- List/Reports of titles received against orders
- Letters of intimation (on arrival of documents)
- Printouts of accession register entries
- Printouts of barcode labels
- List of supplier/publishers
- List of currency and exchange rates
- Budget with commitments
- Report of detailed annual budget of library
- Report on amount received in different budget heads
- Report/statistics of vendor performance
- List of recent additions
- Generation of book cards (in case of integrated ordering and cataloguing system)

Self Check Exercise

2) Write down the operations in computerised acquisition subsystem.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.6 COMPUTERISED CATALOGUING SUBSYSTEM

Automated cataloguing systems are useful to control the clerical and technical processes involved, and to promote the exchange and use of cataloguing data locally, regionally, nationally and globally. Over the last 20 years, databases and Online Public Access Cataloguing (OPAC) have gradually replaced card, sheaf and microform catalogues. The catalogue records have become the central bibliographic record for the exchange management system. A computerised catalogue record requires to be standardised for the interchange of cataloguing records. Standardisation and structuring of cataloguing records are based on content designators such as MARC 21 (Machine Readable Cataloguing), CCF (Common Communication Format), UNIMARC (Universal MARC) or other similar standards.

5.6.1 Basic Requirements for Catalogue Module

In view of the recent developments in ICT sector, a modern LMS should provide appropriate facilities in its catalogue module. The basic necessities of the module for machine-readable cataloguing may be formulated as follows:

a) *Bibliographic Requirements*

- Should support a variety of classification schemes and vocabulary control devices (e.g., subject headings lists or thesauri and electronic resource description formats)
- Must enable coverage of standard bibliographic and authority record formats (e.g. MARC, CCF, etc.)
- Must enable records to be exported or imported to and from tape, disk or other devices
- Must enable a bibliographic record to be retrieved and modified at any time, once it has been entered
- Must enable coverage of items like monographs, serials, government documents or any other type of documents.
- Must enable catalogue data to be added, validated, updated and removed online from the workstations
- Must support withdrawal of items, export and import of records

b) *Authority Control Requirements*

- Must enable incorporating changes in the MARC authority format or other national/international standard formats

- Must generate various kinds of references from authority records
- Must accommodate:
 - Personal, corporate and topical names headings in a name authority file
 - Title, uniform title, and series entries in a title authority file
 - Subject headings in a subject authority file

OPAC Requirements

- Must allow both simple and expert searching
- Must support users to enter multiple words or phrases to be searched on one , more than one or all fields
- Should support Boolean operators, within and across all fields, such as
 - OR (either one or both terms must be in the record)
 - XOR (either term, but not both, must be in the record)
 - AND (both terms must be in the same record)
 - NOT (following term must not appear in any record)
- Should support positional operators
 - SAME (terms must be in the same field)
 - WITH (terms must be in the same sentence within a field)
 - NEAR (terms must be adjacent to one another, but in either order)
 - ADJ (terms must be immediately adjacent to one another)
- Should support relational operators (less than, greater than, equal to, etc.)
- Should include user self-service features such as reservation of items, cancellation of reserved items.
- Must enable users to limit searches by: publication year, language, item type, item category, location and access.
- Must support multilingual features
- Must enable searchers to specify which fields are to be displayed and to indicate brief or full display of hits

c) Downloading Requirements

- Must be Z39.50 compliant cataloguing system [ANSI/NISO Z39.50 (1995) or ISO 239.50 (1998)]
- Should enable to capture bibliographic and authority records from any Z39.50 server through Z39.50 client
- Should allow local manipulation of captured data.

d) Reports and Backup Requirements

- Must produce a count of all records added, edited by a specific operator or over a specified time period
- Must generate lists, statistics of items added or tabulated by call number, document type, adjustment location etc.

- Must produce a list of all citations with authority file violations
- Must support backup of all cataloguing records in suitable media (magnetic, optical, etc.) and easy recovery of records at the time of need.

5.6.2 Computerised Cataloguing Jobs

The process of computerised catalogues usually covers the following jobs :

- Titles in Process:** These are the documents, which have been received from the suppliers and technical processing.
- Retrospective Conversion:** This involves preparation of machine-readable catalogue of old and existing library stock. Here the cataloguing data is first entered on a worksheet or datasheet designed by following any standard content designator scheme and then data transferred to the database through manual keying. However, if the cataloguers have enough experience catalogue entries can be directly keyed into the system as per a standard input format.
- Existing Library Catalogue in Machine-Readable Form:** Development of library OPAC started long back in many libraries of the world and at that time integrated library automation packages were not available. As a result, cataloguing data in digital format are available in many libraries, which can be merged into the catalogue database created using the newly installed LMS.
- Union Catalogue:** Union files of the stock of several libraries, or another shared database may be imported, converted into local standard format and finally merged into the catalogue database.
- Commercially Available Files of MARC Records:** In this process records from external databases may be downloaded directly from the relevant websites or CD-ROMS.

5.6.3 Basic Jobs in Cataloguing

The key jobs of a cataloguing module are:

- Authority control
- Data entry
 - For newly acquired document
 - For existing old stock
- Downloading

These jobs should also be supplemented by regular backup of the catalogue database in suitable media. In the family of magnetic storage devices, Digital Audio Tape (DAT) is the most popular in libraries all over the world. DVD-ROM – an optical storage device with a typical storage capacity of 17 GB is coming a big way to replace all other backup media for the storage of library records.

a) Authority File Creation and Maintenance

Authority file is essential to control form of index terms or headings, such as author headings, or subject index terms for better retrieval efficiency. Records in this file may be created locally or drawn from externally available files such as the name and subject authority files of the Library of Congress or other agencies. Alternatively, one or more suitable thesauri may be used to build up the authority file. Library automation packages provide a facility to create and maintain authority file in the catalogue module. This file acts as a master database, where entry is to be made once. This gets reflected in various modules of the package. The master file containing authority entries can be consulted during cataloguing, possibly by display in a separate window and new headings are immediately added to the authority file with an opportunity to review subsequently.

b) Data Entry for Cataloguing

This facility of the catalogue module is utilised for generating bibliographic data elements to the newly acquired documents and also for entering bibliographical data of existing old stock of the library. An integrated automation package uses the bibliographical details created for the acquisition module for the catalogue module also. In the catalogue module the record is standardised by entering additional data elements and rendering of access points with the help of authority file. The transformation of bibliographical data elements of existing stock of any library into machine-readable form is called Retrospective Conversion or simply RECON.

c) Copy Cataloguing and Downloading

In this process, cataloguing as particular document, the cataloguing data is downloaded from an external OPAC and necessary changes in some fields like those of accession numbers, keywords, etc., are made. This approach helps in reducing cataloguing. The OPAC of US Library Congress is perhaps the best source for downloading bibliographic data pertaining to books. For other types of documents like reports, research papers, etc., relevant databases may have to be accessed. This mode of shared cataloguing is popularly termed as copy cataloguing.

Self Check Exercise

3) What are the steps involved in generating computerised catalogue database?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.7 COMPUTERISED SERIALS CONTROL SUBSYSTEM

Management of serials requires frequent and repetitive record addition or amendment. Computerisation is an attractive proposition for serials control because of this reason. The serials control module of any Library Management Software (LMS) should provide the following basic facilities to cope with the complexities of the management of periodicals and other serial publications.

- Handling new subscriptions
- Handling renewal of subscriptions
- Cancellation of subscription
- Budget control
 - Department/unit-wise budget
- Invoice processing
 - Invoice for individual issues, or for annual (or other period) subscription

- Recording the receipt of journal issues
 - Formula for generating expected issues
- Managing (sending claims for) missing issues
 - Sending reminders
- Dealing with “special editions”, supplements, and indexes
- Dealing intelligently with name changes (of publication, publisher) and merges or splits (i.e, one journal becomes two, or two merge together)
- Binding control
- Accessioning bound volumes
 - Barcoding of accession numbers
- Report generation
- Listing titles of the periodicals received (alphabetical and subjectwise)
 - Hyperlinking the e-journals
- Editing & updating of records
- Searching by title, publisher, distributor, sorting by date or volume/issue number
- Option for Web-OPAC
- Table of contents and other personalized information services
- Article indexing
 - Indexing of journal articles by author, title, and subject keywords. There should be option for bulk import in some format
- Union list and union catalogue
 - The module should support creation of union list and union catalogs. The union list is the list of journals currently subscribed in the different affiliated libraries. The union or net worked catalogues the complete holdings information along with all its missing issues, discontinuation in subscription, changes in title, etc.

5.7.1 Creation and Maintenance of Master Database

Modern LMSs are based on relational data model. Each module of these packages centres on a master database. Any number of additions, modifications and deletions are possible in the master database and these changes are automatically reflected in all the sub-modules under that module. Serials control module also follows the same architecture. Records created in the master database are available to all the sub-modules under serials control. It reduces data entry work and ensures standardisation. Generally, this database includes following master entries:

- a) *Title master*: In this file, bibliographical details of new serials are entered after the selection and approval process. This file also includes details of all the existing serials on the basis of some standard bibliographic format. Any number of additions and deletions in the title master file is possible through appropriate options.
- b) *Country master*: This file contains names of countries and their corresponding codes for entering details of country of publication in sub-modules of serials control. Country code is generally based on ISO-3166 where each country is represented by two unique characters e.g. the code of India is *in* as per ISO-3166.
- c) *Language master*: This file contains entries for languages and their three digit codes as per the ISDS manual and CCF manual.

- d) *Supplier/Publisher/Binder master*: This master file contains details of all local and foreign subscription agents, publishers of serials and binders along with their corresponding codes. These codes are generally created locally.

Subject master: This master file holds lists of subject descriptors and are available in various sub-modules for entering subject group(s) of serials. This file also includes class numbers, corresponding to subjects covered.

Frequency master: Individual issues of serials are published at different time intervals e.g., weekly, monthly, semimonthly, bi-monthly, quarterly etc. This master file enables one to record serials frequencies and also to create codes for new serials frequencies.

Budget master: This file enables maintaining various financial data necessary for serials acquisition. This is required for fund accounting of serials budget and includes entries for main budget head, sub-head, period, account name and code, dates, opening balance, grant amount, amount spent, credit note amount, closing balance, etc.

Currency master: This file contains currency description, codes and exchange rate for foreign currencies in terms of Indian rupees. The conversion will be calculated automatically on the basis of data given in this file. Frequent updating of conversion rate is necessary (if required, daily) on the basis of bank notification.

Delivering Mode Master: Publisher and vendors arrange supply of serials by different modes of delivery. These are:

- Air freight (bulk)
- Air Mail
- Shipment/postal
- Hand delivery

Generally publishers charge extra payment for air mail. This file includes different mode of supply for making these information available for data entry work in various sub-modules.

Physical Media Master: Serials are available in many forms, formats and media. The available physical media are print, CD-ROM, magnetic tape, online, DVD-ROM, etc. This file includes description and corresponding codes for different physical media.

Binding type master: Individual parts or issues of journals are bound volume-wise at the end of year or subscription period. Binding of back volumes of journals is a regular work of serials control section of a library. This file contains different modes of binding (e.g., standard, leather binding, cloth and rexin binding etc.) and their corresponding codes.

Letter Master: This file includes formats for every type of letter required for the generation of outputs such as order letter, cancellation of order letter, reminder letter, claim letter, etc. The terms and conditions formulated by the library may be included in the order letter format.

5.7.2 Subscription and Acquisition

This group of activities involves the following basic jobs:

- 1) Selection of serials for new subscription
- 2) Renewal or discontinuation of existing journals/serials
- 3) Selection of delivery mode
- 4) Selection of subscription mode
- 5) Formulation of terms of supply

- 6) Selection of vendors
- 7) Approval from authority
- 8) Ordering and renewal
- 9) Payment
- 10) Receiving and registration
- 11) Reminder generation
- 12) Adjustment of advance payment for non-receipted issues.

All these may be organised into three groups and may be represented diagrammatically

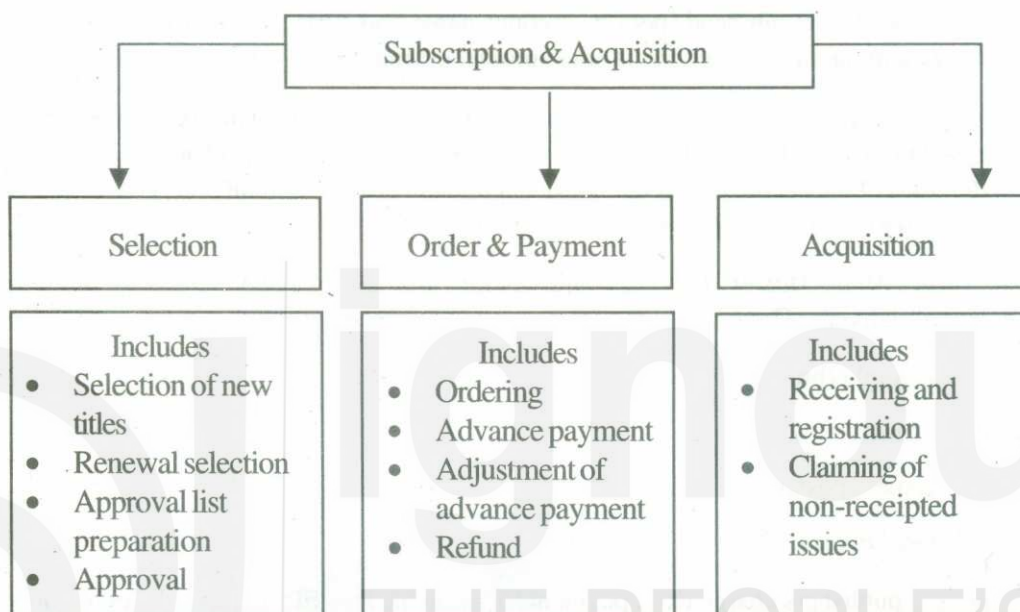


Fig. 5.2: Subscription and Acquisition Processes of a Journal

Some of these initial jobs are similar to book ordering system. The differences and difficulties arise with the continuing nature of subscriptions to serials. Renewals must be decided and communicated at the appropriate time.

5.7.3 Circulation and Binding

These activities usually come under maintenance of periodicals. As a general principle, journals/periodicals should not be issued like books. However, to meet specific and essential requirements in special libraries, bound volumes or loose issues of periodicals may be issued only for short periods overnight or for a day. This restriction is necessary to avoid gaps in journal volumes or issues, as replacements for last issues are not easily available. So, for such short-term issue, the LMS must have provision. Otherwise, the librarian must get the software suitably customised to get this facility incorporated.

5.7.4 Cataloguing and Article Indexing

For providing easy access to the articles published in the journals received in the library, the library or documentation staff builds databases covering the bibliographic details of the articles. For inputting the bibliographic details, a standard data input format is used. Such format can be the one prescribed for MARC, CCF, etc. Library Management Softwares (LMSs) usually have a module or submodule for building article index databases also.

Usually article indexing becomes necessary when externally published indexing and abstracting services do not cover the core journals on discipline of interest.

- a) The absence of publisher produced indexing services is serious enough for a library to consider producing one itself. The first step of article indexing is to determine the bibliographic format of articles to be entered into the system. As per CCF, a bibliographic (input) format for component part in a serial such as article of a periodical should include some mandatory data elements. A library must follow a subject thesaurus for the standardisation of keywords or descriptors. A number of computer generated services like current awareness list (CAL), abstracting journal, subject bibliographies, author bibliographies, SDI service, recent arrival list etc. can be provided to users on the basis of article indexing database.
- b) Binding: A further valuable feature of computer based serials control subsystems is their ability to inform the library staff of volumes that have been completed and are now ready for binding. It is a very helpful feature to assist in work scheduling and to spread the binding workload evenly throughout the year. The binding sub-module includes following functions:
 - Binding set selection: This step involves the work of selecting the relevant issues meant to be sent for binding. The selection follows entering of binding type, colour and embossing type by using drop down menu or as per the provisions.
 - Order process: In this step, the software has to generate order letter giving details of the types of binding no. of volumes for each type of binding, expected date of delivery etc.
 - Receiving: This function includes receiving of bound volumes against binding order checking of the quality of binding and accessioning of bound volumes. It also has provision to enter receipt date classification number, location, etc. manually and to send reminders for defaults in completing the binding jobs and delivering.
 - Payment: Payment work starts with invoice processing, generates payment release note or order and finally records all the details of payment (e.g. cheque/DD number, date, amount, budget head, payment acknowledgement receipt number and date etc.)
 - Reminder: It is meant for generating reminder letters to be sent to the binders in case of any delay in receiving the bound volumes. Binding sub-module should generate list of overdue items.

5.7.5 Information Products

Information products of serials control section of a library are generated in offline mode and form the base of various user services. These products are originated either from article indexing activities or serials catalogue database and produced on demand.

- i) Information Products from Serials Catalogue Database
 - List of recent received journal issues
 - List of journals received available on a particular discipline –wise and alphabetically
 - Holding list of serials-discipline wise and alphabetically
- ii) Information Products from Article Indexing
 - Table of contents service of a group of journals (as per user selection)
 - Compilation of subject bibliographies
 - CAS and SDI services in online and offline mode
 - Online database searching and retrieval of articles
 - Altering services on specific topics (online/off-line)

- Compilation of abstracts bulletin on any given topic
- News compilation services
- Product compilation services

5.7.6 Advantages of Computerised Serials Control Subsystem

Serials management is possibly the most complex process in libraries. It involves a lot of time bound routine activities that require accuracy and precision. In special libraries, periodicals are the most important resources and take the lion's share from the library budget. The optimum utilisation of these resources demands application of ICT as a management tool. The advantages features of automated serials control subsystem are:

- It improves the speed and accuracy in serials control activities
- It can generate various reports in required formats for MIS activities
- It ensures timely reminders generation and better binding control
- It offers easy creation and maintenance of article indexing database and thereby helps in providing a number of user services on demand
- It helps library staff in quick production of serials holdings and list of recent arrivals in many forms.
- It facilitates online access to the serials database from anywhere, anytime in any format
- It predicts the arrival of journal issues and generates schedules for receiving journal issues
- It helps in the export and import of cataloguing data for serials received by the library
- In an integrated serials control module, the master database supplies most of the bibliographical, financial and administrative data at the time of data entry. This approach ensures standardisation of entries and also eliminates data redundancy
- It supports generation of a variety of information products and user services in the desired form and formats

Self Check Exercise

- 4) Comment on the information products generated using computerised serials control subsystem.

Note: i) Write your answer in the space-given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

.....

.....

5.8 COMPUTERISED CIRCULATION SUBSYSTEM

Circulation work in library involves a group of operations that are specific, repetitive and systematic. As a result, automated circulation systems have been fairly successful from the early days of library automation. Such systems require a minimum set of essential data for carrying out circulation activities and data may be captured in a variety of ways. Generally, circulation modules of Library Management Softwares (LMSs) include the facilities meant for Inter Library Loan (ILL), group library circulation (in case of branch libraries and common borrowing card for a group of libraries) and also provides options for the maintenance of library resources. Circulation systems are designed to capture and manipulate three kinds of datasets:

- Information about the borrower
 - Name
 - Address
 - Telephone number
 - Borrower code or identification number
 - Borrower category and privileges.
- Information about the resources being borrowed
 - Call number
 - Resource identification number (e.g., accession number)
 - Title
 - Author
 - Publication year
 - Edition
 - Reservation details.
- Information about the loan itself
 - Date of issue
 - Due date
 - Date of return

In the first set, borrower code is the most important element as it individualises each borrower and their respective category. The resource identification or accession number is the key element in the second set of data. Circulation module of modern LMSs requires no data entry work for the third set. The date of issue is captured from the system date and due date is calculated automatically as per the privileges given to the borrower.

5.8.1 Functions of a Circulation Subsystem

Computerised circulation subsystems generally perform a group of functions utilising three types of information, as listed above, as a base. Apart from the issue and return provision, circulation module in an integrated LMS also performs the following functions:

- To locate circulating documents
 - on loan
 - reserved by user.

- at binding
- being reprocessed
- To identify documents on loan to
 - a particular borrower
 - a specific class of borrowers
- To record 'personal reservations' for documents on loan
- To print recall notices
 - returning overdue documents
 - renewing of items documents
- To calculate fines or overdue charges and generate
 - printout of fine notices
 - receipts for fines payments
- To calculate and print statistical reports
- To extend provision for handling special categories of borrowers and special types of materials
- To generate and print gate pass and due date slips
- To interface with various data capturing devices e.g. barcode readers, smart card and RFID equipment
- To extend facilities for ILL and maintenance activities

In summary we can specify a total of nine functional groups of activities performed by a circulation module and these are:

- Circulation control to perform all charging and discharging activities at a workstation in real time
- Fixing of reservations on any titles that can be issued
- Loan period calculation to determine automatically the loan period for each item at the time it is charged
- Overdue processing to automatically produce appropriate notices for all the outstanding loans
- Cash management function to allow users to be billed automatically for recovering overdue charges and value of lost materials
- Fiscal report generation to record daily financial transactions including fines and fees levied, fines and fees waived and fines and fees collected
- Transaction control to block users who exceed overdue threshold (e.g. too many overdue items) set by the library
- User registration and maintaining borrower profiles
- Report generation to produce lists, reports and statistics in relation to various facets of circulation management.

The transaction or loan database is the core of circulation system. This database comprises a series of records, one for each transaction. Each record includes a brief dataset that specifies details of the document (through document number), details of the user (through membership code) and transaction details (e.g., date of issue and date of return are extracted from the system date, and due date is calculated automatically). With integrated software, the bibliographical details (e.g., author, title edition, place and year of publication) of documents on loan are extracted from the catalogue database and the membership database is utilised for collecting user information. Accession numbers of documents are used as the key data elements in first case, whereas membership codes act as pointer to the member database in the second instance.

The data required for charging and discharging of documents is taken from the bar codes of a accession numbers and membership ID codes. These bar-coded data elements can be read through a portable light-scanning device (e.g., barcode reader). Some automation packages have also started utilizing modern data capturing equipments like RFID (Radio Frequency Identification), smart card etc.

5.8.2 Activities of Circulation Management

The broad groups of activities necessary to manage automated circulation are detailed below:

a) Membership Management

This sub-module is basically meant to create and update membership records in a library. The jobs of this sub-module start with the creation of master database, which contains details about member categories, institutes, departments, courses/designations and so on. A systematic organisation of activities related to this sub-module will give us three subgroups of facilities:

Master database creation and maintenance: This includes the creation of various master files necessary for membership management in particular and circulation management in general. The details covered in this database are:

- **Category and privileges determination:** This will enable to assign category code, duration, membership charges if any, and to determine various privileges associated with the respective membership categories. The privileges include types of borrowing items, their issue period and reservation facility along with overdue charges.
- **Departments/Divisions file creation:** It allows adding a department under institute by entering department code and its name, address etc. Some packages also allow creating course/designation master file under each department. This option is particularly helpful for academic libraries.
- **Calendar:** This is meant for developing library calendar by entering list of holidays as per the calendar of events of the parent institute. It helps to automatically shift the date of return to a working day, if it falls on a closed holiday. This will also take care of calculating overdue charges accordingly.

Member enrolment: This facility is meant to create and update membership records in a library. It includes

- **New membership record creation:** It is to enroll members to the library by entering member's personal information viz., last name, first name and middle name, type of membership (selected from master database), membership status, category of member, name of department/division, institute and course (if applicable) and contact number, address etc. Membership code may be generated automatically or entered manually into the database. Member ID issued by the administration to staff is often used as the Library Member ID used at various places like identity card, reminder and document transactions.

- **Modification of membership record:** This is required for editing or updating a member's record using member's ID or code.
- **Deletion of membership:** This facility will enable the library staff to delete membership either group wise or single at the end of membership period. This step requires the clearance of all dues for each individual member of the group.
- **Renewal of membership:** This provision is required for renewing the date, when the validity of membership expires.

Output generation: This facility allows:

- Searching of membership details against name or member ID
- Generation of 'No-Dues' certificate group-wise or single at membership termination
- Generation of membership ID card, if necessary, and printing before distribution.

b) **Transaction Management**

Transaction sub-module includes all the day-to-day activities of the circulation section of library viz. issue, return, renewal, reservation, reminders for overdue books, searching document availability and listing of items issued to a member. The details are as follows:

- **Issue/Check-in:** To issue a document, accession number of that document is to be entered either manually or by using a data scanner. Then the bibliographical information of the document along with its status (whether available for issue or reserved) is displayed on the screen. The next step is to enter member code of the borrower. Then the screen will show details of the member along with category, number of documents issued against the member and overdue charges etc. The library staff can then issue the document by selecting appropriate option. The system after checking various parameters settled by the library, will generate appropriate message along with date of issue and due date of return. This sub-module generally supports extension of due date, if required.
- **Return/Check-out:** This facility is meant for receiving back the documents lent out to users. The work starts with entering accession number of returned document. It will display the status of the document along with details of returning member. The library staff may then mark returns the document by clicking necessary option. If the book has been returned within due date a successful return message will be displayed. In case of late return (after due date) the system will be asking for fine collection.
- **Renewal:** This is meant to issue a document to the same member again by renewing the transaction record. Due date gets changed automatically in the process. The renewal process will be successful, if someone has not reserved the document.
- **Reservation:** It supports booking of already issued document for a member. Reservation facility is programmed to generate message after returning a reserved document. The reservation process also allows change of reservation priority and cancellation of reservation.
- **Recall:** This facility is useful if the document issued needs to be recalled and a reminder is to be sent to the member. It allows printing of recall letter and storing of the letter for record.
- **List of document borrowed:** This facility helps to display details of documents borrowed by a member and information about overdue charges, etc.
- **Document availability:** Circulation staff has to search sometimes details of the document for checking of bibliographical information or document availability. This

facility provides a link to the OPAC module from the transaction sub-module in an integrated architecture.

c) **Reminder Generation**

This facility is meant for generating reminders for overdue documents

- To a group of members
 - To individual members
 - For a particular due date
 - To all defaulters
- The format and text of reminder letter may be modified by using this facility or by using the master database.

d) **Fiscal Management**

It provides option to manage outstanding dues against a member. This facility allows entering of overdue charges or amount against a member and generation of payment receipt. This facility should also allow printing of fine statement if a member wants to have a statement of fines.

e) **Inter-Library Loan (ILL)**

Inter library loan method simply means that documents of a library can be issued to other libraries. No library in the world can procure all the published documents on all subject areas. So, libraries have to depend on each other for meeting of the information requirement of their users. The success of ILL clearly depends on the availability of union catalogues. The circulation module of LMSs should support ILL to help library staff in smoothly operating ILL method. This sub-module generally includes the following facilities:

- **ILL borrowing:** It involves tasks to procure document(s) for a member of the library who has asked for document on ILL from some other library. The steps are:
 - Receiving request and entering of relevant information into the database (member code, item details, type of item etc.)
 - Generation and printing of request letter addressed to the library from where the document is to be borrowed
 - Receiving of document(s) on ILL by selecting the received status and entering date of arrival in the module
 - Sending intimation about arrival of document(s) on ILL to the requester
 - Issuing document(s) to the member who has requested it
 - Returning the item(s) borrowed on ILL
 - Generating reminder letter in case of overdue document(s)
 - Sending ILL document back to the owner library
- **ILL lending:** It is meant for lending of library documents to member libraries on ILL through the use of ILL sub-module. The tasks to be performed for the service are:
 - Issuing of item(s) to the member library through entering necessary details (accession number of document and name and code of the member library)
 - Receiving back the item(s) on ILL

- Sending reminder to the library concerned in case of delay in returning
- Generating list of documents which have been loaned to other libraries on ILL

f) Maintenance

This sub-module is generally attached with circulation module for recording information about lost documents, documents sent for binding, damaged documents, missing documents and documents withdrawn from library. The activities of this sub-module include following groups of tasks:

- **Lost documents:** Loss of a document is managed through two options—replacement of document and recovery of amount. In the first case replacement is confirmed after entering code of the concerned member and accession number of the lost document. If amount is to be recovered from a member for the lost document, the 'amount recover' option is to be activated by entering member code and price to be recovered. Payment receipt may also be generated by selecting appropriate option.
- **Binding:** The reasons for binding documents are two-fold i.e., their preservation for posterity and to keep them in a usable state for users of the day. Binding facility of maintenance sub-module is a combination of five basic steps:
 - Entering accession numbers of the documents to be sent for binding
 - Order generation for binding of selected documents through inputting of order number, expected date of delivery, price, name of binder, budget head and binding type
 - Receiving of documents from binder and processing of invoice
 - Release of payment
 - Generation of reminder for overdue items
- **Withdrawal:** Stock revision or weeding out of documents is essential in order to make enough space for the usable stock. The catalogue database is also to be modified through recording details of weeded out documents. In computerised circulation, the task of weeding out is performed by entering accession numbers of withdrawn documents in appropriate sub-module. Such a system has the provision to reintroduce an already weeded out document by simply checking/clicking the necessary button.

5.8.3 Outputs of Computerised Circulation Jobs

The circulation module would usually have the facility to generate reports, lists and statistics on various aspects of activities performed by the circulation section of a library. LMSs vary considerably from each other in terms of output generation. However, the typical list of outputs from computerised circulation activities includes the following:

- **Member ID card:** Member ID card with name of the member, membership code, department, institute, category, branch and year may be printed by utilising appropriate facility. LMSs also support inclusion of photograph of members on the card and conversion of membership code into bar-coded form for easy and error free data capturing.
- **Reminder:** Generation of reminder letters for overdue document(s) is a regular work of the circulation section. Reminders are produced and sent to a group of members or an individual member or to all defaulters. Reminder letters for ILL over dues are also generated as output.
- **List of library members:** The list of members can be printed either by name or by member code. Member list can also be generated by department, category or

for a particular time period (membership duration). The list can be sorted on any required sequence or order.

- **Items issued over a period :** This is a list of documents issued either on a particular date or between two dates.
- **Items returned over a period:** It gives a list of documents returned either on a particular date or between two dates.
- **Items reserved over a period:** It produces a list of documents reserved either on a particular date or between two dates.
- **Item's transaction history:** Entering accession number or resource identification number of the desired document may generate the transaction history of any particular document displaying who are the members to have got issued this document since its acquisition.
- **Membership expiry list:** This facility generates a list of memberships expiring either on a particular date or between two dates.
- **Member history:** It generates or displays a list of documents got issued and returned by a member during his/her membership period. The membership code of the user is required to be entered for the generation of such a list.
- **Fiscal report:** It gives details of the fines collected by the library either on a particular date or between two dates i.e., in a particular period.
- **Library usage:** It is a very interesting report that shows a graph displaying department/division/section of the organisation on the Y-axis and number of documents used by them section on X-axis. Similar reports to show usage by deferent category of library members may also be generated.
- **Most frequently issued items:** This facility is to generate a list of most frequently issued documents by displaying accession numbers, titles of the books and number of times they has been issued
- **Most frequent members:** It displays a list of most, frequent users by showing member details and how many times each has used the library.

These are the most useful forms of outputs generated by circulation modules of library automation packages. Apart from the abovementioned outputs they may also generate the following:

- List of items issued to a member:
- 'No dues' certificate
- ILL reports (arrival intimation, reminder, list of items on ILL, overdue charges and payment receipts)
- Transaction details undertaken by a staff member working at circulation desk.
- List of lost, missing or damaged documents
- List of lost documents for which money was recovered
- List of documents sent for binding
- Order letter for binding
- List of withdrawn documents

5.8.4 Advantages of Computerised Circulation Subsystem

Computerised circulation subsystems apart from performing all the basic activities related to circulation, offer extensive control of stock. Transaction records can be entered and saved into the main database through a terminal. The central transaction database is updated immediately and subsequent consultation of the database will communicate the current situation. Some of the important issues may be enumerated as:

- Fines can be calculated on demand
- Reservation and other modifications to document records can be made instantly
- Automatic identification of over borrowing and problem borrowers
- Error-free data capturing through barcode, RFID and smart card technology
- Back-up provision to cover telecommunication or computer system failures.

The circulation modules of modern LMSs provide many advantages over the manual system. Such a system can:

- Identify and trap reserved documents on their return from loan
- Generate (in print and electronically) overdue and recall notices
- Collect and display issue statistics
- Print checkout and check-in slips
- Manage a variety of categories of stock and borrowers, each with its own library defined circulation parameters (limits)
- Keep records of fine payment in details (limits)
- Manage multi-site provision, with each site having its own stock, borrowers and associated circulation parameters
- Change status of an item in the OPAC after it has been issued or returned
- Circulate un-catalogued items and to add such records to the catalogue database, not to the circulation file
- Support multiple items renewals in one step
- Alert operator if an item being charged is already issued to another user
- Allow users to place reservation on an available or charged/issued item without library intervention.

A recent trend of LMSs is to incorporate self-checking or self-issue option in the circulation module. RFID (Radio Frequency Identification) is being used to manage un-manned self-service counters for issue and return of documents. An RFID system comprises three components: a tag, a reader and an antenna. The tag is paper-thin chip, which stores necessary bibliographic data. The tag is to be fixed on the inside cover of the corresponding document. RFID reader and antenna are often integrated into patron self-checkout machines or inventory readers. The reader powers the antenna to generate RF field to decode information stored on the chip. Reader sent information to the central server, which in turn communicates with the library automation software. RFID, apart from self-issue facility, also supports stock verification, theft detection, and identification of misplaced books and inventory counts (see unit 06 of this course).

5) Explain the use of data capturing devices in circulation subsystem.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

5.9 SUMMARY

Library housekeeping operations are routine chores that are to be performed to maintain day-to-day functions of a library. Application of ICT helps in the mechanisation of these labour-intensive works and thereby ensures more productive use of library staff. Automation of library housekeeping operations requires analysis of the library system and subsystems in terms of procedures, activities and jobs. Such analysis would help to identify gaps and overlaps in the existing system. Automation of library housekeeping operations may be divided into four subsystems –computerised acquisition subsystem, computerised cataloguing subsystem, computerised serials control subsystem and computerised circulation subsystem. The responsibility of acquisition subsystem is selection, ordering, receiving and accessioning of library resources. It also includes the works related to budget allocation, fund accounting and generation of outputs in relation to MIS support and user services. In an integrated automation package, acquisition and cataloguing works in harmony. The bibliographical data of newly acquired documents are transferred from acquisition module to the cataloguing module. In the cataloguing module, the bibliographical data elements of documents are standardised through necessary addition and modification. The bibliographical data format is based on any internationally adopted content designator scheme to allow exchange of cataloguing data.

Computerised cataloguing includes three groups of work namely, authority control, data entry and downloading. The catalogue records act as the central bibliographic database in a library system. Regular backup of the catalogue database on suitable media and its easy recovery at the time of need is another important area of computerised cataloguing work. The computerisation of serials control activities helps library staff in the management of frequent and repetitive record addition and modification related to serials. Serials control module of modern Library Management Software (LMS) attempts not only to mechanise ordering, receiving, claiming, binding, and other such functions but also performs predication of arrival of issues, schedule preparation and auto-reminder generation. Serials control work starts with the creation of the master database and records of the master database are made available to all the sub-modules. It follows three groups of activities – subscription and acquisition, cataloguing and article indexing and circulation and binding. Computerised serials control subsystems are able to generate a variety of outputs and these are very useful for the design and development of information products and user services in the required forms and formats. Apart from the issue, return and renewal of documents, a circulation system performs reservation of documents, inter library loans, maintenance of statistics and other supportive activities. Computerised circulation systems can manage all these primary operations along with the additional tasks of member card generation, notification of document status (issued or not), reminder generation, automatic calculation of overdue charges, and maintenance of records related to lost, damaged or missing

documents. The circulation module of LMSs centres on the transaction file as the central database. It draws necessary dataset from the document file (catalogue database) and user file (member database). Modern circulation modules extend their support to a variety of data-capture devices of which RFID, smart card, barcode reader and other light scanning devices are quite useful.

5.10 ANSWERS TO SELF CHECK EXERCISES

- 1) The procedural model of library housekeeping involves three hierarchical levels - Procedures, Activities and Tasks. Each library system may be divided into some operational subsystems depending on the nature and objective of the library. The operational subsystems include eighteen common procedures (as identified by ASLIB). Each procedure consists of six possible activities and activities are carried out by a set of fifteen jobs. This procedural model may be applied to any library, as the model is independent of nature, objective, size, and number of users and staff of the library.
- 2) The primary activities associated with computerized acquisition process are of three types – pre-acquisition works, acquisition works and generation of outputs. In the first group i.e. pre-acquisition group the essential activities include development of various master files such as supplier file, currency file, budget file, letter file and member file. The actual acquisition work includes suggestion processing, pre-order searching, ordering, receiving and accessioning. The processing of payment follows these steps. In the third stage, library professionals generate order letters, various lists and reports. Computerised acquisition works can also generate some services like list of recent arrivals, generation of books cards etc.
- 3) The steps involved in generating a computerised catalogue database are: (A) Authority file creation and maintenance; (or using a suitable thesaurus for the authority file), by (b) data entry for cataloguing, and (c) Copy cataloguing and downloading (i.e., downloading the catalogue entries for the documents to be catalogued, from a suitable OPACs of libraries and making necessary changes and merging the data with a library's catalogue database.
- 4) Information products from computerised serials control subsystems are designed on the basis of users' requirements. Generally, such systems produce, discipline-wise journals and issues arrival list, discipline-wise or alphabetical holdings lists of serials, union catalogue, etc. The article-indexing database may produce CAS, SDI, subject bibliographies and other value added information services. Computerised serials control system can generate number of reports, lists and statistics. These are very helpful to the authority and library staff for planning products and services and making decisions on renewal and new subscriptions. A representative list of such reports includes payment reports, list of issues received, list of issues not received, list of missing issues, binding reports, vendor performance reports, etc.
- 5) Barcode is a system of encoding characters by means of a series of parallel bars. This method uses a number of bars (lines) of varying thickness and spacing between them to indicate the desired information. An optical barcode reader can read such bars and convert them into electrical pulses to be processed by a computer. In a barcode-based circulation system accession number of a document is converted into bar-coded form and pasted on the book itself after taking print out of the same. Similarly, bar-coded form of member ID is placed on the member card. A handheld barcode reader is required to translate the barcodes into useful information. These barcodes then acts as pointers to the respective databases. Such a method saves data entry work and eliminates human error in input stage. RFID based circulation is meant for the self-issue and self-return facility in an automated environment. An RFID system comprises three components: a tag, a reader and an antenna. The tag is paper-thin chip, which stores vital bibliographic data. The tag is generally fixed on

the inside cover of the corresponding document. RFID reader and antenna are often integrated into patron self-checkout machines or inventory readers. The reader powers the antenna to generate Radio Frequency field to decode information stored on the chip. RFID reader sends information to the central server, which in turn communicates with the library automation software. RFID, apart from self-issue facility, also supports stock verification, theft detection, and identification of misplaced books and inventory counts.

5.11 KEYWORDS

Integrated Library System : A library automation system in which data entered in one module by other modules to avoid data redundancy.

Intranet : A network belonging to an organisation, accessible only by the organisation's members, employees, or others with appropriate dthat uses Internet technologies (TCP/IP and others) for connectivity.

KOHA : An open source library management software available for downloading from Internet at no cost. It uses some companion softwares and those softwares are also freeware.

LAN : A network workstations that are linked together within a limited geographical areas. Each node (individual computer) in a LAN has its own CPU with which it executes programs, but it is also able access data and devices anywhere in the LAN. This means that many users can share executive devices, such as printers, as well as data. User can also use LAN to communicate with each other.

Library Automation : Library automation is the mechanisation of housekeeping operations and information handling activities mainly by using computer and communication technologies.

Library Housekeeping : A group of basic routine operations related to acquiring, preparing, circulating and preserving the library documents for their maximum use.

Library Management Software : A software package used for the carrying out the functions and services performed by libraries.

OPAC : Online Public Access Catalogue (OPAC), as the name suggests, is a catalogue of all the library resources which can be a library accessed onling through a computer network.

SOUL : SOUL stands for Software for University Libraries. It is a library automation software package developed by INFLIBNET (Information and Library network), Ahemdabad.

System Analysis : A powerful technique for the analysis of an organization and its functions.

WAN : A computer network that operates across large distancesnationwide or worldwide by utilising telephone line, microwave and satellite links. It is also used to interconnect LANs.

5.12 REFERENCES AND FURTHER READING

- Bilal Meghabghab, D. (1997). *Automating Media Centers and Small Libraries: A Microcomputer-based Approach*. Englewood, CO: Libraries Unlimited.
- Boss, R. W. (1997). *The Library Administrator's Automation Handbook*. Medford, NJ : Information Today.
- Cohn, John M. (2001). *Planning for Integrated Systems and Technologies: A How-to-do-it Manual for Librarians*. London: Neal-Schuman.
- Cooper, Michael D. (1996). *Design of Library Automation Systems: File Structures, Data Structures, and Tools*. New York: Wiley & Sons.
- Cortez, E. M. and Smorch, T. (1993). *Planning Second Generation Automated Library*. Westport: Greenwood Press.
- Dempsey, L. (1994). Distributed Library and Information Systems: The Significance of Z39.50. *Managing Information*. 1(6) 41-42.
- Furrie, Betty. (2003). *Understanding MARC: Bibliographic Machine-Readable Cataloging*. Available online at <http://www.loc.gov/marc> (accessed on 03.01.04).
- Gates Library Foundation. (1998). *Gates Library Foundation Homepage*. Available at: <http://www.glf.org/default.htm>. (accessed on 22.11.03).
- Haseltine, R. (1994). New Perspectives on Library Management System. *Program*. 28(2); 53-61.
- INFLIBNET. (1998). *Training Manual for 17th Training Course on Application of Computer to Library and Information Services*. Ahmedabad: INFLIBNET. P.177.
- Mcqueen J. and Boss, R.W. (1984). Serials Control in Libraries: Automated Options. *Library Technology Reports*. 89-282.
- Mishra, R.K. (1999). Library Software Selection: An Analysis. *IASLIC Bulletin*. 44(3); 125-132.
- Murihead, G. (1994). *The System Librarian*. London: Library Association Publishing.
- Murray, I.R. (1997). Assessing the Effect of New Generation Library Management Systems. *Program*. 31(4); 313-327.
- Rayward, W.B. (2002). A History of Computer Applications in Libraries: Prolegomena. *IEEE Annals of the History of Computing*, April-June 2002, 4-15.
- Reynold, D. (1985). *Library Automation: Issues and Applications*. London: Bowker.
- Rowley, J. (1998). *The Electronic Library*. London: Library Association Publishing.
- Simmons, P. and Hopkinson, A. (ed.) (1992). *CCF/B: The Common Communication Format for Bibliographic Information*. Paris: UNESCO.
- Sharma, C.D. (1988). *Advances in Library and Information Science*. Vol.1. Jaipur: R.B.S.A. Publishers.
- Swan, J. (1996). *Automating Small Libraries*. Wisconsin: High Smith Press LLC.
- Tedd, L.A. (1987). *Introduction to Computer Based Library System*. London: Heyde.
- Thomas, P.A. (1975). *Task Analysis of Library Operations*. London: Aslib.

Thurman-Day, T., Flanders, B. and Zuck, G. ed. (1994). *Automation for School Libraries: How to do it from Those Who Have Done It*. Chicago: American Library Association.

U.S. National Commission on Libraries and Information Science. (1998). *U.S. National Commission on Libraries and Information Science Home Page*. Available at: <http://www.nclis.gov>. (accessed on 22.11.03).

WINNEBAGO. (1997). *The Guide to Library Automation*. Caledonia, Minnesota: Winnebago Software Company.

Yates, R. (1996). Library Automation: The Way Forward? *Program*. 30(3), 239-53.

<http://www.libraryhq.com/> (accessed on 02.12.03).

<http://www.ala.org/> (accessed on 09.12.03).

<http://www.koha.org> (accessed on 12.02.04).

<http://www.biblio-tech.com> (accessed on 22.12.03).

<http://www.inflibnet.ac.in> (accessed on 12.12.03).

<http://www.infolibrarian.com> (accessed on 03/02/04).

<http://www.ukoln.ac.uk> (accessed on 23.11.03).

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 6 SOFTWARE PACKAGES: FEATURES

Structure

- 6.0 Objectives
- 6.1 Introduction
- 6.2 Evolution of Library Automation Software
- 6.3 General Functions of Library Automation Software
- 6.4 Requirements for Library Automation Software
 - 6.4.1 System Requirements
 - 6.4.2 Functional Requirements
- 6.5 Implementation of Library Automation Software
- 6.6 Library Automation Software Packages Available in India
 - 6.6.1 Softwares of Foreign Origin
 - 6.6.2 Softwares Developed over Foreign Packages
 - 6.6.3 Softwares of Indian Origin
- 6.7 Evaluation of Library Automation Software
 - 6.7.1 Criteria for Evaluation
 - 6.7.2 Comparison of Automation Packages
- 6.8 Trends and Future Directions
- 6.9 Summary
- 6.10 Answers to Self Check Exercises
- 6.11 Keywords
- 6.12 References and Further Reading

6.0 OBJECTIVES

After going through this Unit you will be able to:

- understand functions and requirements of library automation packages;
- evaluate a library automation package; and
- know features of library automation packages available in India and trace their path of development.

6.1 INTRODUCTION

Software is defined as a set of related computer programs (stored set of instructions) that governs the operation of computer system and makes the hardware run. The software for a computer system may be classified as system software (e.g. Operating System) and application software. System software is responsible for the overall management of the computer resources whereas application software is designed to perform certain tasks and thereby make computers able to perform different predefined jobs. Library

automation software, as application software, performs day-to-day library activities through human interventions. Library automation packages are developed in view of the two most essential activities of any library – housekeeping and information retrieval.

An automated library is one where a computer system is used to manage the library's key functions such as acquisitions, serials control, cataloguing, circulation and the public access catalogue. There are basically three strategies that can be followed for automation or mechanisation of library operations:

- Adopting a piecemeal approach, converting individual operations one at a time.
- Going directly for a fully automated system.
- Working towards an integrated, computerised system progressively, using a 'planned installation' approach.

Library Management Systems (LMSs) are now established as an essential tool in the support of effective customer service, stock management and management of library services. These are based on knowledge and experience of library professionals over the centuries. The rapid growth in the use of hardware, software, connectivity, and reduced costs has resulted in the development of LMSs. Current LMSs are integrated systems, based on relational database architecture. In such systems, files are interlinked so that deletions, additions and other changes in one file automatically activate appropriate changes in related files. The market for LMSs is now a mature one in India. Almost all special libraries and larger academic libraries in India have either adopted a computer-based system or planning actively to go for library automation.

6.2 EVOLUTION OF LIBRARY AUTOMATION SOFTWARE

Software upgradation is a continuous process. LMSs are no exception. A critical study of development of LMSs over the years suggests the developments that LMSs may be divided into four generations on the basis of sophistication of their capabilities, facilities for integration and interconnectivity. The LMSs developed in all parts of the world from mid 1970s to till date may be fitted into one of these four periods.

- The first generation LMSs were module-based systems with no or very little integration between modules. Circulation module and cataloguing module were the priority for these systems and were developed to run on specific hardware platform and proprietary operating systems.
- The second generation LMSs became portable between various platforms with the introduction of UNIX and DOS based systems. The LMSs of this generation offered links between systems for specific functions and were command driven or menu driven systems.
- The third generation LMSs were fully integrated systems based upon relational database structures. They were based on a range of standards, which was a significant step towards open system interconnection. Colour and GUI features, such as windows, icons, menus and direct manipulation have become standards and norms in this generation. They operate in client-server setup.
- The fourth generations LMSs are based on web architecture and facilitate access to other servers over the Internet. These systems allow accessing multiple information sources from single window user interface and support digital archiving.

Thus, the progress of LMSs through the generations has provided us an effective and straightforward user interface, which supports access to multiple sources and services from one multimedia interface. Moreover, the latest LMSs allow customised report generation and manipulation of data and investigate various scenarios and therefore they have all the potential to be a decision support tool. A comparison of features and functionalities of LMSs in four different generations is given in Table 6.1.

Table 6.1: Comparison of the Features of Four Generations of LMSs

Sl. No.	Features	1 st Generation	2 nd Generation	3 rd Generation	4 th Generation
1	Programming Language	Low level language	COBOL, PASCAL, C	4GL	OOPS
2	Operating System	In house	Proprietary, MSDOS	UNIX, Windows,	UNIX, Windows, Linux
3	DBMS	Non-standard	Hierarchical and Network model	Entity Relation model	Object oriented model
4	Import/Export	None	Limited	Standard	Fully integrated and seamless
5	Communication	Limited	Some interface	Standard	Full connectivity across Internet
6	Portability	Machine dependent and hardware specific	Machine independent but Platform dependent	Multi-vendor	Multi-vendor and Platform independent
7	Reports	Fixed format and limited fields	Fixed format and unlimited fields	Customised report generation	Customised report generation with e mail interface
8	Colour	None	None	Available	Fully available with Multimedia
9	Capacity of record holding	Limited	Improved	Unlimited	Unlimited
10	Module Integration	None	Bridges	Seamless	Seamless
11	Architecture	Stand-alone	Shared	Client-Server	Web-centric/ Distributed
12	Interface	Command driven (CUI)	Menu driven (CUI)	Icon driven (GUI)	Icon driven with Web and Multimedia
13	User Support	Single user	Limited number of users	Unlimited number of users	Unlimited number of users
14	Multi-lingual support/ UNICODE	None	Limited (through Hardware support)	Standard	Full support through UNICODE

1) Compare 3rd and 4th generation automation packages.

- Note: i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

6.3 GENERAL FUNCTIONS OF LIBRARY
AUTOMATION SOFTWARE

Software may be viewed as digital version of human skills. LMSs are based on skills, knowledge and experiences acquired by library professionals on the jobs. These (LMSs) are used as intelligent tools for performing housekeeping operations, information retrieval and MIS (Management Information System) activities. The whole array of functions carried out by a modern LMS is given in Figure 6.1.

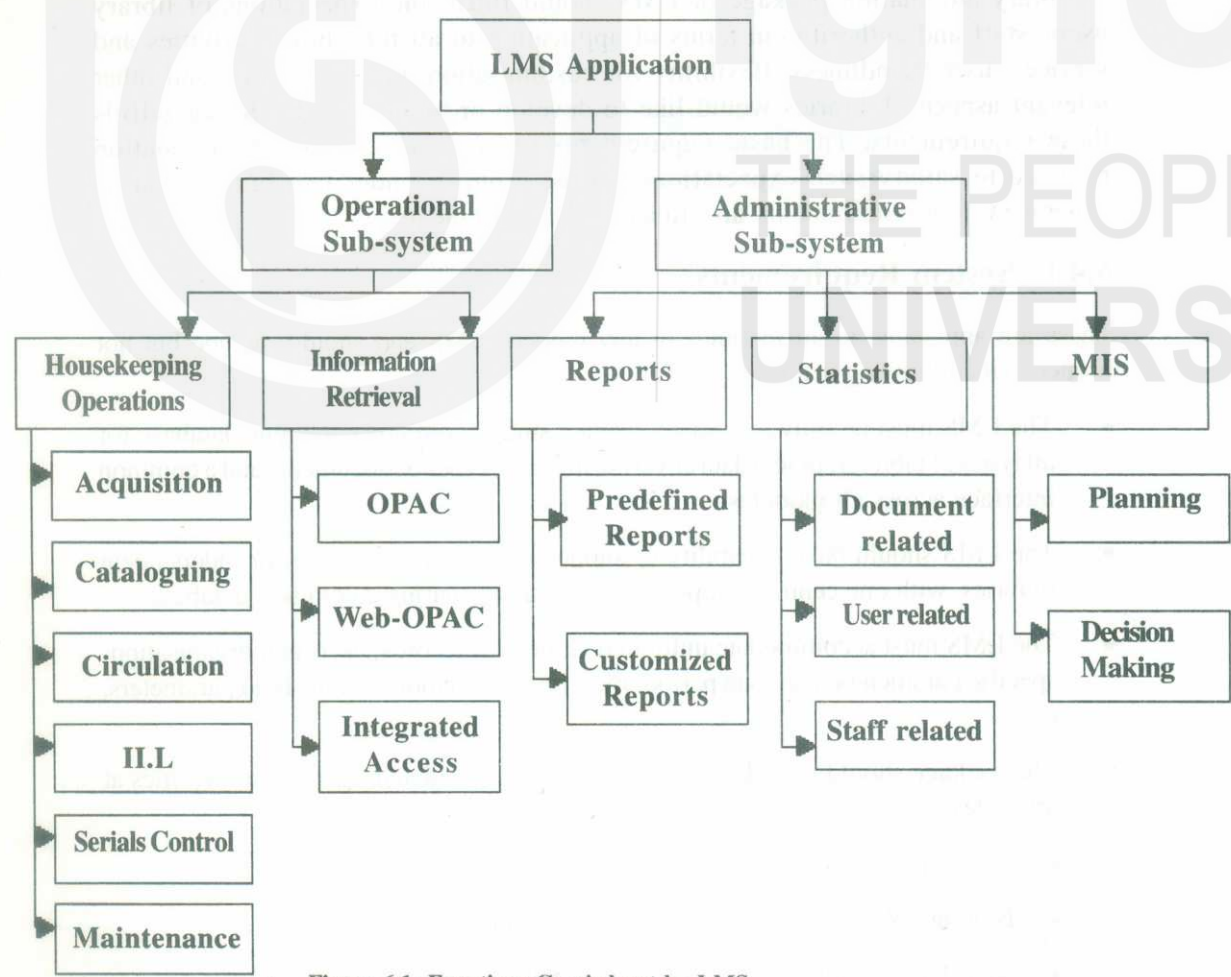


Figure 6.1: Functions Carried out by LMS

LMSs support selection, ordering, acquisition, processing, circulation, serials control, information services and also extend help in library administration, planning and in the decision making process, as a management tool.

The LMSs presently follow a modular approach for the house keeping operations. Generally, the whole package is divided in modules for each operational subsystem. Modules are divided into sub modules and each sub module supports various facilities to carry out tasks related to the library housekeeping operations and information services.

Self Check Exercise

2) Enumerate the role of LMS in library housekeeping.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.4 REQUIREMENTS FOR LIBRARY AUTOMATION SOFTWARE

A library automation package or LMS should fulfill the expectations of library users, staff and authorities in terms of application to all the library activities and services, user-friendliness, flexibility for customisation, of the software and other relevant aspects. Libraries would like to develop or purchase a LMS that fulfills these requirements. The basic requirements for any modern library automation package to satisfy such expectations can be grouped under two broad heads – general system requirements and functional requirements.

6.4.1 System Requirements

These are applicable to all modules of any modern LMS and should include but not limited to the following features:

- The LMS must be fully integrated, using a single, common catalogue database for all types of bibliographical data covering different types of documents and a common interface across all modules
- The LMS should have capability of supporting multiple branches or independent libraries, with one central computer configuration sharing a common database
- The LMS must accommodate unlimited number of records, users and organisation-specific parameters (e.g. loan period rules, fine calculation criteria, hold parameters, etc.)
- The package should include following fully developed and operational facilities at multiple customer sites:
 - Information gateway
 - Bibliographic and inventory control
 - Authority control
 - Public access catalogue
 - Web catalogue interface
 - Circulation control
 - Acquisition management
 - Serials control
 - Electronic Data Interchange (EDI)
 - Reservation and materials booking

- Information gateway (telnet, www, Z 39.50, proxy server, external access, customised web portal)
- One step administrative parameters setting
- Z 39.50 sever (minimum version 3 and bath profile level compliant)
- Z 39.50 OPAC and staff client
- Z 39.50 copy cataloguing client
- MARC 21 bibliographic and authority record import/export utility
- LMS must provide continuous backup in suitable media (as per the choice of libraries) so that all transactions can be recovered to the point of failure
- LMS must be compliant with the following standards:
 - Z 39.2 or ISO 2709 information interchange format
 - MARC 21, UNICODE (UTF-8 OR UTF-16)
 - Z39.71 holdings statements
 - Z39.50 information retrieval service (client and server version 3)
 - EDIFACT (EDI standard)
 - IEEE 802.2 and 802.3 Ethernet
 - HTTP, TCP/IP, Telnet, FTP, SMTP
- The LMS should be based on web-centric architecture and extend support for a range of multi-user and multitasking operating systems and RDBMSs
- The LMS must be compliant with UNICODE standard for multilingual support and RFID for inventory management and self-issue/return facility
- Vendor/Developing group should provide training to enable library staff to become familiar with system functions and operation, should supply full and current system documentation in hard copy and in machine-readable form suitable for online distribution and the LMS should include extensive online help for users and staff
- LMS must support multiple hardware architecture in terms of server, network infrastructure, PC-workstations and peripheral devices
- LMS must be supported with regular maintenance and on-call service, periodical software upgrades, continuous R & D, trouble-shooting of third-party software such as database package and the library automation package, distribution of problem fixes/patches and emergency services for system failures and disaster recoveries
- The package must provide security to prevent accidental or unauthorised modification of records through the establishment of access privileges unique to each user on the system and restriction of specific functions to specific users
- LMS should provide graphical user interface including, but not limited to, extensive online help, user self-service and personalisation features. The software should be supported with PC-based alternative that will allow circulation to continue in the event of system failure, communication failure and downtime required for maintenance.

6.4.2 Functional Requirements

These are the minimum essential features to be supported by each functional unit or module of any modern LMS:

Authority Control

The LMS must have the following capabilities:

- Support for MARC authority format for personal, corporate and topical name headings in a name authority file; title, uniform title and series entries in a title authority file and subject headings in a subject authority file
- Provision for generation of SEE, SEE ALSO references and NT-BT-RT relationships network from authority records and link these references to matching access points in OPAC;
- Must allow any bibliographic field to be authority controlled, should include facilities to search, retrieve, and display print and global editing of authority records by authorised operators; and
- Must include provision for multiple thesauri with the ability to produce a list of all citations with authority file violations.

Bibliographic Control

The master bibliographic record of the LMS should extend support for:

- MARC 21 bibliographic and authority record formats;
- MARC record loader that can accept records input from various sources and from various media like tape diskette or over network;
- Global editing utility that find and replace data within specified fields;
- Data format validation during input of bibliographic information ;
- MARC 21 format for holding and display of holdings on the basis of ANSI Z39.44 serials holdings display format;
- Import and export of bibliographic data through Z39.50 complaint catalogue; and
- Interoperability and crosswalk through incorporation of XML, RDF and metadata schemas (e.g. Dublin Core Metadata).

Online Public Access Catalogue (OPAC)

- OPAC must be fully integrated with other modules and accessible through web-based client
- OPAC should provide browse indexes for author, title, and series and browse index combining all four indexes
- It should allow combined, specific and field level searching for all formats along with phrase searching, nested searching and truncated searching
- It must enable searching by using Boolean operators (OR, XOR, NOT, AND), positional operators (SAME, WITH, NEAR, ADJ) and relational operators ('greater than', 'less than', 'equal to', etc.) within and across all fields
- It should provide facility to see processing status (fully catalogued, in process, lost, withdrawn etc.) and circulation status (in transit, reserve, recalled, on-hold etc.)
- OPAC should support full, brief, standard and customized display of records including relevancy ranking of search results

- OPAC should also support bulletin board, information desk, gateway services (to access external databases) and patron self-service options (e.g. holds, renewals etc.)
- OPAC must track users' preference and interests, organised into a list of favourites. These favourites shall be included in a user's personal online account

Circulation Control

The circulation control activities of any modern LMS must be supported with the following facilities:

- Entry of borrower and item identification by scanner, RFID reader or keyboard entry;
- A template for users registration and automatic display of patron data;
- Automatic blocks for membership expiration, exceeding loan threshold, renewal threshold and fine threshold, placement of hold on any title or item in the database;
- Automatic calculation of loan period in hours, days, weeks or months;
- Generation of notices/reminders to users of overdue items and printing of gatepass for issued items. Automatic calculation of fines and other charges, recording of payment and generation of receipts;
- Generation of variety of pre-defined and user defined reports and usage statistics; and
- Fully operational interlibrary loan module, patron initiated transaction and mechanised inventory control (barcode, RFID, smart card etc.).

Acquisition Control

A fully integrated acquisition module of any modern LMS should support the following activities/functions:

- All traditional acquisition activities like pre-order searching, ordering, claiming, cancellation of order, receipt processing, processing payment, fund accounting, vendor accounting, currency control, statistics and report compilation, etc. Linkage between acquisitions, cataloguing and circulation modules must be seamless;
- Accommodation of a variety of materials, including but not limited to monograph, series, serials, newspapers, cumulative indexes, loose leaf materials, supplements, reports and statutes, musical scores, electronic resources, etc;
- Processing of a variety of materials, including but not limited to print, microfilm, microfiche, film, videotape, audiocassette, CD-ROM, magnetic tape, DVD-ROM, etc.;
- Recording, storing and display of bibliographic information, acquisition type (order, gift, approval etc.), status (ordered, received, etc.), invoice and accounting information, vendor information, requestor information, subject code, etc.;
- Budget-based fund accounting with categorised fund allocation;
- Generation of vendor performance statistics and other reports;

- Accommodation of different order types such as firm order, prepayment, exchange, membership, gift, on approval, blanket order, standing order, subscription, continuation, deposit account, etc; and
- Production of various customised and predefined reports as MIS activities.

Serials Control

- The module should have all the basic capabilities like ordering, check-in, claiming, routing, vouchering, fund accounting, union listing, bindery preparation and report generation.
- It should provide the ability to search for serials records by title, ISSN, publisher, vendor, purchase order, uniform title, editor, conference title, keywords etc.
- Serials control module should have provision for article indexing, online acquisition, e-journal directory service and recording of holding information.
- It should support both predictive and non-predictive mode of check-in and auto-generation of claims for non-receipted issues.
- The module should support routing of journal issues and production of various lists and reports.

Digital Media Archive System (DMA)

The aim of DMA subsystem is to support search, retrieval and viewing of multiple media formats from client machines by using a web browser. It should be able to provide the following facilities:

- To browse and search (full text and metadata based) contents of text and images in ASCII, HTML, SGML, PDF, TIFF, JPEG, GIF, BMP, PCX, DCX etc. formats, audio and video clips and streaming audio and video;
- To link itself with library OPAC through electronic access field (MARC/UNIMARC 856 field);
- To receive and register published documents from an electronic document management system;
- To help a user to import one or more files from the user's system and associate them with a metadata schema within the archive;
- To support metadata harvesting by using Open Archive Initiative (OAI)/ Protocol for Metadata Harvesting (PMH);
- To extend support for various Document Object Identifier (DOI) schemes; and
- To accommodate remote document submission system.

System Administration

The administrator or super user should control the overall administration of LMS through a highly secured module for managing the following activities:

- Access control for individual user, for each module and for each function;
- System security to prevent unauthorised access to databases;
- Standard implementation and setting of system parameters; and
- Keep a log of each transaction, which alters the database.

3) What is digital media archiving?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.5 IMPLEMENTATION OF LIBRARY AUTOMATION SOFTWARE

Library automation is a complex process and should be planned astutely. The entire process of library automation may be divided into the following steps:

- Software selection
- Hardware selection
- Site preparation
- General training
- Customisation
- Defining procedures for
 - Bibliographical data entry
 - Administrative data entry
 - Financial data entry
- Commissioning

It is quite obvious that implementation of these steps requires background study or analysis of the library system. It is a precondition to utilise library automation package for effective results. A library will not be able to take full advantage of automation until and unless its manual functions are streamlined and justified. Therefore, the procedures and tasks (see Unit 05 of this block) followed in different sections should be analysed in terms of the following factors:

- Special features of the library system
- Local variations (their validity and usefulness) and available manpower
- Limitations of the existing system
- Nature and objective of the library
- Total number of collection
- Annual acquisition and procedures followed for acquisition
- Number of periodicals scrutinised
- Number of users and their categories
- Number of daily transactions (issue/return/reservation)
- Availability of multilingual documents
- Need of information services (CAS/SDI, etc.)
- Future plan (in terms of networking and consortia)

Library managers must develop an implementation plan for the library automation project on the basis of above facts, prior to the installation of any equipment and the LMS. The plan shall be incorporated upon written approval of the authority. The plan must include, but is not limited to, the following aspects:

- Identification of all the required tasks
- An indication of which person/group responsible for completion of each task
- Expected start and completion dates for each task
- A site preparation requirements
- Performance testing checklist
- Timeline for all the required tasks
- A software installation requirements checklist
- A training requirements checklist
- Training courses and course outlines,
- Description of post-implementation checklist

Self Check Exercise

- 4) Identify factors to be considered for successful implementation of library automation.

- Note:** i) Write your answer in the space given below.
 ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.6 LIBRARY AUTOMATION SOFTWARE PACKAGES AVAILABLE IN INDIA

The automation of library activities in India started in full swing with the introduction of CDS/ISIS. The CDS/ISIS is a menu-driven generalised information storage and retrieval system designed by a team of experts under UNESCO/PGI programme. It is specifically meant for the structured non-numerical databases. In India, NISSAT with the help of other professional bodies organised a number of training courses on application of CDS/ISIS (DOS & Windows versions) in information organisation activities. As a result, a large pool of trained manpower developed in the country. Partially, some organisations developed their own LMSs e.g. DESIDOC developed DLMS (Defence Library Management System), INSDOC came out with CATMAN (Catalogue Management) and SANJAY was developed by DESIDOC under NISSAT project by augmenting CDS/ISIS (Version 2.3) for library management activities. The LMSs presently available in India may be ranked in 2nd, 3rd and in between 3rd and 4th generation on the basis of their features as listed in the table 6.1. As far as the origin and the application domain are concerned, the LMSs available in India may be placed under three fundamental groups – LMSs of foreign origin, LMSs developed over LMSs or textual database management systems of foreign origin and LMSs of Indian origin. This grouping may again be sharpened by dividing the packages on the basis of size of library systems i.e. large libraries, medium range libraries system and small range libraries system. The grouping of LMSs is given in table 6.2. As it is not possible to discuss every LMS listed in the table, only a few LMSs are selected from each group for discussion on the basis of their popularity and features.

Table 6.2: Grouping of LMSs Available in India

Origin	Application Domain		
	Large System	Medium Range System	Small System
LMSs of foreign origin	Alice for WINDOWS BASISplus & TECHLIB plus VIRTUA ILS	KOHA	Not Available
LMSs developed over LMS of foreign origin	NG-TLMS.NET (over TLMS package)	WINSANJAY (Over CDS/ISIS)	LAMP (Over CDS/ISIS)
LMSs of Indian origin	LIBSUITE LIBSYS MECSYS NEWGENLIB NEXLIB SLIM21 SOUL SUCHIKA TULIPS ULYSIS WILISYS	AUTOLIB DLMS GRANTHALAYA Krvger Library Managca LIBRA LIBRARIAN LISTPLUS NETLIB NIRMALS SLIM ++	ARCHIVES CATMAN E-GRANTHALAYA GOLDENLIBRA LIBMAN LIBRARY- MANAGER LIBRIS LIBSOFT LOAN-SOFT SALIM

6.6.1 Softwares of Foreign Origin

This group includes library automation packages developed by foreign or multinational vendors and distributed in India either through approved agents or value-added resellers. The discussion covers four most popular packages of foreign origin.

ALICE FOR WINDOWS

This LMS developed by Softlink International, Australia, is a global software package and is marketed worldwide through a number of agencies based in America, Australia, Britain, Iceland, India, Malaysia, New Zealand and Singapore. This software is marketed under the name of Embla in Iceland, Alice elsewhere in Europe, OASIS in South East Asia and Australia and Annie in America and other parts of the world. Recently Softlink International decided to call the software *Alice for Windows* all over the world to maintain consistency in nomenclature. The main features of Alice are as follows:

- It has four distinct versions – Public library version, Special library version, Academic library version and School library version
- The package is modular and modules are grouped into one of the three sets:
 - Standard Set: Includes Management; Reports and Utilities; Circulation; OPAC
 - Advanced Set: In addition to Standard Set it includes Acquisition; Periodicals; Journal Indexing; Multimedia; Web Inquiry
 - Special Set: In addition to Standard and Advanced set it includes Reservation; Interlibrary loan; Patron self checking; Rapid retrospective conversion; Multilingual features; Self circulation; Union catalogue
- The LMS is backed by a number of support services which include onsite training programmes, continued R&D, feedback system through user groups, free newsletters, etc.
- Besides traditional library materials, it can be used to manage slides, audio and videocassettes, paper clippings, maps, charts, electronic documents and www sites.

Location of documents in library could also be seen with the help of the library map. It is possible to maintain consistency in recording of items through the use of authority files. Alice has a capacity of holding 99 lakhs records.

- It supports a total of eleven search criteria to search the database from any machine (UNIX/MAC/ Apple, etc.) through Internet or Intranet. It helps to generate customized reports in addition to 800 preformatted reports available with Standard Set.
- It supports barcode technology and has in built communication function. As special features, the LMS provides data protection functions, rapid retro-conversion facility and online tutorial and help system.

BASISPLUS AND TECHLIBPLUS

BASISplus and TECHLIBplus are products of Information Dimensions Inc. (IDI), USA and National Informatics Centre (NIC), New Delhi. NIC is the value-added reseller of the packages in India. BASISplus is a client-server relational database system for text and mixed object documents. It adheres to the fundamental principles of open systems including interoperability, portability and scalability. The database engine provides user authentication, document access control, concurrency control, deadlock protection and recovery. The features of the LMS are as below:

- Relational DBMS
- Client-server architecture
- Active data dictionary
- Enhanced security feature
- Complete backup and restore capabilities
- Power search facility
- Full text retrieval
- Mixed object management
- Thesaurus and controlled vocabulary
- Screen customisation
- Document converters
- Immediate and Deferred updating (online and batch)
- Content based retrieval
- Component-level retrieval and image management
- Networking (LAN & WAN)
- Seamless Internet support
- Open Application Programming Interface (API) to support client access to server databases
- GUI based easy user interface for retrieval, display and data entry
- Intelligent search assistance and thesaurus manager
- Allows user to import and export many word processor file formats

TECHLIBplus is a comprehensive library automation package developed over BASISplus and customised to perform all the operations and activities of a fully electronic library. TECHLIBplus supports OPAC, Catalogue maintenance, Circulation, Serials management, Acquisition, Processing and MARC cataloguing. The LMS provides direct access to information in Current Contents.

KOHA

Koha is an integrated library management system that was originally developed by Katipo Communications Limited of Wellington, New Zealand for the Horowhenua Library Trust (HLT), a regional library system located in Levin near Wellington. In 1999, Katipo proposed developing a new system for HLT using open source tools (Perl, MySQL, and Apache) that would run under Linux and use Telnet to communicate with the branches. The software was in production on 3 January 2000, and released under the GPL for other people to use in July 2000. There has been a high level of interest in Koha internationally, and it is currently being used in New Zealand, Australia, Canada, United States, India, Thailand, United Kingdom, and France. Many of the libraries presently using Koha are small and medium sized, mainly school, college, public and special libraries. Koha has just been implemented at the Nelsonville Public Library in Ohio. The Koha project has attracted

the attention of the developers in a number of different countries, with release Version 1.2.2 being coordinated from Canada and the current stable release, Version 2.0.0 (available both for Linux and Windows), from France. The major features of Koha are:

- General: Free to download, no license fees, fast, web centric, fully customisable, environmental friendly (one can recycle those old PCs), establishing an international community of users and developers giving libraries the freedom to do it themselves or work directly with the system builders, generating an international spirit of co-operation and collaboration, easy staff training, supports both Windows and Linux platform, uses freeware companions like Apache as web server, MySQL as backend RDBMS and PERL as scripting language, supports web OPAC and web interface for staff, branches access main server via ordinary phone lines and modems, can run on PC grade or server grade hardware
- Circulation: Issues (including rentals), renewals, returns and fines; uses barcode scanners or keyboard; can generate a list of over dues for a phone reminder system
- Acquisitions: Multiple book budgets and suppliers, real time budget information
- Catalogue updates fast and slick, support for MARC 21 and UNIMARC
- Searching by keyword, author, title, subject, class number or combinations and customised search interface to suit need of individual libraries
- Memberships - One-stop-shop with all member information on one page
- User-driven reservation facility from OPAC interface (Do-it-yourself reserves in the Library or via the Internet)
- OPAC in the library or via the Internet and stock rotation through branch libraries
- Work in progress: Z39.50 searching, virtual bookshelves, French and German versions, NCIP self checking, port to other operating systems so that it will run natively, new themes, additional book information (e.g., covers), web based reports, Mozilla chromed OPAC integrated with Greenstone Digital Library System, integration with Internet "gateway" system and bill to patron card, printing spine labels
- In most cases, Koha users either undertake the development themselves and contribute the changes back to the project, or they commission a developer to undertake specific enhancements
- The Koha project uses a number of channels to allow members of its community to communicate with each other — there is a general mailing list, as well as separate ones for developers, Windows users, French-speaking Koha users/developers, and German-speaking Koha users/developers. In addition, the developers use Internet Relay Chat (IRC) for real-time scheduled meetings and conversations
- Allows multilingual cataloguing and supports Unicode
- Incorporated Functional Requirements for Bibliographic Records (FRBR) partially

VIRTUA ILS

Virtua ILS (Integrated Library System) is a sophisticated, internationalised library automation solution that addresses the full spectrum of library activities. This LMS is designed and developed by VTLS Inc., Virginia, USA. It uses off-the-Shelf UNIX hardware and the Oracle RDBMS to guarantee continued availability and support. The important features of this world-class software are enumerated below:

- It is a fully parameterised software i.e., libraries can configure the setting to achieve maximum flexibility

- Supports national and international standards for data interchange
- Basic system includes modules for OPAC, circulation, reserves, cataloguing, acquisition, serials control and reporting
- Basic system may be supplemented by companion products like RFID, MARC data processing suite, ILL manager and patron self-check system
- Provides support for excellent security options at different levels of access
- Supports UNICODE and thereby enables the input and display of different languages in their native scripts. In fact, Virtua ILS ensures a true multi-lingual catalogue database
- Helps in designing web-enabled digital media archiving and supports development of digital library database (delivery options include CD-ROM, DLT, DVD and DAT)
- Provides 'security bit' enabled RFID solution to serve both inventory and theft deterrence functions
- Provides comprehensive customisation parameters (over 1000) for global settings and each subsystem (OPAC, cataloguing, circulation, acquisition, serials control, etc.)
- Provides extensive and precise control over user activities and helps creation of rich and customised web interface for various collection components for each patron class
- Ensures management of multiple libraries or branches across a library
- Supports multilingual authority control, and networked multimedia database management and seamless access to multiple databases through Z39.50 client
- Incorporated FRBR model in the design of bibliographic databases

6.6.2 Softwares Developed over Foreign Packages

This group includes automation packages in which some applications are built on the top of either foreign automation packages or general text retrieval packages. The discussion here covers three automation packages of which two are based on CDS/ISIS.

LAMP

Library Automation and Management Package (LAMP) is a freeware and available for downloading from Internet at no cost. The package, developed jointly by Netherlands Library Association, Pakistan-Library Association and UNESCO, is based on CDS/ISIS (version 3.07). This MS-DOS-based package is well suited for small libraries like school and college libraries. The package contains five main modules with sub-modules under each. Facilities are grouped under each sub-module. The features of LAMP include the following:

- Supports creation of authority files for books and serials (supplier authority, name authority, subject authority, etc.)
- Supports use of LCSH subject headings in subject authority file
- The Acquisition module supports all the major tasks related to books and serials acquisition like - creation of budget, purchase order authority file and purchase

order data entry; Generation of purchase order; receiving of books and serials; display of order status; reminders to suppliers; Payment and reordering

**Software Packages:
Features**

- Cataloguing module supports data entry for monographs and serials, catalogue card generation, creation of bibliographies, Binding and write off functions
- Circulation module has the facilities like creation of member database, edit/updating of member database, Issue and return of books and serials, renewal and reservation, checking of availability of documents and checking of outstanding documents, generation of statistics on issued items and library members
- Supports a number of other utilities like a generation of gate pass, shelf-card, accession registers, production of various statistical reports, global editing and replacement, etc.

NG-TLMS.net

NG-TLMS.net is state of the art library automation software based on TLMS (Total Library Management Service). TLMS has developed in Germany by TRANCE group. The route of development for this LMS is TLMS -> NGTLMS -> NGTLMS.net. NG-TLMS.net is designed on the top of TLMS by WebOPAC Applications Pvt. Ltd. It is SQL-backed client-server system, based on Microsoft's .NET platform. It supports CCF, USMARC, Indian UNIMARC (as recommended by Central Secretariat Library, New Delhi) and Z39.50. The NG-TLMS.net supports the following activities and facilities:

- Printing of accession register; AACR II card generation; article scan management; authority files creation; auto cataloguing from web sites; auto export & import; Auto keyword generation; automatic barcode generation; letterhead creation; dropdown matching, etc.
- Barcode based issue & return and serials control; auto status generation for progress of processing of documents; bulletin board facility; Kardex generation for serials control; complete Intranet support; automatic claim generation for overdue and missing journal issues
- RTF, Dial-up networking, e mail and printing of gate pass; ID card generation; arrival list generation; multi-lingual support, web access of OPAC; power search facility; Fine calculation and receipt generation; reservation of books; retrospective data conversion; SDI service; search refining; security enhancement; statistics & graphs; stock verification
- UNIMARC input sheet generation; UNIMARC cataloguing; virtual library creation; Z39.50 client & server; UNICODE support (all languages of the world are supported)
- Web centric architecture i.e, requires only installation on server. Client uses browser to access all information
- No restrictions on the number of records; acquisition module includes accounting software and is optional; basic software covers all areas e.g., OPAC, cataloguing and circulation; all Indian Languages supported
- Includes facilities for - WebOPAC, Internet and Intranet via browser, union cataloguing supported in distributed and replication environments, US MARC 21, UNIMARC, CCF and Z39.50 and can be installed by relatively less experienced computer users.

WINSANJAY

This LMS is originated from SANJAY. SANJAY is based on CDS/ISIS (version 2.3). It was developed by DESIDOC under a NISSAT project to meet the requirements of library management activities. It includes a set of 35 Pascal programs and 25 special menus. The features of WINSANJAY are as follows –

- Windows-based and more user friendly than WINISIS and CDS/ISIS for library house keeping operations. Suitable for medium range libraries
- Effective interlinking of databases (it is a great achievement because WINISIS or CDS/ISIS does not support relational database design model). Interlinks book databases, member databases, vendor databases and budget databases
- Maintenance module restricts the access right to a limited set of users and thereby provides security measures and user module helps library staff to carry out daily routine in circulation, acquisition and online catalogue functions.

6.6.3 Softwares of Indian Origin

This group includes packages designed and developed by Indian vendors and software agents. The features of nine automation packages are discussed here. Packages are selected for discussion on the basis of their customer base and popularity and arranged alphabetically.

AUTOLIB

Autolib is fully integrated multi-user software on Windows environment, designed to automate various activities of university libraries, college libraries, public libraries and special libraries. The software is developed by AutoLib Software Systems, Chennai and the product range includes the following LMS:

- MS-Access with VB version
- MS-Access with VB & W version
- MS-SQL server with VB version
- MS-SQL server with VB & Web version
- MS-SQL server with VB (public library version)

This LMS is module based system and designed and developed by a team of library and information specialists, system analysts, software professionals, network specialists and database designers. The features of AutoLib are given below:

General features: Module based; user friendly; GUI environment; based on client server architecture, uses Visual Basic 6.0 as front end and MS-SQL V.7.0 RDBMS as back end; uses TSQL query Language, module level security; Z39.50 protocol support; export/import of data in ISO 2709 format, cataloguing of digital resources, implementation of AACR, CCF, Dublin Core, TCP/IP and dial-up network support, web based reports, menu based operations, incorporated mandatory fields of CCF, powerful search facility/Query builder, printing various reports in several formats, Simple data entry, User ID and password protection, Online help/documentation, Continuous product upgradation, Customer support & maintenance.

Minimum Hardware & Software Requirements: Server configuration- Pentium II/III 64 MB RAM/4.2 GB HD, 32x CD-Drive/1.44 FDD, network accessories, SVGA monitor, Windows NT Server 4.0 and SQL Server 7.0. Client configuration - Pentium II with 32 MB RAM/4.2 GB HD Windows NT workstation/Windows 9x.

Modules: Cataloguing, OPAC/Advanced OPAC, Transaction (Circulation) Management, Book ordering/Acquisition control, Journal/Serials Control, System Administration, Report Management, Article Indexing, Digital library, Web OPAC.

Functional features: 1) Database Management—Data entry/Updating of database for user, author, publishers, suppliers, member, book, journal issues & back volumes, article report, thesis, standard, non-book materials, budget, subject, department etc.; 2) OPAC—Powerful and versatile search facility, simple search for beginners, query builders for advanced users, Query windows for complicated search, Boolean search, field level search (single field/multiple fields, author / title / keyword / subject, accession No / class No., journal name / article name etc); 3) Circulation – Transaction, Issue, return, renewal of books, Journals, back volumes, recall, reservation, cancellation,

reminders, reports); 4) Serials Control - Subscription of new journals, Renewal of journals, receipts of new issues, Reminders for missing issues, Invoice processing, Payments, browsing issues, Reports generation, 5) Acquisition Control— Duplicate checking, Indent processing for new books, Book ordering, Reminders, Receipts of books, Invoice processing, payment, Budget management; 6) System Administration module- User ID & encrypted password protection, Module level Security, Budget management, Stock verification, Global updating, Fixing due dates, Overdue charges etc., Holiday maintenance, Reports, New additions, Catalogue (main/author/title), Accession register/bibliography, List of books by author/ title/ publisher/year, subject/call number (by any order), Books by unique titles, Frequently issued books, frequently accessed books, Books issued/returned/reserved, Receipt for fine amount/ deposit/ loss of book, etc, List of user/publisher/supplier/ departments, No-due certificates, stock verification report, Budget details, Orders, Journal list, Journal subscription /order report/missing issues; 7) Article Indexing – provides for creating journal article database, Allows to create author index and keyword index, Searching and retrieval of journal articles, Allows to create index and abstracts, Allows to publish CAS bulletin, Allows to generate contents pages; 8) Digital Library module - cataloguing multimedia digital resources such as text, images, audio file, video clippings, etc, Allows to catalogue based on Dublin Core standard, Also provides for to handling various file formats such as .bmp, .jpeg, pdf, .doc, .avi, etc.

E-GRANTHALAYA

This LMS was developed by National Informatics Centre (NIC), Bangalore centre to suit the requirements of small and medium sized libraries. It is an easy-to-use software package and supports all the routine library operations. The general and special features of the LMS are as follows:

- The package includes functional modules for administration, serials control, acquisition, circulation, OPAC, reports and index
- Generates customised reports and statistics of library usage
- OPAC allows simple and advance search option, supports web-enabled searching
- The package has bilingual capabilities and can be customised to suit all Indian languages supported by ISM 2000 developed by C-DAC, provides options to control access through login and password and supports quit-in privileges for users and staff
- Supports both standalone and networked operation mode. Recommended sever configuration is Pentium III processor, 128 MB RAM and 4.3 GB Hard disk
- Requires Windows NT/Windows 2000 for server and Windows 98/XP/2000 for client machines and uses MS SQL sever as backend database
- Requires ISM2000/Leap office 2000 as bilingual tool

GRANTHALAYA

This CUI based (DOS and UNIX) modular LMS was developed in FoxPro by INSDOC (now NISCAIR) for medium range libraries. It includes all the modules required for day-to-day library operations. The package has seven modules – Library administration; Query; Circulation; Acquisition; Serials Control; Technical Processing and Data Administration. The salient features of the LMS are:

- Based on object oriented design
- Supports CCF and ISO 2709 for import & export
- Supports Boolean operators and range searching
- Provides online help through screen messages
- Generates a dictionary for various data elements for easy searching

LIBSUITE

This LMS, developed by SOFT-AID Computer Ltd., Pune, is based on web-centric architecture and designed to work with different media. LIBSUITE is based on three-tier web centric architecture in which server machine uses Windows NT/2000 and Internet Information Sever (IIS) – as web server. The database server relies on Oracle 8i and clients use web interface for accessing the server through Internet or Intranet. LIBSUITE extensively uses latest technologies like Active Server Pages (ASP) and Component Object Modeling (COM). The web-centric architecture ensures that any machine with a web browser can be a client. It also ensures working independent of operating systems in client machines. The significant features are:

- It provides all the standard modules and supports customised report generation and standard protocols like Z 39.50
- The package bundles the fully featured modules – Acquisition, Cataloguing, Circulation, Queries, Serials Control, setup and maintenance
- Cataloguing module, apart from supporting regular activities manages multi-format materials, generates entire status i.e., total number of books, number of books issued and number of books on the shelf.
- Circulation module supports all the required operations including ILL and generation of photograph of the member in circulation, panel
- System administration module supports controls over the access, creation of authority entries and setting of parameters for cataloguing, circulation etc.
- Supports stock verification and global addition and deletion
- Acquisition module supports all media and production of accession register
- Web-centric architecture ensures use of any machine as client as it does not require the installation of client-side software
- Provides easy user interface and ensures seamless navigation through Internet and Intranet and ensures login and password - based access as security measure
- Includes various utilities like calculator and calendar

LIBSYS

LIBSYS is a fully integrated multi-user LMS based on client-server model and supports open system architecture, web-based access and GUI. This indigenous software package is designed and developed by LibSys Corporation, New Delhi. LIBSYS has seven basic modules – Acquisition; Cataloguing; Circulation; Serials; OPAC; Web-OPAC and Article Indexing. The leading features of different LIBSYS products such as LIBSYS 4.0, LS-Premia, LS-Digital, LSmart and LSEase are enumerated below:

- Based on client-server model and TCP/IP for communication and networking
- Provides ANSI Z39.50 compliant web access for making the server accessible through Internet/Intranet
- Supports web-OPAC for accessing bibliographic databases through Internet/Intranet
- Supports standard bibliographic formats like MARC 21, UNIMARC, CCF, etc.
- Includes images and multimedia interfaces with LIBSYS search engine
- Supports barcode technology for membership card production and circulation
- Offers SDI, CAS, fine calculation, e-mail reminders, etc. and utilities

- Provides flexibility in choosing operating platforms (UNIX, Windows NT, Novell NetWare) and backend RDBMS (SQL sever, Oracle)
- Supports Web-OPAC through PERL/CGI access mechanism
- Offers a range of products suitable for different types of libraries e.g. LSEase is an affordable solution for small and medium sized libraries.
- Provides RFID technology (in cooperation with TAGSYS – the largest smart card solution provider) for inventory management and for smart card technology integration for identification of individuals
- LSPremia provides full UNICODE support and management of multi-site libraries
- Offers digital resource management system through LS-Digital suite (It supports resource structure definition, scanning, PDF conversion, multimedia database management and metadata based searching and retrieval)

NEWGENLIB

NewGenLib is the result of collaboration between a charitable trust called Kesavan Institute of Information and Knowledge Management (KIIKM), Hyderabad and Venus Solutions Pvt. Ltd. The features of this LMS are:

- Completely web based and adheres to international standards, supports web services and allows networking of unlimited number of libraries, database and operating system independent and uses open-source, n-tier, and Java-based technologies for scalability, reliability and efficiency
- Includes seven modules: Cataloguing, Circulation, Acquisition, Serials management, OPAC, Network Configuration, Administration and Setup. The package is available as single user (small library) version, multiuser-single library LAN/Intranet version, Multiuser-single library Web version, Multiuser-multi-library consortium version
- NewGenLib adheres to international standards like MARC21, ISO2709 (communication format), and AACR-2R. Cataloguing databases is based on well proven database design to adhere to MARC and also supports Unicode 3.0 and UTF-16 encoding format by which it can support all the possible languages (if available in unicode 3.0 code charts)
- Supports functionalities like Import of MARC data from sources such as OCLC and freely available web-based resources, extensive use of setup parameters in configuring the software to suit specific needs, e.g., in management of fines, Multi-user and multiple security levels, automated email facility integrated into different functions of the software to ensure efficient communication between library and users, vendors etc., module-specific querying in all modules
- Acquisition supports - online requests by users, firm orders, on-approval purchases, standing orders, solicited gifts, unsolicited gifts, exchange-triggered acquisitions, web service interfaces to supply sources such as *amazon.com*, Management information reporting to enable better decisions in acquisitions management
- Cataloguing supports data-entry using MARC tags, fields, sub-fields, etc., or simple label and form-based data-entry; Import of MARC records from sources such as OCLC or from free MARC download sites on the web; access to authority files during data entry and catalogue database searching, catalogue record attachments enabling access to related data, e.g., multimedia, web-based resources, scanned images, and full text digital documents, Provision of a search engine to search full text documents, Plug-ins for specialized thesauri, Automatic validation etc.

- Network functionalities supports sharing of hardware, server and application software between the host and one or more associate libraries. It helps users of branch libraries - to download metadata or the full text of records, where records are available, into their desktops; in acquisition of new publications from the host library, To access their circulation records, To access electronic journals across all the libraries in the network, To improve services to both the end user and the library staff
- OPAC module supports - browser-based access to the library's catalogue database; Extensive search, retrieval, display, print, download and formatting options for patrons (customized, text format (brief), text format (Full), MARC tagged, ISO 2709, MARC-XML, Dublin core) Patrons can request new additions, access the circulation data, make reservations and go to the web via the OPAC. Patrons can trigger interlibrary loans, interact with library staff via instant messages/email
- Circulation, apart from traditional functions, supports setting up of a wide range of circulation options, fines, user privileges, etc.; needed in different library environments, Rapid charging, discharging, renewal and reservation operations; built-in traps for delinquent users, reservations, etc.; on-the-fly circulation, Interlibrary transactions, binding management, MIS support for better management of collection and assistance in stock verification
- Serials Control module includes facilities like Integrated management of serials subscriptions, registration, cataloguing and binding, rapid registration of incoming serials issues using a kardex-like interface, Batch and on-demand claiming for missing issues, support for union catalogues, MIS reporting for better serials management.

NEXLIB

This window based LMS was designed and developed by NexEvolve Logic Solutions Pvt. Ltd. It provides a simple point-and-click navigation interface backed by a powerful database engine capable of maintaining millions of records. Nexlib provides all the basic utilities required for the management of libraries, namely, Acquisition, Cataloguing, Circulation, Serials Control and OPAC. The important features of NexLib are as follows:

- Acquisition module is fully integrated with the cataloguing module. It can manage a variety of library materials and also fund accounting
- Cataloguing module helps to define location of items by floor, shelf number etc.
- The entire circulation task can be carried out from one screen
- Provides facility to create unlimited number of user types and member data can be transferred from any existing user information system
- OPAC can be accessed through any standard web browser
- Generates over 50 pre-defined reports related to library MIS
- Ensures easy-to-use library staff interfaces
- Full-featured serials control with the support for variety of formats and advanced search option for OPAC
- Digital Media Archive (DMA) module for the management of full text articles, newspaper reports, images, etc.
- DMA allows incorporation of standard metadata schemas and metadata-based retrieval.
- NexLib is a platform-specific LMS and depends on Microsoft products. The hardware and third party software requirements are as follows: Server should be Pentium III or

higher with minimum of 256 MB RAM and preinstalled with windows 2000 sever. Client machine should be Pentium II or higher with 128 MB RAM and preinstalled with windows 98/NT/2000/XP.

- Follows client/sever architecture and security of library database access is based on Windows NT model
- Supports MS Access, MSSQL or Oracle 8i as backend RDBMS and uses Visual Basic (VB) as scripting language.

SLIM 21

SLIM (System for Library Information Management), a software suite from Algorhythms Consultants Pvt. Ltd., Pune is a module-based LMS that offers wide range of functionality for library management. SLIM 21 is the latest product of the series that succeeded SLIM, SLIM++, SLIMEX and SLIMLX. SLIM 21 supports multiple operating systems and backend databases. The relevant features of SLIM 21 are: as follows

- SLIM21 is a module-based system. The basic modules are Acquisition, Cataloguing, Circulation, Serials Control, OPAC and Article Indexing.
- Enterprise module of SLIM21 supports usage statistics, current awareness service (CAS Publish), web aware OPAC (WAOPAC), web proposals for new books, interlibrary loan (ILL) and selective dissemination of information (SDI)
- Supports export/import through MARC/CCF/ISO-2709 standards and downloading of bibliographic data from online databases through DB Bridge module and Z39.50
- Generates customized reports on screen/printers/RTF or as text/PDF/HTML files with auto e-mailing facility
- Unicode based LMS that supports multi-script sequencing for Indian scripts
- Generates shelving order for documents as per colon classification, supports smart card/RFID based circulation and touch chip (biometric) interface for user authenticity
- Creates library map for easy location of items
- Provides user-friendly online help and reference manual
- Supports digital library environment and transformation of bibliographic data into XML, XHTML and DCMES
- Supports both standalone and network architecture. Minimum requirement for sever machine is P-4 processor and 256 MB RAM with Windows XP/2000 as OS and SQL sever as backend RDBMS. Client machine may be any P-2 machine with window 9x/2000/ME/XP or NT workstation.

SOUL

The story of SOUL (Software for University Libraries) started with the development of ILMS (Integrated Library Management Software) by INFLIBNET in collaboration with DESIDOC. Two versions of ILMS (DOS and UNIX) were developed for university libraries in India. But with the introduction of GUI-based system and other revolutionary changes in the field of computer software, INFLIBNET decided to develop a state-of-the art, user-friendly, Window-based system that will contain all the features/facilities available with other LMSs in the market. As a result INFLIBNET came out with a LMS called 'SOUL'. The package was first demonstrated in February 1999 during CALIBER-99 at Nagpur. SOUL uses RDBMS on Windows NT operating system as backend to store and retrieve data. The SOUL has six modules – Acquisition; Cataloguing; Circulation;

Serials Control; OPAC and Administration. The modules have further been divided into sub-modules to take care of various functions normally handled by the university libraries. The features of SOUL are:

- Window-based user-friendly system with extensive help messages at affordable cost
- Client-server architecture-based system allowing scalability to users
- Uses RDBMS to organise data
- Supports bibliographic standards like CCF and AACR II and ISO 2709 for export & import facility
- Provides facility to create, view & print records in regional languages
- Supports LAN & WAN environment
- Multi-user software with no limitation for simultaneous access
- User-friendly OPAC with web access facility
- Available in two versions – The university library version and the college library version

Self Check Exercise

5) Compare features of any two LMSs of Indian origin.

- Note:** i) Write your answer in the space given below.
 ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.7 EVALUATION OF LIBRARY AUTOMATION SOFTWARE

The selection of library management software is a complex and time-consuming one. Apart from thorough knowledge of the library system, sub-systems, procedures, activities and tasks, it requires the knowledge of LMSs features and trends in the development of ICT. The selection process should be based on seven basic steps – Evaluation, Comparison, Demonstration of package, Feedback from authority, staff & users, Preliminary selection, Modification and Customisation and Final selection. The process of evaluation should be based on some predefined criteria. After evaluation against checklist, the LMS should be compared with other packages on the basis of services, utilities and features.

6.7.1 Criteria for Evaluation

The following factors should be taken into consideration at the time of evaluation of any library automation package

Vendor validity: The reputation of software development group or the vendor is very important. The following questions should be raised to judge the validity -

- a) Is the vendor also the software developer, or is the vendor a distributor or agent for the software developer?

- b) Is there an international presence or is the company localised?
- c) How long has the software developer been in the library systems industry?
- d) How long has the library software you are interested in been in the market?
- e) Who uses this software? (Look for someone in close proximity and contact him or her with questions. If possible, make an on-site visit to see the product in action.)

Services Availability Checklist: The services and utilities of any LMS should be checked for the availability of following core, enhanced and value-added services –

- Core services: Acquisition, Cataloguing, Circulation, OPAC, Serials control, Bibliographic format support, Data exchange format support, Article indexing, Retro-conversion, Standard reports and System administration
- Enhanced services: Customised report generation, GUI-based user interface, Reservation facility, Interlibrary loan module, Multi-lingual support, Union catalogue, Authority file support and controlled vocabulary, Online help, Online tutorial, Power search facility, Internet support, Intranet support, Web access OPAC, Multimedia interface, Barcode support and Backup utility
- Value-added services: Patron self service through RFID and Smart card (self-circulation, self reservation, etc.), Online user training/orientation, Stock verification facility, Members photo ID card generation, Barcode generation, Fine calculation & receipt generation, Gate pass generation, Bulletin board services & e-mail reports, Electronic SDI, CAS support, Digital media archiving support

Functional Checklist: The following general features are part of software module testing, and each should be tested or conducted during the evaluation process:

- | | |
|--|-----------------------------------|
| • Searching Capabilities (All modules) | • Fines and Fees (Circulation) |
| • Data Entry and Editing (All modules) | • Notice Production (Circulation) |
| • Bibliographic/Item File and Maintenance | • Holds (Circulation) |
| • Bibliographic Interface Software | • Recalls (Circulation) |
| • Authority Control | • Patron File (Circulation) |
| • Inventory (Circulation) | • Reserves (Circulation) |
| • Check-out (Circulation) | • Portable Back-up Units |
| • Renewal (Circulation) | • Report Writer |
| • Circulation/Management Reports (Circulation) | • Acquisitions |
| • Check-in (Circulation) | • Serials |
| • Z39.50 Server | • Electronic Databases |
| • Inter-Library Loan | • Gateways |
| • Web Accessibility | • Network Operations |
| • Integrated Archiving | • Self-Registration |
| | • Statistics Generation |
| | • Export and Import of data |
| | • Fund Accounting |

Data conversion and backup utility: The ability of the package in terms of support for data conversion from other library systems and adherence to the international bibliographic

data standards and protocols should be checked extensively. In this age of shared cataloguing systems and web integration, the LMS should also support metadata schemas and interoperability issues like XML, RDF and OAI/PMH. Backup facility in suitable media is also to be checked in view of data recovery at the time of need.

Training, Documentation and Customer support: The vendor must provide the following:

- Adequate training without fees for supervisor and operators:
 - To manage and operate the system on a day-to-day basis
 - To run file backup operations, software utilities and cataloguing utilities
 - To troubleshoot and solve simple problems and load software enhancements received from the vendor
- Complete documentation (in hard copy and machine-readable form) must be available with the package along with regular documentation updates and release notes available for local printing or downloading via www.
- The package must have support from the software vendor for hardware and software maintenance, data conversion, emergency and on-call support and disaster management.

Hardware and Third Party Software Requirements: The vendor should provide a complete list of hardware requirements (processor type and RAM) for server and client machines, operating system requirements and backend RDBMS (with version details) requirements. Evaluation should be based on total cost for minimum hardware and third party software requirements of the package.

Performance testing: Any LMS should be evaluated by checking some performance testing like transaction throughput capacity and response time, hardware functionality, module functionality, conversion testing, database loading, index building, etc.

6.7.2 Comparison of Automation Packages

A comparative study of LMSs selected on the basis of abovementioned checkpoints may be done by taking into account five aspects – hardware requirements, intrinsic features of packages, available services and facilities, customer support services and price.

- Hardware and backend software requirement for LMSs: Any LMS is application software, which requires some system software and hardware support for proper functioning. Thus, the selection of LMS should be done on the basis of careful analysis of the basic hardware and software requirements for the package.
- Intrinsic features of LMSs: The factors like data storage techniques, programming language(s) used in the development of software, database structure, file organisation etc., must also be taken into consideration for the comparative study because these factors will determine very important issues of maintenance, upgradation and customisation in future.
- Services and facilities available with LMSs: The suitability and superiority of any LMS depends on the available services, facilities and coverage of library/information activities in various modules. The whole range of services available in selected LMSs may be divided into three groups – Core services, Enhanced services and Value added services, as stated earlier
 - **Core Services:** These are the basic services necessary for day-to-day library activities and must be available with LMSs. These services may be tabulated for the comparative study as follows:

Table 6.3: Core Services in a Library

SL. No.	Core Services	Score (1 presence of service, 0 absence of service)
1	Acquisition	
2	Cataloguing	
3	Circulation	
4	OPAC	
5	Serials Control	
6	Bibliographic Format Support	
7	Data Exchange Format Support	
8	Article Indexing	
9	Retroconversion	
10	Standard Reports/Administration	
	Total Score	

- **Enhanced Services:** These are the additional set of services, which will make the work of a librarian easy, smooth and seamless, and at the same time these will help users in efficient information retrieval, cross-domain searching and easy navigation.

Table 6.4: Enhanced Services in a Library

SL. No.	Enhanced Services	Score (1 presence of service, 0 absence of service)
1	Customised Report Generation	
2	GUI-based User Interface and Colour	
3	Reservation Facility	
4	Interlibrary Loan Module	
5	Multilingual Support	
6	Union Catalogue	
7	Authority File Support and Controlled Vocabulary	
8	Online Help	
9	Online Tutorial	
10	Power Search Facility	
11	Internet Support	
12	Intranet Support	
13	Web Access OPAC	
14	Multimedia Interface	
15	Barcode Support / RFID Support	
16	Backup in suitable media	
17	Z 39.50 Client and Server (Target)	
	Total Score	

- **Value Added Services:** These are the essential work and services of library management and generally not included within the scope of a LMS. But with the development in hardware, software and connectivity, LMSs are presently trying to provide software solution for this type of work and services.

Table 6.5: Value-Added Services in a Library

SL. No.	Value Added Services	Score (1 presence of service, 0 absence of service)
1	Patron Self-Service (Self-Circulation, Self Reservation, etc.)	
2	Online User Training/Orientation	
3	Stock Verification Facility (RFID)	
4	Members Photo ID Card Generation	
5	Barcode Generation	
6	Fine Calculation & Receipt Generation	
7	Gate pass Generation	
8	Bulletin Board Services & E Mail Reports	
9	Electronic SDI, CAS & CAL Support	
10	Digital Camera Support	
	Total Score	

- **Customer Support Services:** The support services from the software developers/ agent at the right time and at the right place is a critical factor to be kept in mind at the time of selection of any LMS.

Table 6.6: Customer Support Services in a Library

Sl. No.	Customer Support Services	Score (1 presence of service, 0 absence of service)
1	On Call & On Site Support	
2	Continued R & D and Software Updating	
3	Live Internet Support and Updates	
4	Training	
5	User Group and Newsletter Services	
	Total Score	

- **Price:** Last but not least, we have to take the price of the packages also into consideration for cost-benefit analysis of available services in the short-listed LMSs. This factor plays an important role in our country because financial crunch is a regular feature in Indian libraries and information centres. As a general trend, commercial LMSs are costlier than the packages developed by government organisations. The price of the software rises with the number of value-added services and enhanced services incorporated in the software. In general, the pricing model of commercial LMSs may be divided into three options – Basic version (Price range: Rs. 10,000/- to 1,00,000/-), Standard version (Price range: Rs. 1,00,000/- to 5,00,000/-) and Full version (Price range: Rs. 5,00,000 onwards). The basic version supports limited collection size and does not have networking features. The standard version supports unlimited collection size and limited number of network users. The full version supports unlimited collection size and unlimited users. Moreover, commercial LMS developers also claim additional charges for customisation, onsite training,

Self Check Exercise

- 6) Discuss the process of evaluation for LMSs.

Note: i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.8 TRENDS AND FUTURE DIRECTIONS

The rapid developments of ICT have changed the libraries over the last few decades. The library systems all over the world are going through a process of transformation to address the effects and implications of technological change. In response to the needs of the hour, library automation packages are gradually upgraded to satisfy diversified demands of library authority, staff and users by incorporating various epoch-making features. Some of these features are selected for discussion here on the basis of their importance and utilities in library management.

Unicode: Unicode enables the input and display of different languages of the world in their native scripts. Unicode compliant LMSs are able to dynamically change language at any point without affecting other system users. For example, a librarian could catalogue a record in English, and then change languages to enter record in Hindi, Bengali, Tamil, Marathi, etc. Unicode is a character representation standard like ASCII. ASCII is one byte (8 bits) code and can represent only 2^8 i.e., 256 characters, where as Unicode is two byte code (16 bits) and can represent 2^{16} i.e., around 65,000 characters. As a result Unicode standard can represent all the scripts of the world including some obsolete scripts like Bramhi and Kharosti. Unicode provides two encoding formats – UTF-16 (default) and UTF-8 and the present standard (Unicode 4.0) can represent 50,000 characters. A Fully functional multi-lingual system requires the Unicode support by operating system, programming languages, application software like DBMS and word processors. Unicode support for LMS is essential in a multilingual country like India.

Z39.50 Information Retrieval Protocol: The growth of shared cataloguing and cooperative cataloguing initiatives allow capturing of bibliographic data from remote library servers over the Internet. It reduces unit cost of cataloguing and saves a lot of time for individual libraries. However, the major problem is of variation in software and hardware. Library professionals have to learn the specific features of each system. The more the electronic resources grow, the more will be the confusion on how to access information from diverse databases. ANSI/NISO Z39.50 standard was developed to share the bibliographical information electronically and to overcome the problems of database searching with different search languages. Z39.50 is a session-oriented program-to-program open communication protocol based on client-server computing model. LMS incorporated with Z39.50 copy-cataloguing client (called origin in the standard) submits a search request to any Z39.50 server (called target), which then processes the request

and returns the result in desired standard. LMS will then place the captured record in the catalogue editor for manipulation.

Web-centric Architecture: Web-centric LMSs allow web-based access for staff and users and thereby ensures searching, browsing, data entry and system administration from anywhere anytime against user authentication. In such a system, there is no requirement to install client-side software in client computers. Any machine with standard web browser may be used as client for accessing library database. This architecture applies Common Gateway Interface (CGI) and Hypertext Transfer Protocol (HTTP) to ensure platform-independent access to library services. It also helps to overcome space and time barriers.

Integrated Access Interface: Integrated access interface refers to the ability of LMSs to combine multitude of resources and media type in a single and seamless search mechanism. Such interface should support hypermedia environment to include the following facilities services:

- Library catalogue
- Collection acquired in digital form
- Collection digitized in-house
- E-journals and e-books
- Purchase a datasets on CD ROMs
- Subject gateways
- Other library's OPAC
- Bulletin board, Listserv & discussion forum
- Information desk
- Community information
- Manuscripts and local history collection

FRBR-based Bibliographic Data Model: FRBR stands for Functional Requirements for Bibliographic Records. It is a conceptual model, proposed and designed by IFLA (International Federation for Library Associations and Institutions) for the management of bibliographic databases. The model uses entity-analysis techniques to identify entity, attributes and relationships in the bibliographic universe. It also identifies the relevance of each attribute and relationship to the generic tasks performed by users of bibliographic data. Packages are incorporating FRBR model for the design of central catalogue database (See Unit 6 under Block 2 of MLIS 03).

Interoperability and Crosswalk: Interoperability means the ability of multiple systems (with different hardware and software platforms and data structure interface) to exchange data with minimal loss of content functionality. A crosswalk is a mapping of the elements, semantics and syntax from one metadata schema to those of another. It allows metadata created by one community to be used by another group that employs a different metadata standard. Interoperability and crosswalk ensure exchange of bibliographic data among heterogeneous systems across the globe. LMSs are now supporting various standards and protocols like Z39.50, OAI/PMH, METS (Metadata Encoding and Transmission Standard) and MARC-XML to achieve interoperability.

RFID and Smart Card Based Inventory Control: Radio Frequency Identification (RFID) is the technology that is slated to replace barcodes in library applications. The RFID tags are placed in books and usually covered with the sticker. RFID reader and antenna are often integrated into patron self-checkout machines or inventory readers. The reader powers the antenna to generate RF field to decode information stored on the

chip. Reader sends information to the central server, which in turn communicates with the library automation software. LMSs are incorporating RFID technology for performing self-issue and return, stock verification, theft detection, identification of misplaced books and inventory counts. RFID compliant LMS increases staff productivity and ensures foolproof security. Smart card technology is used in libraries to manage public access resources. It makes the process user-friendly for librarians as well as for patrons. It supports self-checkout, payment of fees and fines and use of public access resources through using one smart card by patrons. The system also provides excellent privacy, security options and personalisation of services for library users. RFID solution works through four parts:

- **RFID Tags** — Flexible, paper-thin smart labels that are applied directly to library documents. RFID tag contains a tiny chip, which is both readable and writable and can store information to identify items in library collection. In library applications, it also stores a security bit and if needed, information to support sorting systems.
- **Antenna** — A conduit between RFID tags and the coupler. RFID antennas emit radio waves that activate RFID tags as they pass through the activation field. After a tag is activated, it can send information to or receive information from the coupler.
- **Coupler** — The link between RFID tags and the PC. The coupler can send information in two directions: It can read information from a tag and send it to the PC (read mode), or it can read information from the PC and send it to an RFID tag (write mode).
- **PC** — The link between the coupler and the library automation system. Library automation vendors have already developed software that runs on PC to provide an interface between the RFID hardware and library automation system.

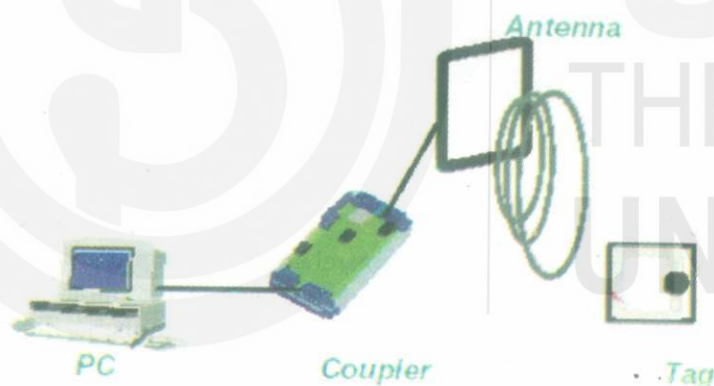


Fig. 6.2: RFID Components

Open Source Solutions: Open Source Software (OSS) is software for which the source code is freely available. It means that anyone can access the source code and make changes. Such facilities are not available with proprietary or closed source programs. Some examples of open source software are — Linux operating system, Mozilla web browser, MySQL RDBMS, Apache web server and PERL. The open source movement has its roots in the 1970s and is continuing to grow in popularity. A number of integrated library automation packages are available as OSS for downloading and use in libraries all over the world, such as KOHA, MyLibrary, Avantika, etc. Generally, these packages are based on LAMP architecture i.e., Linux operating system, Apache web server, MySQL RDBMS and PERL/PHP as scripting language. Many libraries, faced with budgetary crunches and the resultant lack of technological resources, have opted for open source solutions. The advantages of using OSS for library management are:

- Open source systems, when licensed in the typical “general license” manner, cost nothing (or next to nothing) to use - whether they have one or one thousand users. Thus use of OSS offers substantial cost savings for library institutions.

- Open source product support is not locked in to a single vendor. The community of developers for a particular open source product tends to be a powerful support structure because of the pride in ownership. Also, anyone can go into business to provide support for software for which the very source code is freely available. Thus, even if a library buys an open source system from one vendor, it might choose to buy technical support from another company or to arrange for technical support from a third-party at the time of purchase. On top of this flexibility, any library with technical staff capable of understanding source code might find that its own staff might provide better internal support because the staff could have a better understanding of how the systems work
- The entire library community might share the responsibility of solving information systems accessibility issues and OSS can be highly customized to meet individual library needs

Self Check Exercise

7) What are the uses of RFID in automated libraries?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

6.9 SUMMARY

Automation helps to greatly improve the efficiency of the library operations and services through the capabilities extended by ICT in general and LMSs in particular. The process of library automation centres on the automation package or the library management software. Modern packages are integrated in nature and offer all the facilities essential for day-to-day library management. Automation packages available in India are basically of three types – packages of foreign origin, foreign LMS packages on which applications are developed, and packages developed in India. These packages are modular in structure and most of them support all the housekeeping operations, OPAC, Web-OPAC, digital media archiving, Z39.50 based copy cataloguing and modern data capture devices like RFID, smart card, etc. The process of selection of LMS should be based on some well-defined criteria because implementation of LMS is a big investment in terms of money, time and manpower. The concept of computer-based solutions to the range of increasingly difficult problems that libraries were experiencing became potentially attractive to library professionals. ICT seemed to offer librarians the prospect of more efficient processing, improved services to library users, saving money and facilitating resource sharing and inter library cooperation.

6.10 ANSWERS TO SELF CHECK EXERCISES

- 1) The process of development of library automation packages is characterised by four distinct phases or generations. Automation packages of different generations may be compared against some well-defined criteria. The packages belonging to 3rd and 4th generation share many common features such as use of backend RDBMS, multi-

vendor portability, support for import and export of data, graphical user interface and support for unlimited number of documents and users. Apart from all these features, a 4th generation LMS also facilitates multi-lingual support through Unicode, web architecture, multimedia databases and platform-independent access.

- 2) Library housekeeping operations were highly labour intensive and basically routine clerical chores performed slowly and expensively by human beings, until recently. But automation of housekeeping operations is quite helpful for minimizing human clerical routines and thereby making library staff more productive. LMS may be viewed as an intelligent tool in this regard. A library automation package acts as an integrated solution tool to perform related activities in housekeeping such as receiving suggestions, ordering and acquisition, processing, circulation, serials control, ILL, etc. Packages generally follow modular organization of related housekeeping tasks and offer single window control panel for setting up parameters and creation of master databases.
- 3) Digital media archiving supports storing and controlled indexing of locally digitized resources available in different forms and formats. It allows easy and integrated access to full text resources from client machines. The advantages of such a system include – full text and field level searching OPAC integrated access remote submission of digital documents personalized information environment; single interface for local and global resources, and so on.
- 4) Library automation is a complex job and the following factors should be taken into account before implementing automation of a library – objectives of the library, analysis of the existing system, number of users & staff, number of books, bound volumes of periodicals & other documents, number of currently subscribed journals, circulation workload, available financial and human resources and future plan of the library.
- 5) Library automation packages of Indian origin are quite matured now. These packages cover all the essential library services along with various modern aspects of library automation like Z39.50 copy cataloguing, interoperability & cross-walk, RFID support, Unicode support etc. A comparative study of the features of two Indian namely LMSs SOUL and LibSys, in terms of the essential enhanced and value added services other relevant factors and features, reveals that both the packages cover all the core services related to library management. In the area of enhanced services, LibSys supports Z39.50 copy cataloguing, Unicode, RFID, Web-OPAC and web-centric architecture. These facilities are not available with SOUL. The customer support of LibSys is also better than SOUL but SOUL is available at affordable cost whereas LibSys products are priced exorbitantly.
- 6) Selection of a library automation package is a complex task. The process of selection should start with evaluation of packages against some well-defined checklists or criteria. These criteria should include – vendor viability (to judge reputation, customer base of the software and post installation services); availability services (to determine coverage in terms of core services, enhanced services and value added services); utility of the facilities under each module) availability of effective training and documentation (manual) including to check availability of staff training, manuals and online updates as bundled offer); Hardware and third party software requirements (to know the configuration requirements for server and client machines, operating system and RDBMS) and the throughput and operational response time of the software).
- 7) RFID based circulation is meant for the self-issue and self-return facility in an automated library. Apart from self-issue facility, it also supports stock verification, theft detection, and identification of misplaced books and inventory counts.

6.11 KEYWORDS

- Authority Record** : A collection of information about one name, uniform title, or topical term heading. An authority record can contain the established form of heading, see from references, see also from references, and notes.
- FRBR** : A conceptual data model for bibliographic database design and management. The model, proposed by IFLA in 1998, is based on entity-analysis techniques.
- Integrated Library System:** : An automated library system in which data entered in one module by other modules can be to eliminate data redundancy.
- ISO-2709** : An international standard for bibliographic information interchange on magnetic tape, developed in 1981. Most of the content designator schemes are based on this standard.
- Master Database** : A main file of information that acts as the core database. Entries made once in the master file are available for data entry work in related files.
- Module** : Functions specific to a particular system capability such as the online public access catalog, cataloging, acquisition, serial control, circulation, etc.
- Open Source Software** : Open source software is typically created and maintained by developers crossing institutional and national boundaries, collaborating by using internet-based communications and development tools. The products are usually free or sometimes through a license (GPL).
- RFID** : RFID stands for Radio Frequency Identification. RFID system comprises three components: a tag, a reader and an antenna. The tag is paper-thin chip, which stores vital bibliographic data. Reader decodes the information contained in the chip and sends to the server through antenna. This technology is now used extensively by LMSs for patron self-checking function.
- Tag 856** : A field in UNIMARC and MARC 21 bibliographic format to accommodate electronic location and access related to a bibliographical item. The field is optional and repeatable.
- Third party Software** : These are products manufactured or developed by a corporate entity independent from vendor and provided by vendor on a non-exclusive licensing or other distribution agreement with the third party manufacturer.
- Unicode** : Unicode is a two byte oriented code that can represent a total of 2^{16} characters i.e, 65,536 characters. This standard is designed by Unicode Consortium and offers an ideal solution to deal with multilingual processing work.

6.12 REFERENCES AND FURTHER READING

- Chowdhury, G. G. (2001). Digital Libraries and Reference Services: Present and Future. *Journal of Documentation*. 58(3), 258-283.
- Fisher, S. and Rowley, J. (1994). Management Information and Library Management Systems: An Overview. *The Electronic Library*. 12(2), 149-117.
- Haseltine, R. (1994). New Perspectives on Library Management System. *Program* 28(2), 53-61.
- IFLA. (2004). *IFLA Digital Reference Guidelines*. Available at: www.ifla.org (Accessed on 06.05.04).
- Levitan, K. B. (1982). Information Resources as 'Goods' in the Life Cycle of Information Production. *JASIS*. 33(1), 44-54.
- Mishra, R.K. (1999). Library Software Selection: An Analysis. *IASLIC Bulletin*. 44(3), 125-132.
- Mukhopadhyay, P.S. (2001) The Progress of Library Management Software: An Indian Scenario. *Vidyasagar University Journal of Library Science*. 6, 51-69.
- Murray, I.R. (1997). Assessing the Effect of New Generation Library Management Systems. *Program*. 31(4), 313-327.
- Reynold, D. (1985). *Library Automation: Issues and Applications*. London: Bowker.
- Rowley, J. (1998). *The Electronic Library*. London: Library Association Publishing. P. 313-320.
- Tedd, L. A. (1987). *Introduction to Computer Based Library Systems*. London: Heyde.
- U.S. National Commission on Libraries and Information Science Home Page. Available at : <http://www.nclis.gov>. (accessed on 22.02.04).
- Yates, R. (1996). Library Automation: The Way Forward? *Program*. 30(3), 239-53.
- <http://www.libraryhq.com/> (accessed on 02.03.04).
- <http://www.ala.org/> (accessed on 09.12.03).
- <http://www.koha.org> (accessed on 12.02.04).
- <http://www.biblio-tech.com> (accessed on 22.12.03).
- <http://www.oss4lib.org> (accessed on 12.04.04).
- <http://www.ilsr.com/> (accessed on 22.11.03).
- <http://www.infolibrarian.com> (accessed on 02.03.04).
- <http://www.librarytechnology.org> (accessed on 03.03.04).

UNIT 7 DIGITISATION: CONCEPT, NEED, METHODS AND EQUIPMENT

Structure

- 7.0 Objectives
- 7.1 Introduction
- 7.2 Digitisation: Basics
 - 7.2.1 Definition
 - 7.2.2 Need for Digitisation
- 7.3 Selection of Materials for Digitisation
- 7.4 Steps in the Process of Digitisation
 - 7.4.1 Scanning
 - 7.4.2 Indexing
 - 7.4.3 Storing
 - 7.4.4 Retrieving
- 7.5 Digitisation: Input and Output Options
 - 7.5.1 Scanning as Image Only
 - 7.5.2 Optical Character Recognition (OCR) and Retaining Page Layout
 - 7.5.3 Retaining Page Layout using Acrobat Capture
 - 7.5.4 Re-keying
- 7.6 Technology of Digitisation
 - 7.6.1 Bit Depth or Dynamic Range
 - 7.6.2 Resolution
 - 7.6.3 Threshold
 - 7.6.4 Image Enhancement
- 7.7 Compression
 - 7.7.1 Lossless Compression
 - 7.7.2 Lossy Compression
 - 7.7.3 Compression Protocols
- 7.8 File Formats and Media Types
 - 7.8.1 Formats and Encoding used for Text
- 7.9 Tools of Digitisation
 - 7.9.1 Scanners
 - 7.9.2 Scanning Software
- 7.10 Digitisation of Audio and Video
- 7.11 Organising Digital Images
- 7.12 Digital Library Softwares
- 7.13 Planning and Implementation
 - 7.13.1 Feasibility
 - 7.13.2 Planning the Project
 - 7.13.3 Purchase of Hardware and Software
 - 7.13.4 Selection of Material for Digitisation and 'Born Digital'

7.14 Summary

7.15 Answers to Self Check Exercises

7.16 Keywords

7.17 References and Further Reading

7.0 OBJECTIVES

After reading this Unit you will know about the following concepts relate to the technology of digitisation:

- digitisation: basics, concept and need;
- steps in the process of digitisation;
- technology of digitisation;
- compression;
- Optical character recognition (OCR);
- file formats;
- tools of digitisation; and
- organising digital images.

7.1 INTRODUCTION

All recorded information in a traditional library is analogue in nature. The analogue information can include printed books, periodical articles, manuscripts, cards, photographs, vinyl disks, video and audiotapes. However, when analogue information is fed into a computer, it is broken down into 0s and 1s changing its characteristics from analogue to digital. These bits of data can be re-combined for manipulation and compressed for storage. Voluminous encyclopaedias that take-up yards of shelf-space in analogue form can fit into a small space on a computer drive or stored on to a CD ROM disc, which can be searched, retrieved manipulated and sent over the network. One of the most important traits of digital information is that it is not fixed in the way that texts printed on a paper are. Digital texts are neither final nor finite, and are not fixed either in essence or in form except, when it is printed out as a hard copy.

Flexibility is one of the chief assets of digital information. An endless number of identical copies can be created from a digital file, because a digital file does not decay by copying. Moreover, digital information can be made accessible from remote location simultaneously by a large number of users.

Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) into digital format. In most library applications, digitisation normally results in documents that are accessible from the web site of a library and thus, on the Internet. Optical scanners and digital cameras are used to digitise images by translating them into bit maps. It is also possible to digitise sound, video, graphics and animations, etc.

Digitisation is not an end in itself. It is the process that creates a digital image from an analogue image. Selection criteria, particularly those, which reflect user needs, are of paramount importance. Therefore, the principles that are applicable in traditional collection

development are applicable when materials are being selected for digitisation. However, there are several other considerations related to technical, legal, policy, and resources that become important in a digitisation project.

Digitisation is one of the three important methods of building digitised collections. The other two methods include providing access to electronic resources (whether free or licensed) and creating library portals for important Internet resources.

7.2 DIGITISATION: BASICS

7.2.1 Definition

The word “digital” describes any system based on discontinuous data or events. Computers are digital machines because at their most basic level they can distinguish between just two values, 0 and 1, or off and on. All data that a computer processes must be encoded digitally as a series of zeroes and ones.

The opposite of digital is analogue. A typical analogue device is a clock in which the hands move continuously around the face. Such a clock is capable of indicating every possible time of the day. In contrast, a digital clock is capable of representing only a finite number of times (every tenth of a second, for example).

As mentioned before, a printed book is analogue form of information. The contents of a book need to be digitised to convert it into digital form. Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) to digital formats.

Digitisation refers to the process of translating a piece of information such as a book, journal articles, sound recordings, pictures, audio tapes or videos recordings, etc. into bits. Bits are the fundamental units of information in a computer system. Converting information into these binary digits is called digitisation, which can be achieved through a variety of existing technologies. A digital image, in turn, is composed of a set of pixels (picture elements), arranged according to a pre-defined ratio of columns and rows. An image file can be managed as a regular computer file and can be retrieved, printed and modified using appropriate software. Further, textual images can be OCRed so as to make its contents searchable.

An image of the physical object is captured using a scanner or digital camera and converted into digital format that can be stored electronically and accessed via computers. The process of digitisation, however does not stop at scanning of physical objects, a considerable amount of work is involved in optimising usage of digitised documents. Sometimes, these post-scanning processes are often assumed in the meaning of digitisation. At other times the word “digitisation” is used in restricted sense to include only the process of scanning.

7.2.2 Need for Digitisation

Digitising a document in print or other physical media (e.g., sound recordings) makes the document more useful as well as more accessible. It is possible for a user to conduct a full-text search on a document that is digitised and OCRed. It is possible to create hyperlinks to lead a reader to related items within the text itself as well as to external resources. Ultimately, digitisation does not mean replacing the traditional library collections and services; rather, it serves to enhance them.

A document can be converted into digital format depending on the objective of digitisation, end user, availability of finances, etc. While the objectives of digitisation initiatives differ from organisation to organisation, the primary objective is to improve the access. Other objectives include cost savings, preservation, keeping pace with technology and information sharing. The most significant challenges in planning and execution of a digitisation project relate to technical limitations, budgetary constraints, copyright considerations, lack of policy guidelines and lastly, the selection of materials for digitisation.

While new and emerging technologies allow digital information to be presented in innovative ways, the majority of potential users are unlikely to have access to sophisticated hardware and software. Sharing of information among various institutions is often restricted by the use of incompatible software.

One of the main benefits of digitisation is to preserve rare and fragile objects by enhancing their access to multiple users simultaneously. Very often, when an object is rare and precious, access is only allowed for a certain category of people. Going digital could allow more users to enjoy the benefit of access. Although, digitisation offers great advantages for access like, allowing users to find, retrieve, study and manipulate material, it cannot be considered as a good alternative for preservation because of ever changing formats, protocols and software used for creating digital objects.

There are several reasons for libraries to go for digitisation and there are as many ways to create the digitised images, depending on the needs and uses. The prime reason for digitisation is the need of the user for convenient access to high quality information. Other important considerations are:

Quality Preservation: The digital information has potential for qualitative preservation of information. The preservation-quality images can be scanned at high resolution and bit depth for best possible quality. The quality remains the same inspite of multiple usages by several users. However, caution needs to be exercised while choosing digitisation for preservation of information.

Multiple Referencing: Digital information can be used simultaneously by several users at a time.

Wide Area Usage: Digital information can be made accessible to distant users through the computer networks over the Internet.

Archival Storage: Digitisation is used for restoration of rare material. The rare books, images or archival material are kept in digitised format as a common practice.

Security Measure: Valuable documents and records are scanned and kept in digital format for safety and security.

Self Check Exercise

1) Define digitisation. What are the major benefits of digitisation?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.3 SELECTION OF MATERIALS FOR DIGITISATION

To begin the process of digitisation, first of all, we need to select documents for digitisation. The process of selection of material for digitisation involves identification, selection and prioritisation of documents that are to be digitised. If an organisation generates contents, strategies may be adopted to capture data that is "born digital". If documents are available in digital form, it can be easily converted into other formats. If the selected material is

from the external sources, IPR issues need to be resolved. If material being digitised is not available in public-domain, then it is important to obtain permission from the publishers and data suppliers for digitisation. The IPR issues must be addressed early in the selection process. Getting permissions from the publishers and individuals could be time consuming, difficult and may involve negotiation and payment of copyright fees. Moreover, decision may be taken whether to OCR the digitised images. Documents selected for digitisation may already be available in digital format. It is always economical to buy e-media, if available, than their conversion. Moreover, over-sized material, deteriorating collections, bound volumes of journals, manuscripts, etc. would require highly specialized equipment and highly skilled manpower.

The documents to be digitised may include text, line art, photographs, colour images, etc. The selection of documents needs to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and much-in-demand documents and images are selected as first priority without considering the quality. Factors that may be considered for selecting appropriate media for digitisation include the following:

Audio: The sound quality has to be checked and required corrections made together by the subject expert and computer sound editor.

Video: The video clippings are normally edited on Beta max tapes, which can be used for transferring on to digital format. While editing colour tone, resolution is checked and corrected.

Photographs: The selection of photographs is very crucial process. High resolution is required for photographic images and slides. Also, the quality and future needs are to be checked and the copyright aspects are to be taken care of.

Documents: Documents which are much in demand, too fragile to handle, and rare in availability are reviewed and selected for the process. If the correction of literary value demands much input, then documents are considered for publication rather than digitisation. Moreover, the purpose of all digitisation is related to increased access to digitised materials and value addition. The first consideration for digitisation of documents should be intellectual significance of contents in terms of quality, authority, uniqueness, timeliness, and demand. The intellectual contents, physical nature of the source materials, number of current and potential users are therefore, major considerations.

Self Check Exercise

- 2) Describe criteria used for selection of material for digitisation.

Note: i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

.....

.....

7.4 STEPS IN THE PROCESS OF DIGITISATION

Digitisation: Concept,
Need, Methods and
Equipment

The following four steps are involved in the process of digitisation. Software, variably called Document Image Processing (DIP), Electronic Filing System (EFS) and Document Management System (DMS), provides all or most of these functions.

7.4.1 Scanning

Electronic scanners are used for acquisition of an electronic image into a computer from its original that may be a photograph, text, manuscript, etc. An image is "read" or scanned at a predefined resolution and dynamic range. The resulting file, called "bit-map page image" is formatted (image formats described elsewhere) and tagged for storage and subsequent retrieval by the software package used for scanning. Acquisition of image through fax card, electronic camera or other imaging devices is also feasible. However, image scanners are most important and most commonly used component of an imaging system for the transfer of normal paper-based documents.

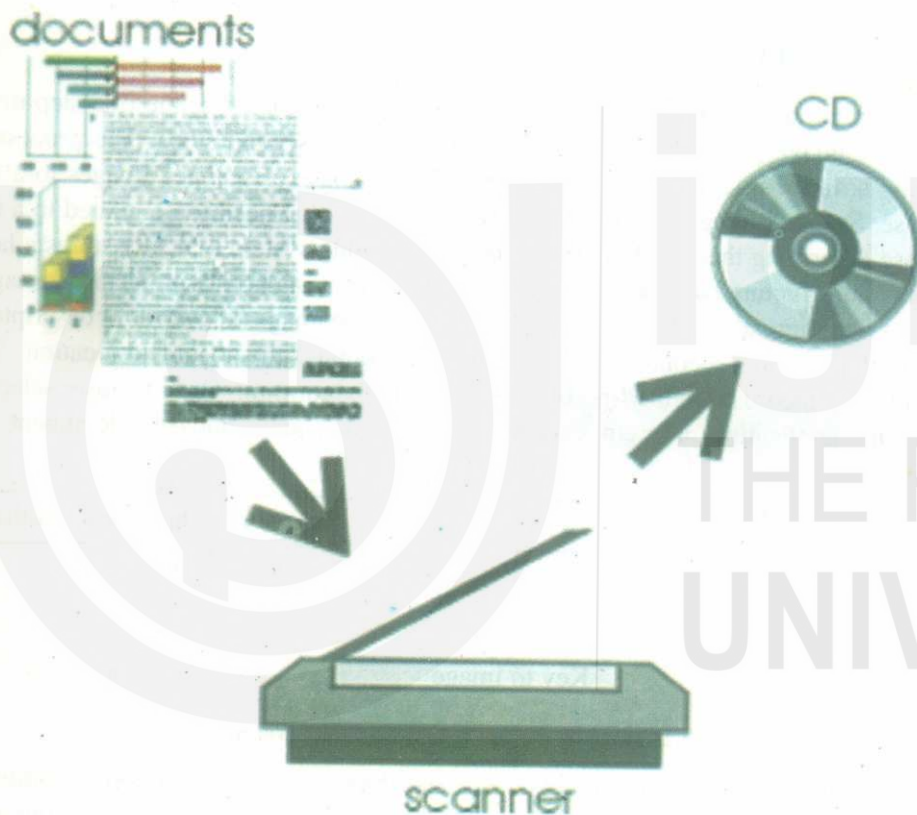


Fig. 7.1: Scanning using a Flatbed Scanner

Steps in the Process of Scanning using a Flatbed Scanner

- Step 1 Place picture on the scanner's glass
- Step 2 Start scanner software
- Step 3 Select the area to be scanned
- Step 4 Choose the image type
- Step 5 Sharpen the image
- Step 6 Set the image size
- Step 7 Save the scanned image using a desirable format (GIF or JPEG)

Self Check Exercise

- 3) Describe the steps involved in the process of scanning a document using a flatbed scanner.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.4.2 Indexing

If converting a document into an image or text file is considered as the first step in the process of imaging, indexing these files comprises the second step. The process of indexing scanned images involves linking of the database of scanned images to a text database. Scanned images are just like a set of pictures that need to be related to a text database describing them and their contents. An imaging system typically stores a large amount of unstructured data in a two file system for storing and retrieving scanned images. The first is traditional file that has a text description of the image (keywords or descriptors) along with a key to a second file. The second file contains the document location. The user selects a record from the first file using a search algorithm. Once the user selects a record, the application program keys into the location index, finds the document and displays it.

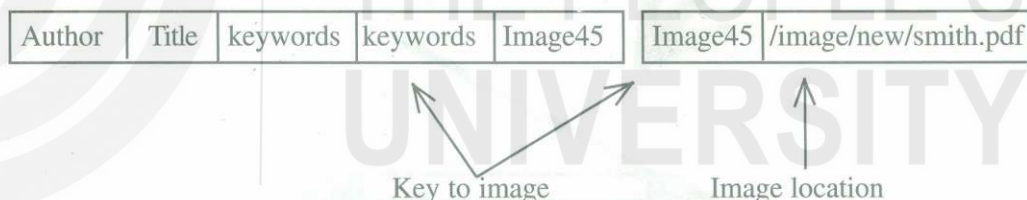


Fig. 7.2: Two File System in a Image Retrieval System

Most of the document imaging software packages through their menu driver or command driven interface, facilitate elaborate indexing of documents. While some document management systems facilitate selection of indexing terms from the image file, others allow only manual keying in of indexing terms. Further, many DMS packages provide OCR capabilities for transforming the images into standard ASCII files. The OCR'd text then serves as a database for full-text search of the stored images.

7.4.3 Storing

The most tenacious problem of a document image relates to its file size and, therefore, to its storage. Every part of an electronic page image is saved regardless of the presence or absence of ink. The file size varies directly with scanning resolution, the size of the area being digitised and the style of graphic file format used to save the image. The scanned images, therefore, need to be transferred from the hard disc of scanning workstation to an external large capacity storage devices such as an optical disc, CD ROM/DVD ROM disc, snap servers, etc. While the smaller document imaging system may use offline media, which need to be reloaded when required, or fixed hard disc drives allocated for image storage, larger document management systems use auto-changers such as optical jukeboxes and tape library systems. The storage required by the scanned images varies

and depends upon factors such as scanning resolution, page size, compression ratio and page content. Further, the image storage device may be either remote or local to the retrieval workstation depending upon the imaging system and document management system used.

7.4.4 Retrieving

Once scanned images and OCR'd text documents have been saved as a file, a database is needed for selective retrieval of data contained in one or more fields within each record in the database. Typically, a document imaging system uses at least two files to store and retrieve documents. The first is traditional file that has a text description of the image along with a key to the second file. The second file contains the document location. The user selects a record from the first-file using a search algorithm. Once the user selects a record, the application program keys into the location index, finds the document and displays it. Most of the document management systems provide elaborate search possibilities including use of Boolean and proximity operators (AND, OR, NOT) and wild cards. Users are also allowed to refine their search strategy. Once the required images have been identified their associated document image can quickly be retrieved from the image storage device for display or for getting printed output.

7.5 DIGITISATION: INPUT AND OUTPUT OPTIONS

A document can be converted into digital format depending on the objective of digitisation, end users, availability of finances, etc. There are four basic approaches that can be adapted to transform from print to digital:

- Scanning as Image Only
- OCRing and Retaining Page Layout
- Retaining Page Layout using Acrobat Capture; and
- Re-keying the Data

7.5.1 Scanning as Image Only

'Image only' is the lowest cost option in which each page is an exact replica of the original source document. Several digital library projects are concerned with providing digital access to materials that already exists in printed media in traditional libraries. Scanned page images are practically the only reasonable solution for institutions such as libraries for converting existing paper collection (legacy documents) without having access to the original data in computer processible formats convertible into HTML / SGML or in any other structured or unstructured text. Scanned page images are natural choice for large-scale conversions for major digital library initiatives. Printed text, pictures and figures are transformed into computer-accessible forms using a digital scanner or a digital camera in a process called document imaging or scanning. The digitally scanned images are stored in a file as bit-mapped page images, irrespective of the fact that a scanned page contains a photograph, a line drawing or text. A bit-mapped page image is a type of computer graphic, literally an electronic picture of the page which can most easily be equated to a facsimile image of the page and as such they can be read by humans, but not by the computers, understandably "text" in a page image is not searchable on a computer using the present-day technology. An image-based implementation requires a large space for data storage and transmission.

Capturing page image format is comparatively easy and inexpensive, therefore, it is a faithful reproduction of its original maintaining page integrity and originality. The scanned textual images, however, are not searchable unless it is OCR'd, which in itself, is highly error prone process specially when it involves scientific texts. Options of technology for converting print to digital are given separately.

If OCR is not carried out, the document is not searchable. Most scanning softwares generate TIFF format by default, which, can be converted into PDF using a number of software tools. Scanning to TIFF / PDF format is recommended only when the requirement of project is to make documents portable and accessible from any computing platform. The image can be browsed through a table of contents file composed in HTML that provides link to scanned image objects.

7.5.2 Optical Character Recognition (OCR) and Retaining Page Layout

The latest versions of both Xerox's TextBridge and Caere's OmniPage incorporate technology that allow the option of maintaining text and graphics in their original layout as well as plain ASCII and word-processing formats. Output can also include HTML with attributes like bold, underline, and italic which are retained.

Retaining Layout after OCR

A scanned document is nothing more than a picture of a printed page. It cannot be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred to by their labels rather than characters in the documents. OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generate a file containing that text in ASCII code or in a specified word processing format leaving the image intact in the process.

7.5.3 Retaining Page Layout using Acrobat Capture

The Acrobat Capture 2.0 provides several options for retaining not only the page layout but also the fonts, and to fit text into the exact space occupied in the original, so that the scanned and OCR'd copy never over- or under-shoots the page. Accordingly, it treats unrecognisable text as images that are pasted in its place. Such images are perfectly readable by anyone by looking at the PDF file, but will be absent from the editable and searchable text file. In contrast, ordinary OCR programs treat unrecognised text as wild or some other special character in the ASCII output. Acrobat Capture can be used to scan pages as images, image +text and as normal PDF, all the three options retain page layout.

- i) Image Only: Image only option has already been described in Section 7.5.1
- ii) Image + Text: In image+text solutions, OCR'd text is generated for each image where each page is an exact replica of the original and left untouched, however, the OCR'd text sits behind the image and is used for searching. The OCR'd text is generally not corrected for errors since; it is used only for searching. The cost involved is much less than PDF Normal. However, the entire page is a bitmap and neither fonts nor line drawings are vectorised, so the file size of Image + Text PDFs is considerably larger than the corresponding PDF Normal files and pages will not display as quickly or cleanly on screen.
- iii) PDF Normal: PDF normal gives the clear view on-screen display. It is searchable, with significantly smaller file size than Image+Text. The result is not, however, an exact replica of the scanned page. While all graphics and formatting are preserved, substitute fonts may be used where direct matches are not possible. It is a good choice when files need to be posted on to the web or otherwise delivered online. If during the Capture and OCR process, a word cannot be recognised to the specified confidence level, Capture, by default, substitutes a small portion of the original bitmap image. Capture "best guess" of the suspect word lies behind the bitmap so that searching and indexing are still possible. However, one cannot guarantee that these

bitmapped words are correctly guessed. In addition, the bitmap is somewhat obtrusive and detracting from the 'look' of the page. Further, Capture provides option to correct suspected errors left as bit-mapped image or leave them untouched.

7.5.4 Re-keying

A classic solution of this kind would comprise of keying-in the data and its verification. This involves a complete keying of the text, followed by a full re-keying by a different operator, the two keying-in operations might take place simultaneously. The two keyed files are compared and any errors or inconsistencies are corrected. This would guarantee at least 99.9% accuracy, but to reach 99.955% accuracy level, it would normally require full proof-reading of the keyed files, plus table lookups and dictionary spell checks.

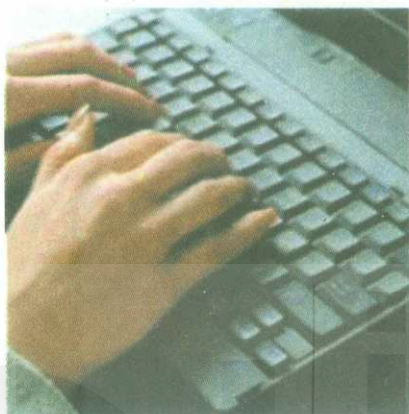


Fig. 7.3: Re-keying-in as an Option for Digitisation

7.6 TECHNOLOGY OF DIGITISATION

Digital images, also called "bit-mapped page image" are "electronic photographs" composed or set of bits or pixels (picture elements) represented by "0" and "1". A bit-mapped page image is a true representation of its original in terms of typefaces, illustrations, layout and presentation of scanned documents. As such, information or contents of "bit-mapped page image" cannot be searched or manipulated unlike text file documents (or ASCII). However, an ASCII file can be generated from a bit-mapped page image using an optical character recognition (OCR) software such as Xerox's TextBridge and Caere's OmniPage. The quality of digital image can be monitored at the time of capture by the following factors:

- Bit depth / Dynamic Range
- Resolution
- Threshold
- Image Enhancement

Terminology associated with technological aspects of digitisation described below is given in the keywords. Students are advised to understand the terminology, specially bit, byte and pixel before going through the Unit.

7.6.1 Bit Depth or Dynamic Range

The number of bits used to define each pixel determines the bit depth. The greater the bit depth, the greater the number of gray scale or colour tones that can be represented. Dynamic range is the term used to express the full range of total variations, as measured by a densitometer between the lightest and the darkest of a document. Digital images can be captured at varied density or bits per pixel depending upon (i) the nature of source material or document to be scanned; (ii) target audience or users; and (iii) capabilities of the display and print subsystem that are to be used. Bitonal or black & white or binary

scanning is generally employed in libraries to scan pages containing text or the drawings. Bitonal or binary scanning represents one bit per pixel (either "0" (black) or "1" (white)). Gray scale scanning is used for reliable reproduction of intermediate or continuous tones found in black & white photographs to represent shades of grey. Multiple numbers of bits ranging from 2-8 are assigned to each pixel to represent shades of grey in this process. Although each bit is either black or white, as in the case of bitonal images, but bits are combined to produce a level of grey in the pixel that is, black, white or somewhere in between.

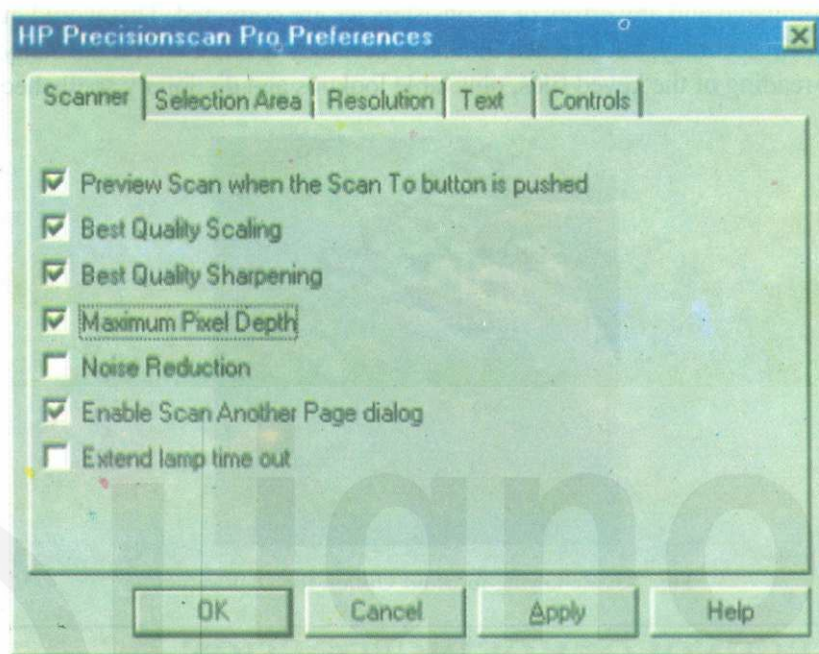


Fig. 7.4: Setting Bit Depth in Precisionscan Pro Scanning Software

Lastly, colour scanning can be employed to scan colour photographs. As in the case of grey-scale scanning, multiple bits per pixels typically 2 (lowest quality) to 8 (highest quality) per primary colour are used for representing colour. Colour images are evidently more complex than grey scale images, because it involves encoding of shades of each of the three primary colours, i.e., red, green and blue (RGB). If a coloured image is captured at 2 bits per primary colour, each primary colour can have 2^2 or 4 shades and each pixel can have 4^3 shades for each of the three primary colours. Evidently, increase in bit depth increases the quality of image captured and the space required to store the resultant image. Generally speaking, 12 bits per pixel (4 bits per primary colour) is considered minimum pixel depth for good quality colour image. Most of today's colour scanners can scan at 24-bit colour (8 bit per primary colour).

Table 7.1: No. of Bits Used for Representing Shades in Colour and Gray-scale Scanning

Sl. No.	No. of Bits	No. of bits/shades	No. of shades	No. of Shades/pixel
1	2	2	$2^2 = 4$	$4^3 = 64$
2	4	3	$2^3 = 8$	$8^3 = 512$
3	8	4	$2^4 = 16$	$16^3 = 4096$
4	16	5	$2^5 = 32$	$32^3 = 32768$
4	32	6	$2^6 = 64$	$64^3 = 262144$
5	64	7	$2^7 = 128$	$128^3 = 2097152$
6	128	8	$2^8 = 256$	$256^3 = 16777216$

4) Why is "Bit depth" not important for bitonal scanning?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

7.6.2 Resolution

The resolution of an image is defined in terms of number of pixels (picture elements) in a given area. It is measured in terms of dots per inch (dpi) in case of an image file and as ratio of number of pixels on horizontal line x number of pixel in vertical lines in case of display resolution on a monitor. The higher the dpi set on the scanner, the better the resolution and quality of image and larger the image file.

Regardless of the resolution, the quality of an image can be improved by capturing an image in grayscale. The additional gray-scale data can be processed electronically to sharpen edges, file-in characters, remove extraneous dirt, remove unwanted page strains or discoloration, so as to create a much higher quality image than possible with binary scanning alone. A major drawback in gray scale is that there is large amount of data capture. It may be noted that continuing increase in resolution will not result in any appreciable gain in image quality after some time, except for increase in file size. It is thus important to determine the point where sufficient resolution has been used to capture all the significant details present in the source document.

The black and white or bitonal images (textual) are scanned most commonly at 300 dpi that preserve 99.9% of the information content of a page and can be considered as adequate access resolution. Some preservation projects scan at 600 dpi for better quality. A standard SVGA/VGA monitor has a resolution of 640×480 lines while the ultra-high monitors have a resolution of about 2048×1664 (about 150 dpi).

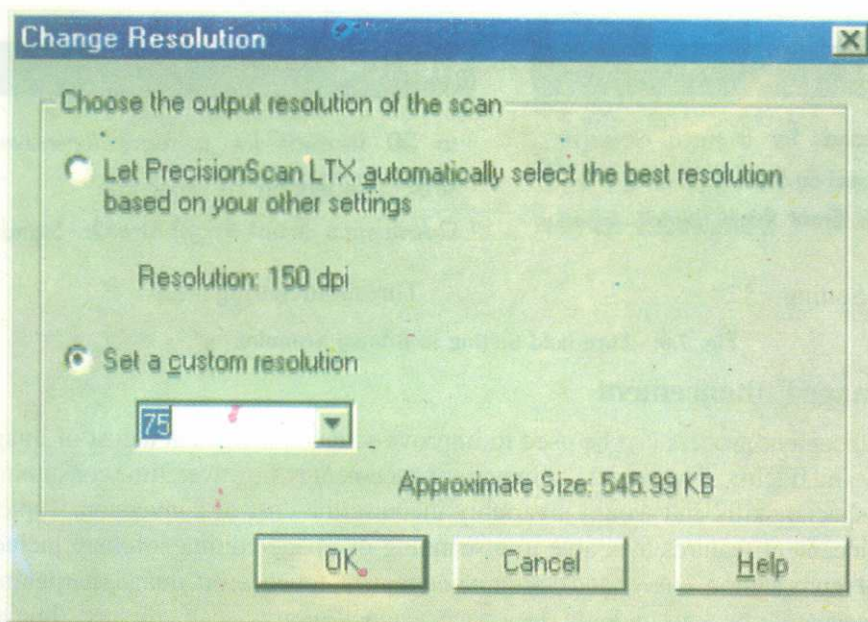


Fig. 7.5: Setting-up Resolution Manually

Self Check Exercise

5) What is resolution? Is print resolution different from scanning resolution?

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

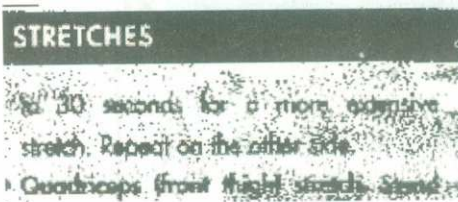
.....

7.6.3 Threshold

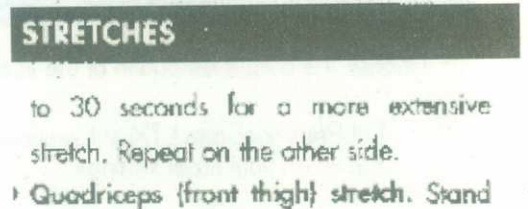
The threshold setting in bitonal scanning defines the point on a scale, usually ranging from 0 - 255, at which gray values will be interpreted as black or white pixels. In bitonal scanning, resolution and threshold are the key determinants of image quality. Bitonal scanning is best suited to high-contrast documents, such as text and line drawings. Gray scale or colour scanning is required for continuous tone or low contrast for documents such as photographs. In gray scale/colour scanning both resolution and bit depth combine to play significant role in image quality.

In line art mode, every pixel has only two possible values. Every pixel will either be black or white. The Line art **Threshold** control determines the decision point about brightness determining if the sampled value will be a black dot or a white dot. The normal threshold default is 128 (the midrange of the 8-bit 0 - 255 range). Image intensity values above the threshold are white pixels, and values below the threshold are black pixels. Adjusting threshold is like a brightness setting to determine what is black and what is white.

Threshold for text printed on a coloured background or cheap-quality paper like newsprint has to be kept at lower range. Reducing threshold from 128 to about 85 would greatly improve the quality of scan. Such adjustments would also improve the performance of OCR software.



Threshold Setting = 128



Threshold Setting = 85

Fig. 7.6: Threshold Setting in Bitonal Scanning

7.6.4 Image Enhancement

Image enhancement process can be used to improve scanned images at a cost of image authenticity and fidelity. The process of image enhancement is, however, time consuming; it requires special skills and would invariably increase the cost of conversion. Typical image enhancement features available in a scanning or image editing software include filters, tonal reproduction, curves and colour management, touch, crop, image sharpening, contrast, transparent background, etc. In a page scanned in gray-scale, the text /line art and half tone areas can be decomposed and each area of the page can be filtered separately

to maximise its quality. The text area on page can be treated with edge sharpening filters, so as to clearly define the character edges, a second filter could be used to remove the high-frequency noise and finally another filter could fill-in broken characters. Gray-scale area of the page could be processed with different filters to maximise the quality of the halftone.



Fig. 7.7: Sharpening Image using HP Precisionscan Pro

7.7 COMPRESSION

Image files are evidently larger than textual ASCII files. It is thus necessary to compress image files so as to achieve economic storage, processing and transmission over a network. A black & white image of a page of text scanned at 300 dpi is about 1 mb in size whereas a text file containing the same information is about 2-3 kb. Image compression is the process of reducing size of an image by abbreviating the repetitive information such as one or more rows of white bits to a single code. The compression algorithms may be grouped into the following two categories:

7.7.1 Lossless Compression

The conversion process converts repeated information as a mathematical algorithm that can be decompressed without losing any details into the original image with absolute fidelity. No information is "lost" or "sacrificed" in the process of compression. Lossless compression is primarily used in bitonal images.

7.7.2 Lossy Compression

Lossy compression process discards or minimises details that are least significant or which may not make appreciable effect on the quality of image. This kind of compression is called "lossy" because when the image that is compressed using "lossy" compression techniques is decompressed, it will not be an exact replica of the original image. Lossy compression is used with gray-scale/colour scanning.

Compression is a necessity in digital imaging but more important is the ability to output or produce the uncompressed true replica of images. This is especially important when images are transferred from one platform to another or are handled by software packages under different operating systems.

Uncompressed images often work better than compressed images for different reasons. It is thus suggested that scanned images should be either stored as uncompressed images or at the most as lossless compressed images. Further, it is optimal to use one of the standard and widely supported compression protocols than a proprietary one, even if it offers efficient compression and better quality. Attributes of original documents may also be considered while selecting compression techniques. For example ITU G-4 is designed to compress text whereas JPEG, GIF and ImagePAC are designed to compress pictures. It is important to ensure migration of images from one platform to another and from one hardware medium to another. It may be noted that highly compressed files are more prone to corruption than uncompressed files.

Self Check Exercise

- 6) What is image compression? Describe the types of image compression.

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.7.3 Compression Protocols

The following protocols are commonly used for bitonal, gray scale or colour compression:

TIFF-G4

International Telecommunication Union (ITU Group 4) is considered as de facto standard compression scheme for black and white or bitonal images. An image created as a TIFF and compressed using ITU-G4 compression technique is called a Group-4 TIFF or TIFF-G4 and is considered as defacto standard for storing bitonal images. TIFF- G-4 is a lossless compression scheme. Joint Bi-level Image Group (JBIG) (ISO-11544) is another standard compression technique for bitonal images.

JPEG (Joint Photographic Expert Group)

JPEG (Joint Photographic Expert Group) is an ISO-10918-I compression protocol that works by finding areas of the image that have same tone, shade, colour or other characteristics and represents this area by a code. Compression is achieved at loss of data. Preliminary testing indicates that a compression of about 10 or 15 to one can be achieved without visible degradation of image quality.

LZW (Lempel-Ziv Welch)

LZW compression technique uses a table-based lookup algorithm invented by Abraham Lempel, Jacob Ziv, and Terry Welch. Two commonly-used file formats in which LZW compression is used are the Graphics Interchange Format (GIF) and Tag Image File Format (TIFF). LZW compression is also suitable for compressing text files. A particular LZW compression algorithm takes each input sequence of binary digit of a given length (for example, 12 bits) and creates an entry in a table (sometimes called a "dictionary" or "codebook") for that particular bit pattern, consisting of the pattern itself and a shorter code. As input is read, any pattern that has been read before the results in the substitution of the shorter code effectively compresses the total amount of input to something smaller. The decoding program that uncompresses the file is able to build the table itself by using the algorithm as it processes the encoded input.

Self Check Exercise

- 7) How does LZW compression protocol work? Which file formats use LZW compression protocol?

- Note :** i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

OCR (Optical Character Recognition)

OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generate a file containing that text in ASCII code or in a specified word processing format leaving the image intact in the process. The OCR is performed in order to make every word in a scanned document readable and fully searchable without having to key-in everything in the computer manually. Once a bit-mapped page image has gone through the process of OCR, a document can be manipulated and managed by its contents, i.e., using the words available in the text.

OCR does not actually convert an image into text but rather creates a separate file containing the text while leaving the image intact. There are four types of OCR technology that are prevailing in the market. These technologies are: matrix matching, feature extraction, structural analysis and neural network.

- i) **Matrix / Template Matching:** Compares each character with a template of the same character. Such a system is usually limited to a specific number of fonts, or must be “taught” to recognise a particular font.
- ii) **Feature Extraction:** Can recognise a character from its structure and shape (angles, points, breaks, etc.) based on a set of rules. The process claims to recognise all fonts.
- iii) **Structural analysis:** Determines characters on the basis of density gradations or character darkness.
- iv) **Neural Networking:** Neural networking is a form of artificial intelligence that attempts to mimic processes of the human mind. Combined with traditional OCR techniques plus pattern recognition, a neural network-based system can perform text recognition and “learn” from its success and failure. Referred to as “Intelligent Character Recognition”, a neural network-based system is being used to recognize hand-written text as well as other traditionally difficult source material. Neural network ICR can contemplate characters in the context of an entire word. Newer ICR combines neural networking with fuzzy logic.

The image scanner optically captures text images to be recognised. Text images are processed with OCR software and hardware. The process involves three operations: document analysis (extracting individual character images), recognising these images based on (i) their template stored in the OCR database; (ii) structure and shape (angles, points, breaks, etc.) (iii) density gradations or character darkness and (iv) contextual processing. The output interface is responsible for communication of OCR system that results, to the outside world.

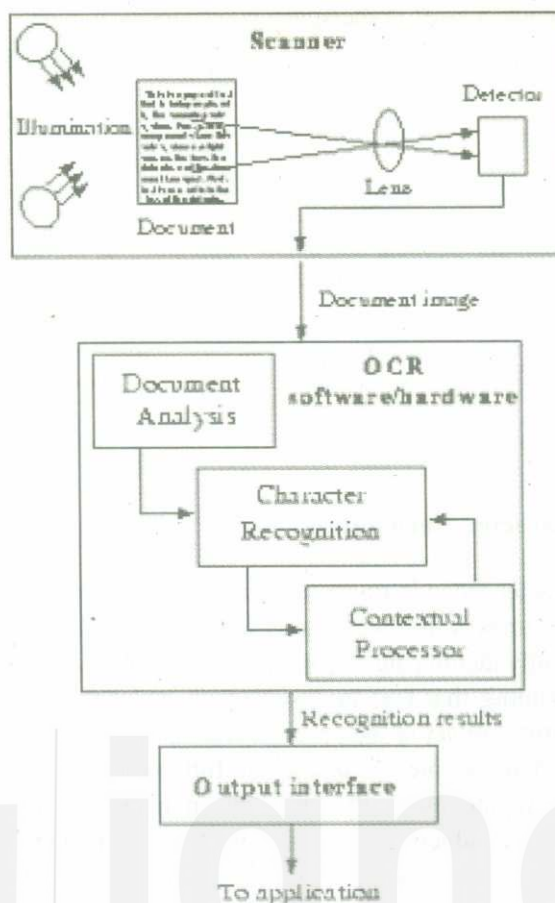


Fig. 7.8: OCR Technology

Several software packages now offer facility of retaining the page layout after it has been OCR'd. The process for retaining the page layout is software dependent. Caere's Omnipro offers two ways of retaining page layout following OCR. It calls them True Page Classic and True Page Easy. True Page Classic places each paragraph within a separate frame of a word processor into which the OCR output is saved. If one wishes to edit anything subsequently, then the relevant paragraph box may need to be resized. However, Easy Edit facilitates editing of pages without the necessity of resizing the boxes although there are greater chances of spillage over the page. Xerox Text Bridge offers similar feature called DocuRT which is broadly equivalent to True Page Easy edit. The process of OCR dismantles the page, OCRs it, and then reassembles it in such a way that all the component parts such as tabs, columns, table, graphics can be used in a text manipulation package such as word processor.

There is a little doubt about the fact that OCR is less accurate than rekeying-in the data. At an accuracy ratio of 98%, a page having 1800 characters will have 36 errors per page on an average. It is therefore, imperative to cleanup after OCR unless original scanned image will be viewed as a page and OCR is being used purely to create a searchable index on the words that will be searched via a fuzzy retrieval engine like Excalibur, which is highly tolerant to OCR errors.

Another possibility for cleaned-up OCR is use of a specialist OCR system such as, Prime Recognition. With production OCR in mind, Prime OCR licenses leading to recognising engine and passes the data through several of them using voting technology along with artificial intelligence algorithms. Although it takes longer initially, but saves time in the long run and Prime contends that it improves the result achieved by a single engine by 65 – 80 %. The technology is available at a price depending upon the number of search engines that one would like to incorporate. Michigan Digital Library production services used Prime OCR for placing more than two million pages of SGML – encoded text and the same number of page images on the web.

- 8) What is OCR? Why is it important to OCR a digitised image?

Note : i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.8 FILE FORMATS AND MEDIA TYPES

A defined arrangement for discrete sets of data that allow a computer and software to interpret the data is called a file format. Different file formats are used to store different media types like text, images, graphics, pictures, musical works, computer programs, databases, models and designs, video programs and compound works combining many types of information. Although, almost every type of information can be represented in digital form, a few important file formats for text and images typically applicable to a library-based digital collections are described here. However, every object in a digital library needs to have a name or identifier which distinctly identifies its type and format. This is achieved by assigning file extensions to the digital objects. The file extensions in a digital library typically denote formats, protocols and rights management that are appropriate for the type of material. Names of file formats applicable in digital library and their file extensions are given in Table 7.2.

Self Check Exercise

- 9) What are file formats? How is “unstructured text” different from “structured text”?

Note : i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.8.1 Formats and Encoding Used for Text

Text and image-based contents of a digital library can be stored and presented as (i) simple text or ASCII (American Standard Code for Information Interchange); (ii) unstructured text; (iii) Structured text (SGML or HTML or XML); (iv) page description language and (v) page image formats.

Simple Text or ASCII

Simple text or ASCII (American Standard Code for Information Exchange) is the most commonly used encoding scheme used for facilitating exchange of data from one software

to another or from one platform to another. "Full-text" of articles from many journals are available electronically through online vendors like DIALOG and STN in this format since more than two decades.

Simple text or ASCII is compact, economic to capture and store, searchable, inter-operable and is malleable with other text-based services. On the other hand, the simple text or ASCII cannot be used for displaying complex tables or mathematical formulas. Photographs, diagrams, graphics, special characters cannot be displayed in ASCII. ASCII format does not store text formatting information, i.e., italics, bold, font type, font size or paragraph justification information. Simple text or ASCII in many ways is inadequate to represent many journal articles because of the reasons mentioned above. Although simple text or ASCII is extremely useful for searching and selection, its inability to capture the richness of the original makes it an interim step to structured text formats.

Structured Text Format

Structured text format attempts to capture the essence of documents by "marking-up" the text so that the original form could be recreated or even produce other forms such as ASCII. Structured text formats have provision for imbed images, graphics and other multimedia formats in the text. SGML (Standard Generalized Markup Language) is one of the most important and popular structured text format. ODA (Office Document Architecture) is a similar and competing standard. SGML is an international standard (ISO, 1986) around which several related standards are built. SGML is a flexible language that gave birth to HTML (Hyper-Text Markup Language), de facto markup language of the World Wide Web, to control the display format of documents and even the appearance of the user interface for interacting with the documents. Like simple text or ASCII, structured text can be searched or manipulated. It is highly flexible and suitable both for electronic and paper production. Well-formatted text increases visual presentation volume of textual, graphical and pictorial information. Structured formats can easily display complex tables and equations. Moreover, structured text is compact in comparison to the image-based formats, even after including imbedded graphics and pictures.

Creation of structured text, if rekeyed, is always too expensive on a production basis. However, creation of structured text is generally integrated with the production of printed artifacts. SGML is in fact, a format generated as a by-product of printed artifacts generated electronically.

Besides SGML and HTML, there are other formats used in digital library implementation. TeX, used for formatting highly mathematical text is one such format which allows greater control over the resulting display of document, including reviewing the formatting of errors.

Page Description Language (PDL)

Page Description Languages (PDLs), such as Adobe's PostScript and PDF (Portable Document Format) are similar to image but the formatted pages displayed to the user are text-based rather than image-based. PostScript and PDF formats can easily be captured during the typesetting process. PostScript is especially easy to capture since most of the systems automatically generate it and conversion program, called Acrobat Distiller, can be used to convert PostScript file into PDF files. The documents stored as PDF require Acrobat Reader at the user's end to read or print the document. The Acrobat Reader can be downloaded free of cost from the Adobe's Web Site.

Acrobat's Portable Document Format (PDF) is a by-product of PostScript. Adobe's page-description language had become the standard way to describe pages electronically in the graphics world. While PostScript is a programming language, PDF is a page-description format.

PDF can have two formats: (i) Text-based PDF that uses outline font technology of PostScript PDL (Page Description Language) from Adobe to describe format of a page; (ii) raster-scanned image PDF without the text output of OCR (Optical Character

Recognition). The image PDF is essentially equivalent to TIFF or CCITT G4 formats or to a photograph where text characters cannot be manipulated by the computer. Besides, an image-based PDF may be converted into text-based PDF once it goes through the process of OCR. In this process, scanned image is replaced by the text with fonts and layout matching with the scanned document.

Self Check Exercise

10) How is Page Description Language (PDL) different from structured text?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

Page Image Format

The digitally scanned images are stored in a file as a bit-mapped page image, irrespective of the fact that a scanned page contains a photograph, a line drawing or text. The bit-mapped page image can be created in dozens of different formats depending upon the scanner and its software. National and international standards for image-file formats and compression methods exist to ensure that data will be interchangeable amongst systems. An image file stores discrete sets of data and information allowing a computing system to display, interpret and print the image in a pre-defined fashion. An image file format consists of three distinct components, i.e., **header** which stores information on file identifier and image specifications; **Image data** consisting of look-up table and image raster and lastly, **footer** that signals file termination information. While bit-mapped portion of a raster image is standardised, it is the file header that differentiates one format from another.

TIFF (Tagged Image File Format) is the most commonly used page image file format and is considered to be the de facto standard for bitonal images. Some image formats are proprietary developed by commercial vendors and require specific software or hardware for display and printing. Images can be coloured, grey-scale or black and white (called bitonal). They can be uncompressed (raw) or compressed using several different compression algorithms.

Table 7.2: File Formats Used in a Digital Library

Abbreviation	Format	File Extension
File Format for Unstructured Text		
ASCII	American Standard Code for Information Interchange	.txt
File Format for Structured Text		
SGML	Standard Generalized Markup Language	.sgml
HTML	Hypertext Markup Language	.html
XML	Extended Markup Language	.xml
PDF	Portable Document Format (Adobe)	.pdf
PostScript	PostScript (Adobe)	.ps
TEX	Texture Format	.txt

Abbreviation	Format	File Extension
File Format for Images		
PDF	Portable Document Format	.pdf
BMP	Bit Map Page (Windows)	.bmp
IMG	Ventura Publisher	.img
JPEG	Joint Photographic Expert Group	.mpg
JFIF	JPEG File Format	.jfif
PCP	PC Paint (B&W Mode)	.pcp
PCX	PC Paint Brush (Color & B&W)	.pcx
PSD	Photoshop	.psd
TGA	True Vision Targa	.tga
PNG	Portable Network Graphic	.png
TIFF	Taged Image File Format	.tif
TIFF-G4	Taged Image File Format with Group 4 Fax Compression	.tif
SPIFF	Still Picture Interchange File Format	.spf
PCD	Photo CD (Kodak)	.pcd
Audio and Video File Format		
WAVE	Waveform Audio (Microsoft)	.wav
AIFF	Audio Interchange Format	.aif
VoC	Creative Voice	.voc
MIDI	Musical Instrument Digital Interface	.midi
SND	Sound	.snd
AU	Audio (Sun Microsystems)	.au
RAF	Real Audio Format (Progressive Networks)	.ra
AVI	Audio Visual Interleave	.avi
FLA	Macromedia Flash Movie	.fla
FLC	AutoDesk FLIC Animation	.flc
MOV	Quicktime for Windows Movie	.mov
MPEG	Motion Picture Expert Group	.mpg
MP2	MPEG Audio Layer 2	.mp2
MP3	MPEG Audio Layer 3	.mp3

7.9 TOOLS OF DIGITISATION

Digital imaging is an inter-linked system of hardware, software, image database and access sub-system with each having their own components. Tools used for digitisation include several core and peripheral systems. An image scanning system may consist of a stand-alone workstation where most or all the work is done on the same workstation or as a part of a network of workstations with imaging work distributed and shared amongst various workstations. The network usually includes a scanning station, a server and one or more editing, retrieval stations. A typical scanning workstation for a small, production-level project could consist of the following:

- Hardware (Scanners, computers, data storage and data output peripherals)
- Software (image capturing and image editing)
- Network (data transmission)
- Display and Printing technologies

This Unit concentrates on scanners and scanning software as important components of the scanning system.

7.9.1 Scanners

Digital scanners are used to capture digital images from analogue media such as printed pages or a microfiche / microfilm at a predefined resolution and dynamic range (bit range). There are two types of image scanners: vector scanner and raster scanners. The vector scanners scan an image as a complex set of x,y coordinates. Vector images are generally used in Geographical Information Systems (GIS). The display software for the vector image interprets the image as function of coordinates and other included information to produce an electronic replica of the original drawing or photograph. Vector images can be zoomed in portion to display minute details of a drawing or a map. Maps, engineering drawings, and architectural blueprints are often scanned as vector images. Raster images are captured by raster scanners by passing lights (laser in some cases) down the page and digitally encoding it row by row. Multiple passes of lights may be required to capture basic (as a set of bits known as bit map) colours in a coloured image. Raster scanners are used in libraries to convert printed publications into electronic forms. Majority of electronic imaging systems generate raster images. The scanners used for digitizing analogue images into digital images come in a variety of shapes and sizes.

How Scanner Works?

Scanners are equipped with a lamp that moves with the scanner head to light-up the object being scanned. Most scanners use a cold cathode florescent lamp or a xenon lamp. The scan head is made up of the mirrors, lens, filter, and charge-coupled devices (CCD) array. A belt that is connected to the stepper motor makes the scan head move. A stabilizing bar prevents wobbling during the pass. The mirrors reflect what is being scanned into the lens and the image is then focused through a filter on the CCD array. Three smaller images of the original are made by the lens. These images then go through a color filter and onto a section of the CCD array. The data is then combined into a single image.

While selecting a scanner, one should consider resolution, sharpness, and rate of image transfer. The resolution is measured in dots per inch (dpi). The average scanner has at least 300x300 dpi. The number of sensors in a row of the CCD array determines a scanner's dpi. Sharpness depends on how bright the lamp is and the quality of the lens. Image transfer depends on the connection used to connect the scanner to the computer. The slowest is the parallel port. Universal Serial Bus or USB scanners are affordable, easy to use, and have good speed.

The hardware required for a scanner is a connector such as a USB. The software required is a driver. The driver is needed to communicate with the scanner. TWAIN is the language spoken by scanners. Any program that supports TWAIN can acquire a scanned image.

There are following types of Scanners:

- Flatbed Scanners – right angle, prism and overhead flatbed
- Sheet-Feed Scanners
- Drum Scanners
- Digital Cameras
- Slide Scanners
- Microfilm Scanners
- Video Frame Grabbers
- Hand-held scanners

The type of scanner selected for an imaging project would be influenced by the type, size and source of documents to be scanned. Many scanners can handle only transparent material, whereas others can handle only reflective materials.

Flatbed Scanners: Flatbed scanners are most common, and widely used scanners that look like a photocopier and are used in much the same way. Source material in a flatbed scanner is placed face down for scanning. The light source and charge-coupled devices (CCDs) move beneath the platen, while the document remains stationary as in the case of photocopying machine. Flatbed scanner comes in various models like right-angle, prism and planetary/overhead to handle bound volumes and books. Flatbed scanner can scan usually a document at 600 dpi. Many flatbed scanners however, offer higher resolution.

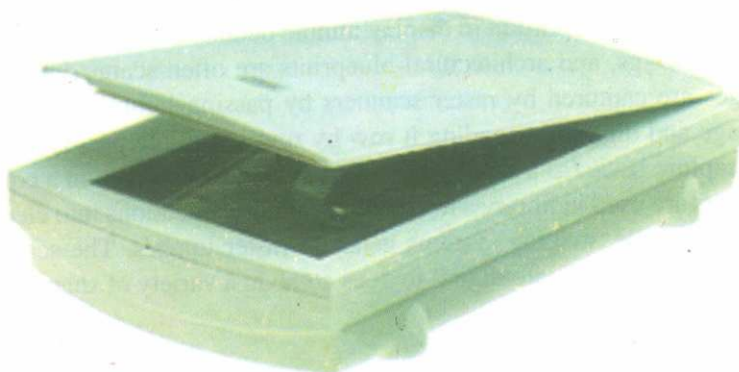


Fig. 7.9: Flatbed Scanner

Sheet feed Scanners: In a sheet-feed scanner, as is indicated in the name, document is fed over a stationary CCD and light source via roller, belt, drum, or vacuum transport. In contrast to a flat-bed scanner, sheet-feed scanner has optional attachment of auto-feed uniform-sized stack of documents to be scanned.

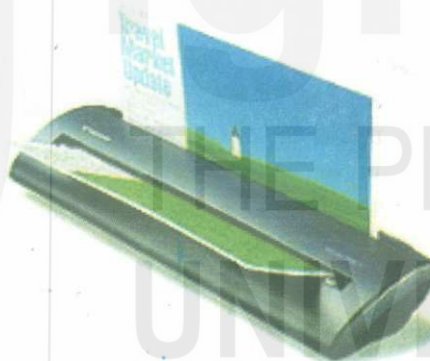


Fig. 7.10: Sheet feed Scanners

Drum Scanners: Source material in a drum scanner is wrapped on a drum, which is then rotated past a high-intensity light source to capture the image. Drum scanners offer superior image quality, but require flexible source material of limited size that can be wrapped around the photosensitive drum. Drum scanners are specially targeted for graphic arts market. Drum scanners offer highest resolution for grey scale and colour scanning. Drum Scanner uses Photo-Multiplier (Vacuum) Tubes (PMTs) instead of CCDs, which offer a greater bit depth (12 to 16 bits).

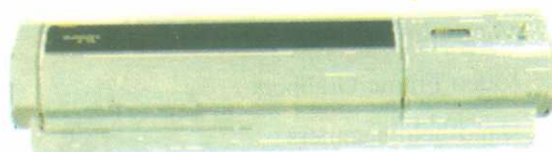


Fig. 7.11: Drum Scanners

Digital Cameras: Digital cameras mounted on copy cradle resemble microfilming stand. Source material is placed on the stand and the camera is cranked up or down in order to focus the material within the field of view. Digital cameras are most promising scanner development for library and archival applications.

Digitisation: Concept,
Need, Methods and
Equipment



Fig. 7.12: Digital Camera

Slide Scanner: Slide scanners have a slot in the side to accommodate a 35mm slide. Inside the box, the light passes through the slide to hit a CCD array behind the slide. Slide scanners can generally scan only 35mm transparent source materials.

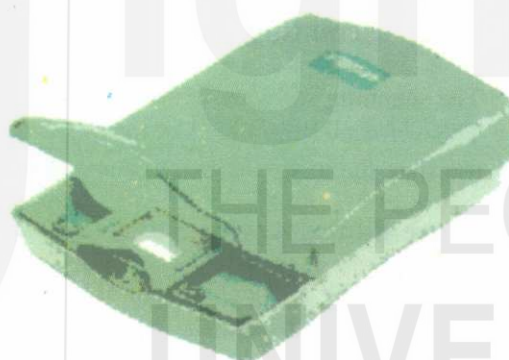
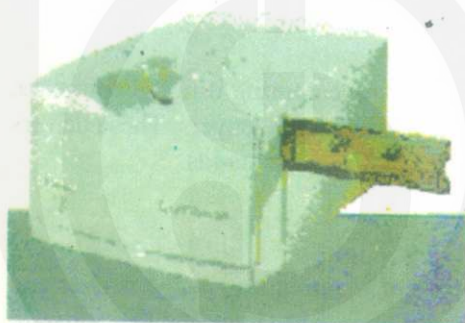


Fig. 7.13: Drum Scanner

Microfilm Scanner: Specially targeted to library/archival application, microfilm scanners have adapters to convert roll film, fiche, and aperture cards into the same model.

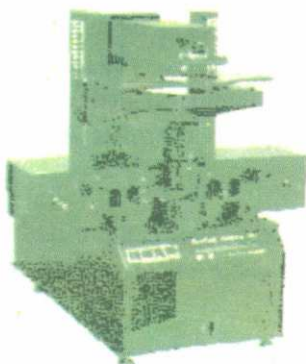


Fig. 7.14: Microfilm Scanner

Video Frame Grabber or Video Digitiser: Video digitisers are circuit boards placed inside a computer and attached to a standard video camera. Any thing that is filmed by the video camera is digitised by the video digitiser.

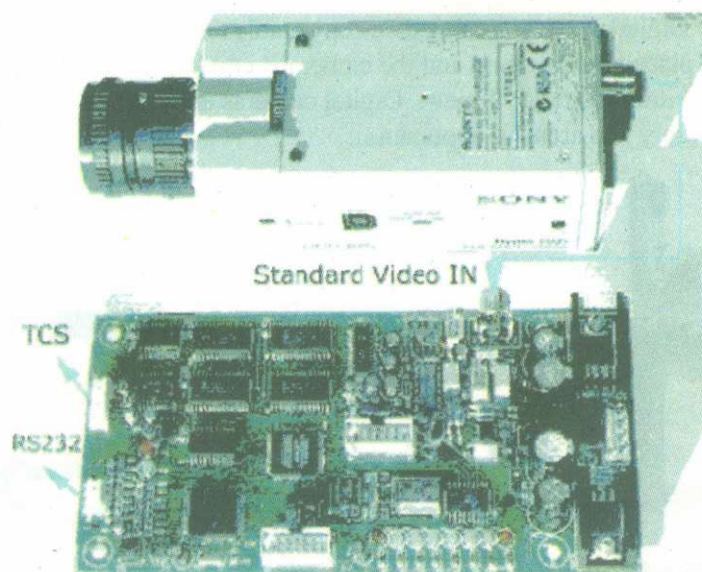


Fig. 7.15: Video Grabber

Hand-held Scanners

Hand-held scanners are used for scanning selective sections of data. It may require multiple pass to capture large area. Moreover, a user should have a steady hand while moving the scanner over the document to be scanned. These scanners are normally used for circulation work in a library.

7.9.2 Scanning Software

The scanning software is used for scanning the image and capturing it in the computer. This software is provided by the manufacturer of the product to the buyers. These drivers translate the instructions into commands, which the scanner understands.

Image Editing Applications

Image editing applications are used once the process of scanning the image is over and the image is available in the computer for further manipulation. Most image editing software offer features like image editing, sharpening, filter, cropping, colour adjustments, forms conversion, resising, etc. Most image editing software can also be used for capturing the images.

7.10 DIGITISATION OF AUDIO AND VIDEO

The songs or speeches that we generally listen from tape recorders or radio are in an analogue form. The analogue sound tracks can be digitized by attaching an audio player to a system through an audio capture card so as to record the sound to the system. The audio files can be saved as wav, mp3, midi, etc. MP3 format is highly compact and the sound quality is better in comparison to other formats. Audio files can be further processed using noise reduction software.

Like audio, video capture also requires a video capture card with input from video cassette player (VCP / VCR), TV antenna, cable or movie camera. The digitised files can be saved as mov, avi, mpg file formats.

7.11 ORGANISING DIGITAL IMAGES

A disc full of digital images without any organisation, browse and search options may have no meaning except for one who created it. Scanned images need to be organised in order to be useful. Moreover, images need to be linked to the associated metadata to

facilitate their browsing and searching. The following three steps describe the process of organising the digital images:

Organise the scanned image files into disc hierarchy that logically maps the physical organisation of the document. For example, in a project on scanning of journals, create a folder for each journal, which, in turn, may have folder for each volume scanned. Each volume, in turn, may have a subfolder for each issue. The folder for each issue, in turn, may contain scanned articles that appeared in the issue along with a content page, composed in HTML providing links to articles in that issue.

Name the scanned image files in a strictly controlled manner that reflects their logical relationship. For example, each article may be named after the surname of first author followed by a volume number and an issue number. For example, file name "smithrkv5n1.pdf" conveys that the article is by "R.K. Smith" that appeared in volume 5 and issue no.1. The file name for each article would, therefore, convey a logical and hierarchical organisation of the journal.

Describe the scanned images file internally using image header and externally using linked descriptive metadata files. The following three types of metadata are associated with the digital objects:

- i) **Descriptive Metadata:** Include content or bibliographic description consisting of keywords and subject descriptors.
- ii) **Administrative or technical Metadata:** Incorporates details on original source, date of creation, version of digital object, file format used, compression technology used, object relationship, etc. Administrative data may reside within or outside the digital object and is required for long-term collection management to ensure longevity of digital collection.
- iii) **Structural Metadata:** Elements within digital objects facilitate navigation, e.g., table of contents, index at issue level or volume level, page turning in an electronic book, etc.

The simplest and least effective method for providing access is through a table of contents and linking each item to its respective object / image. Content pages of issues of journals done in HTML would offer browsing facility. Full-text search to HTML pages or OCR'd pages can be achieved by installing one of the free Internet search engines like Oingo Free Search (http://www.oingo.com/oingo_free_search/products.html); Swish-E (<http://www.berkeley.edu/SWISH-E/>); WhatyUseek (<http://intra.whatuseek.com/>); Excite (<http://excite.com/>) and Google (<http://www.google.com>).

Large scanning projects would, however, require a back-end database storing images or links to the images and metadata (descriptive / administrative). Back-end database used by most document management systems holds the functionality required by most web applications. Important management systems like File Net have now integrated their database with HTML conversion tools. Further, some of the document management systems have also signed up with Adobe to incorporate Acrobat and Acrobat Capture into their web-based document management systems. These databases entertain queries from users through "HTML forms" and generate search results on the fly. Several digital library packages are now available as "open source" or "free-ware" that can be used not only for organizing the digital objects but also for their search and retrieval.

7.12 DIGITAL LIBRARY SOFTWARES

Several digital library softwares are currently available like, Greenstone Digital Library (GSDL), Dspace, Eprints, Fedora, etc., which are available freely for download on the Internet. There are some commercial Digital Library Softwares available but none has been used on large scale in comparison to the ones mentioned above. We will provide here a brief account of digital library softwares available in public domain. The description is an outline of the features of the respective softwares.

Dspace

Dspace (www.dspace.org) has been developed in partnership between Hewlett Packard (HP) and Massachusetts Institute of Technology (MIT). Development is still in progress but as an institutional repository software. Dspace is making its mark, with an increasing number of institutions around the globe installing, evaluating, and using the package. The latest stable version is 1.2 available for download at the Dspace web site.

Currently, the original developers undertake most of the core development, but a growing technical user base is generating suggestions for future releases as well as looking for producing some add-on modules. In addition the Dspace Federation is guiding the transition of this software to a more community-wide open-source development model.

Dspace captures, stores, indexes, preserves, and, redistributes the intellectual output of a university's research faculty into digital formats. Dspace accepts all forms of digital materials including text, images, video, and audio files. Possible content includes: articles and preprints; technical reports; working papers; conference papers; e-theses; datasets (statistical, geospatial, matlab, etc.); images (visual, scientific, etc.); audio files; video files; learning objects; and reformatted digital library collections. The back end technologies used include: Apache, Tomcat, OpenSSL/mod_ssl; Java 1.3, JSP 1.2, Servlet 2.3; PostgreSQL 7, JDBC (RDBMS); CNRI handle System 5 (persistent ids); Lucene 1.2 (index/search).

Prerequisites

Dspace depends upon the Java programming language and the PostgreSQL open source database system. It also requires a number of additional Java-based elements to be installed: Tomcat, which is a Java based server; a number of Java code libraries; and the Ant, a Java compiler. It is recommended that Dspace be installed on a Linux or a Unix machine. It requires an experienced system administrator to do the prerequisite installation.

E-prints

GNU Eprints 2.x is a free software which creates online archives (<http://software.eprints.org/>). The default configuration creates a research paper archive. With its origin in the scholarly communication movement, e-print default configuration is geared to research papers but it can be adapted to other purposes and content. It was developed in the Intelligent Agents, Multimedia Group at the Electronics and Computer Science Department of the University of Southampton.

GNU Eprints is freely distributed to the GNU General Public License. The latest version is 2.3 and is available for download at <http://software.eprints.org/download.php>

Prerequisites

- Any computer capable of running GNU/Linux or similar operating system. The faster, the better, but any Intel Pentium II processor will give good performance.
- A GNU operating system. GNU/Linux (a very advanced and free UNIX-like operating system) works just fine, and is in fact the development platform
- Apache WWW server
- Perl programming language, also a number of additional modules
- mod_perl module for Apache, which significantly increases the performance of Perl scripts
- MySQL Databases

Greenstone Digital Library (<http://greenstone.org>)

Greenstone Digital Library is an open-source software available under the terms of the GNU General Public License. It has the ability to serve digital library collections and build

new collections. It provides a new way of organising information and publishing it on the Internet or on CD-ROM. The Greenstone Digital Library software is produced by the New Zealand Digital Library Project at the University of Waikato, and distributed in cooperation with UNESCO and the Humanities Library Project. The New Zealand Digital Library Web site (<http://nzdl.org>) contains numerous example collections, all created with the Greenstone software, which are publicly available for anyone to peruse. The Greenstone runs on Windows and Unix platforms. The distribution includes ready-to-use binaries for all versions of Windows and for Linux. It also includes complete source code for the system, which can be compiled using Microsoft C++ or gcc. Greenstone works with associated software that is also freely available: the Apache Web server and PERL.

Ganesha Digital Library

Ganesha Digital Library version 3.1 (GDL) (<http://gdl.itb.ac.id/>) is another open source software developed under Indonesian Digital Library Network (IndonesiaDLN). Ganesha Digital Library enables institutions or individuals to share their knowledge and also access and utilise available knowledge in the Indonesian 'giant memory' through the network of IndonesiaDLN digital libraries. The software is available in three publisher editions: Personal, Internet Cafe, and Institution. Released under the terms of the GNU GPL, (GNU General Public License).

7.13 PLANNING AND IMPLEMENTATION

Digitisation is the first step towards building a digital library. It is highly specialised and cost-intensive activity that requires inputs from diverse branches of knowledge. It is important that objectives, needs and the purpose of digitisation are established clearly. The digitisation proposal should therefore define its goals (objectives) scope, feasibility, benefits, costs, time required for the developmental phase, implementation issues, deliverables and target users. It may be desirable to continue with traditional collections and also acquire collections in digital media; (ii) buy access to electronic resources; and (iii) develop subject gateways or library portals, instead of undertaking digitisation project. This approach would save the cost and efforts on digitisation and other recurring administrative costs. However, once a decision for digitisation is taken, due importance should be given to factors such as sustenance, reusability, interoperability, verification and documentation both for users as well as for the developer. The steps given below may be considered as pre-requisites. Careful planning of digitisation would bring activity to the project, save cost, time and human efforts. The Planning would include the following steps:

7.13.1 Feasibility

First, it is necessary to conduct a feasibility study of the digitisation project. The feasibility should be established not only in terms of the availability of tools, and expertise, but also the factors like volume/number of documents to be covered in the process of digitisation, target audience, demand for material to be digitised and user's requirements. The study should also assess whether the library can take-up the project in-house or should it be out-sourced.

7.13.2 Planning the Project

The planning of the project needs to cover the following areas:

Managerial Planning

Managerial planning would essentially involve the process of sequencing various tasks, their time management and project monitoring. Activities that need managerial planning may include conducting feasibility study, procurement of equipment, recruitment of manpower, digitisation (whether out-sourced or done-in-house), IPR and rights management issues, integration and organisation of content, finding market, launching and marketing of services. Flow diagrams, PERT, CPM and SWOT analysis and other management techniques may be deployed at this stage.

Hardware and Software Planning

In this, the requirements of hardware and software for the servers and network components may be worked out with their financial implications and network components. Connectivity and the bandwidth required for hosting the digitised collection may also be planned. The technical specifications should be worked out much before the actual process of digitisation commences irrespective of whether digitisation is done in-house or outsourced. For this, the existing types, formats, standards and practices are to be reviewed first. Draft specifications are to be prepared and these are to be tested with sample data. Necessary modifications may be made in the specifications based on this testing laid down for metadata creation for digital objects as well as for the digital collection. Digital objects and digital collections typically require descriptive (keywords / descriptors), structural (navigation, content pages etc.) and administrative (formats, compression, standards, etc.) metadata.

Human Resources Planning

Human resources needed to be worked out in terms of staff time involved, training of existing staff and recruitment of new staff with desired skills. Human resource planning would depend on whether the library is going for in-house digitisation or for outsourcing the process of digitisation. Project management continues to be an important issue even if the digitisation work is outsourced. The management of the project may be divided in groups with responsibilities defined. Communication between the groups and a reporting structure may be laid down to facilitate unambiguous communication among the groups and the staff.

Financial planning

Financial planning is crucial. Cost of migration from one medium to another and from one computer to another may be built-in. Cost of hosting the services and their maintenance should also be planned besides other aspects mentioned above.

7.13.3 Purchase of Hardware and Software

Choice of technology and the equipment required may be made. These include storage and back up devices, network equipment required, software for search and access and other related items. The software may be acquired or developed inhouse. The following steps may be followed in this regard:

- i) Acquire and install hardware and software;
- ii) Acquire and install the network required for hosting the digitised collection. Consider bandwidth requirements that depend upon the media offered by the digital library. While simple text requires relatively low bandwidth to deliver content, images and video require large bandwidth; and
- iii) Acquire and install other components

7.13.4 Selection of Material for Digitisation and 'Born Digital'

In the process of execution of the project, the first task is to identify, select, and to prioritise the documents that are to be digitised. If the organisation is itself creating contents, strategies are to be laid down to capture 'born digital' data. If documents are available in digital form, they can be easily converted to other formats. If the selected material is from external sources, IPR issues need to be resolved. It is necessary to obtain permission from the publishers and data suppliers for digitisation, if material being digitised is not available in public domain. Moreover, decision may be taken whether to OCR the digitised images. Documents selected for digitisation may already be available in digital format. It is always economical to buy e-media, if available than their conversion. Moreover, oversized material, deteriorating collections, bound volumes of journals, manuscripts, etc., would require highly specialized equipment and highly specialised manpower.

7.13.5 Placement and Training of Manpower

Since the entire job of developing and or maintaining a digital library is a highly skilled one, there should be no compromise or slackening in the quality of intake or selection of manpower for the job. Also, even if good quality manpower is positioned, they usually need training to upgrade and sharpen their skills for this job. So, necessary training, should form a component of the execution of the project.

7.13.6 Content Creation

The steps involved in content creation include the following:-

- Conversion of datasets that are 'born digital', for example, convert MS Word file into PDF;
- Conversion of the existing printed sections into digital format (digitisation); and
- Identification of vendors if the digitisation work is to be outsourced.

7.13.7 Execution of the Project

Once the equipment and software and other infrastructure facilities are installed or positioned, and the priorities of the documents for digitisation laid down, the execution of the project starts. The library may use digital library software like greenstone Digital library, or Dspace, etc.

Self Check Exercise

11) How do you digitise audio and video? What are the devices used for it?

- Note:** i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.14 SUMMARY

Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) into digital format. In most libraries, digitisation normally results in making documents accessible from the web sites of the libraries. Optical scanners and digital cameras are used to digitise images by translating them into bit maps. It is also possible to digitise sound, video, graphics, animations, etc.

Digitisation is the first step in the process of building digital libraries. Digitisation is also used for achieving preservation and archiving although it is not considered a good option for this purpose. It is highly labour-intensive and cost-intensive process that involves several complexities including copyright and IPR issues. However, digital objects offer numerous benefits in terms of accessibility and search. The documents to be digitised may include text, line art, photographs, colour images, etc. The selection of document needs to be made carefully considering all the factors of utility, quality, security and cost. Rare and much-in-demand documents and images are selected as first priority even though the quality is not good.

The process of digitisation involves four steps, namely, scanning, indexing, storage and retrieval. A scanned document is nothing more than a picture of a printed page. It cannot

be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred by their labels rather than characters in the documents. OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing files. OCR or text recognition is the process of electronically identifying the text in a bit-mapped page image or set of images and generate a file containing text in ASCII code or in a specified word processing format leaving the image intact in the process.

The quality of digital image can be monitored at the time of capture by four factors, namely: i) bit depth / dynamic range; ii) resolution; iii) threshold; and iv) image enhancement. The Unit describes these parameters in detail. Image files are evidently larger than textual files. It is thus necessary to compress image files. Image compression is the process of reducing size of an image by abbreviating the repetitive information such as one or more rows of white bits to a single code. The compression algorithms may be grouped as lossless compression and lossy compression. The Unit also describes compression technology and protocols.

Text and image-based contents of a digital library can be stored and presented as: i) simple text or ASCII (American Standard Code for Information Interchange); ii) unstructured text; iii) structured text (SGML or HTML or XML); iv) page description language and v) page image formats. The Unit also describes file formats in digitised collection.

An image scanning system would normally consist of a stand-alone workstation where most or all the work is done on the same workstation or as a part of a network of workstations with imaging work being distributed and shared amongst various workstations. The network usually includes a scanning station, a server and one or more editing and retrieval stations. A typical scanning workstation for a small production-level project, could consist of the following:

- Hardware (scanners, computers, data storage and data output peripherals)
- Software (image capturing and image editing)
- Network (data transmission)
- Display and Printing technologies

This Unit describes scanners and scanning software as important components of the scanning system.

7.15 ANSWERS TO SELF CHECK EXERCISES

- 1) Digitisation is the process of converting the content of physical media (e.g., printed materials, images, photographs, microforms, etc.) into digital format. It refers to the process of translating a piece of information such as a book, journal articles, sound recordings, pictures, audio tapes or videos recordings, etc. into bits. Converting information into these binary digits is called digitisation. Digitising a document in print or other physical media (e.g., sound recordings) makes the document more useful as well as more accessible. It is possible for a user to conduct a full-text search on a document that is digitised and OCRed. It is possible to create hyperlinks to lead a reader to related items within the text itself as well as to external resources. Ultimately, digitisation does not mean replacing the traditional library collections and services; rather, it serves to enhance them.
- 2) The process of selection of material for digitisation involves identification, selection and prioritisation of documents that are to be digitised. The selection of document need to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and much in demand documents and images are selected as first priority without considering the quality. If the selected material is from the external sources, IPR issues need to be resolved. If material being digitised is not available in

public-domain then it is important to obtain permission from the publishers and data suppliers for digitisation. Moreover, decision may be taken whether to OCR the digitised images. Documents selected for digitisation may already be available in digital format. It is always economical to buy e-media, if available, than their conversion. Moreover, over-sized material, deteriorating collections, bound volumes of journals, manuscripts, etc. require highly specialized equipment and highly skilled manpower. The documents to be digitised may include text, line art, photographs, colour images, etc.

) Steps in the Process of Scanning using a Flatbed Scanner are as follows:

- Step 1. Place picture on the scanner's glass
- Step 2. Start scanner software
- Step 3. Select the area to be scanned
- Step 4. Choose the image type
- Step 5. Sharpen the image
- Step 6. Set the image size
- Step 7. Save the scanned image using a desirable format (GIF or JPEG)

- 4) Bitonal or black & white or binary scanning is used in libraries to scan pages containing text or the drawings. Bitonal or binary scanning represents one bit per pixel (either "0" (black) or "1" (white). Gray scale scanning, on the contrary, is used for reliable reproduction of intermediate or continuous tones found in black & white photographs to represent shades of grey. Multiple numbers of bits ranging from 2-8 are used to for each pixel to represent shades of grey. Although each bit is either black or white, as in the case of bitonal images, but bits are combined to produce a level of grey in the pixel that is, black, white or somewhere in between. Since only one bit (either black or white) is used to represent a pixel in bitonal scanning, bit depth is not important in bitonal scanning.
- 5) The resolution of an image is defined in terms of number of pixel (picture elements) in a given area. It is measured in terms of dot per inch (dpi) in case of an image file and as ratio of number of pixel, on horizontal line x Number of pixel in vertical lines in case of display resolution on a monitor. Higher the dpi is set on the scanner, the better the resolution and quality of image and larger the image file. (printed resolution versus screen resolution)
- 6) Image compression is the process of reducing size of an image by abbreviating the repetitive information such as one or more rows of white bits to a single code. The compression algorithms may be grouped into the following two categories:

Lossless Compression: The conversion process converts repeated information as a mathematical algorithm that can be decompressed without losing any details into the original image with absolute fidelity. No information is "lost" or "sacrificed" in the process of compression. Lossless compression is primarily used in bitonal images.

Lossy Compression: Lossy compression discards or minimizes details that are least significant or which may not make appreciable effect on the quality of image. This kind of compression is called "lossy" because when the image that is compressed using "Lossy" compression techniques is decompressed, it will not be an exact replica of the original image. Lossy compression is used with grey-scale/ colour scanning.

- 7) LZW compression technique uses a table-based lookup algorithm invented by Abraham Lempel, Jacob Ziv, and Terry Welch. A particular LZW compression algorithm takes each input sequence of binary digit of a given length (for example, 12 bits) and creates an entry in a table (sometimes called a "dictionary" or "codebook") for that particular bit pattern, consisting of the pattern itself and a

shorter code. As input is read, any pattern that has been read before the results in the substitution of the shorter code effectively compresses the total amount of input to something smaller. The decoding program that uncompresses the file is able to build the table itself by using the algorithm as it processes the encoded input. LZW compression is suitable for compressing both graphical as well as text files. Two commonly-used file formats in which LZW compression is used are the Graphics Interchange Format (GIF) and Tag Image File Format (TIFF).

- 8) OCR (Optical Character Recognition) or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generate a file containing that text in ASCII code or in a specified word processing format leaving the image intact in the process. OCR programs are software tools that are used to transform scanned textual page images into word processing file. The process of OCR is used in order to make every word in a scanned document readable and fully searchable without having to key-in everything in the computer manually. Once a bit-mapped page image has gone through the process of OCR, a document can be manipulated and managed by its contents, i.e. using the words available in the text. OCR does not actually convert an image into text but rather creates a separate file containing the text while leaving the image intact.
- 9) A defined arrangement for discrete sets of data that allow a computer and software to interpret the data is called a file format. Different file formats are used to store different media types like text, images, graphics, pictures, musical works, computer programs, databases, models and designs video programs and compound works combining many type of information. Every file in electronic format needs a name or identifier that distinctly identifies its type and format. The file extensions are used to denote formats, protocols and right management that are appropriate for the type of material. For example file with extension .doc can be recognized as a Microsoft word file and a file with an extension .gif can be recognized as an image file stored in Graphical Interface Format (GIF).

Simple text or ASCII is a compact and economic file format that is used to capture and store textual data. However, the simple text or ASCII can neither be used for displaying complex tables or mathematical formulas nor can it be used to display photographs, diagrams, graphics, and special characters. Text or ASCII format does not store text formatting information, i.e., italics, bold, font type, font size or paragraph justification information. Simple text or ASCII is, therefore, inadequate to represent journal articles because of these reasons. Structured text, on the other hand, attempts to capture the essence of documents by "marking-up" the text so that the original form can be recreated or even produce other forms such as ASCII. Structured text format have provision display italics, bold, font size, font type and paragraph justification. It is capable of imbedding images, graphics and other multimedia formats in the text. SGML (Standard Generalized Markup Language), HTML, .DOC are some of the important and popular structured text format. Like simple text or ASCII, structured text can be searched or manipulated. It is highly flexible and suitable both for electronic and paper production. Well-formatted text increase visual presentation of textual, graphical and pictorial value of information. Structured formats can easily display complex tables and equations. Moreover, the structured text is compact in comparison to the image-based formats, even after including imbedded graphics and pictures.

- 10) Page description Language (PDLs), such as Adobe's PostScript and PDF (Portable Document Format) is a page description format that essentially has all the information required to display or print that page. PDLs retain all the information about what that page is supposed to look like and how it should be printed. In addition to fonts, PDF represents all the other visual aspects of the page including line breaks, layout, white space, graphics, colours; i.e. every visual feature of the page. A PDF is essentially similar to image but the formatted pages displayed to the user are text-based rather than image-based. PostScript and PDF formats can easily be captured

during the typesetting process. A PDL can be generated from almost all structured formats using various software packages although reverse process is not possible. Moreover, a structured text file can be edited and manipulated, a PDL file can not be edited or manipulated directly.

- 11) An audio file that we generally listen from tape recorders or radio is in an analogue form. The analogue sound tracks can be digitised by attaching an audio player to a system through an audio capture card so as to record the sound to the system. The audio files can be saved as .wav, mp3, midi, etc. MP3 format is highly compact and the sound quality is better in comparison to other formats. Audio files can be further processed using noise reduction software.

Like audio, video capture also requires a video capture card with input from video cassette player (VCP / VCR), TV antenna, and cable or movie camera. The digitised files can be saved as .mov, .avi, and .mpg file formats.

7.16 KEYWORDS

ASCII : American Standard Code for Information Interchange or ASCII is a standard coding technique for representing computer information.

Analogue : A term used to describe a signal, such as the human voice and electric current, whose value varies continuously with time or transmission method, such as the traditional telephone network, which carries source signals as electrical waves. Compared with digital systems, an analogue telephone line carries data at low speed; it also requires a modem to convert the computer's digital output into a form (analogue) which it can handle.

Bit and Byte : Bit is short term for binary digit, the smallest unit of information on a machine. A single bit can hold only one of two values: 0 or 1. More meaningful information is obtained by combining consecutive bits into larger units. A byte is composed of 8 consecutive bits.

Bit Depth : The number of bits used to represent each pixel. The greater the bit depth, the more colours or grey-scales can be represented. For example, a 24-bit colour scanner can represent 2 to the 24th power (16.7 million) colours.

Bitonal : Each pixel in a bitonal image is represented by a single bit, i.e. black and white. Textual documents and line drawings are scanned in bitonal, i.e. a pixel is either black or white whereas each pixel for a picture (black and white or colour) may contain 2-8 bit per primary colour.

Capture : A term used in document imaging for scanning of a document or any other artefact.

Crop : The process of elimination of a portion of a picture, illustration or photograph that contain unnecessary material or to highlight a certain area of the image.

Digitisation : The process of converting data and information into digital format is called digitisation. It is synonymous with scanning, it is the conversion from printed paper, film, or some other media, to an electronic form where the page is represented as either black and white dots, or color or grayscale pixels.

- Document Scanning** : Document scanning is the process by which print and film documents are fed into a scanner and converted into electronic documents. During the scanning process documents can be OCR'd and indexed to insure quick retrieval at a later date.
- Dots Per Inch (DPI)** : Dots per inch (dpi) indicates the resolution of images or printers. The more dots per inch, the higher the resolution
- Dynamic Range** : The number of colours or shades of grey that can be represented by a pixel. Dynamic range is a measurement of the number of bits used to represent each pixel in a digital image. 1-bit or bitonal means that a pixel can either be black or white. Bitonal imaging is good for black and white images, such as line drawings and text. However, scanning in grey-scale rather than bitonal may produce a better looking image. 8-bit color or 8-bit grey-scale means that each pixel can be one of 256 shades of colour or one of 256 shades of grey. 24-bit colour means that each pixel can be one of 16.8 million colours.
- Halftone** : A method of generating an image that requires varying densities or shades to accurately render the image. This is achieved by representing the image as a pattern of dots of varying size. Larger dots represent darker areas, and smaller dots represent lighter areas of an image.
- Image Sharpening** : Scanned images can be adjusted to increase edge contrast and artificially enhance the overall quality of image. Most paint and colour manipulation programs have special tools to selectively sharpen isolated areas of an image.
- Metadata** : Data about data, or information known about the image in order to provide access to the image. Usually includes information about the intellectual content of the image, digital representation data, and security or rights management information.
- Optical Character Recognition (OCR)** : The OCR refers to the process of scanning text using a scanning device from a printed page into an image and translate it into a computer processible format, i.e. an ASCII file. OCR systems include an optical scanner for reading text and sophisticated software for analyzing images.
- Pixels** : The term "pix" means "part of a picture" and "el" means "from element". In bitonal (black and white) display, each pixel can have only one bit, i.e. either black or white, whereas in a gray-scale display, each pixel can have three numerical values for three colour, i.e. Red, Green, Blue (RGB) to represent the colour. These three RGB components can be represented by three 8-bit numbers for each pixel. Three 8-bit bytes (one byte for each of RGB) is called 24 bit colour. Each 8 bit RGB component can have 256 possible values, ranging from 0 to 255. For example, three values like (250, 165, 0), meaning (Red=250, Green=165, Blue=0) to denote one Orange pixel. The pixels are most commonly used to represent images as a computer file. Pixel are of a uniform size and shape.

Portable Document Format (PDF) : Portable Document Format is a type of formatting that enables files to be viewed on a variety of computers regardless of the program originally used to create them. PDF files retain the "look and feel" of the original document with special formatting, graphics, and colour intact. A special program or print driver (Adobe Distiller or PDF Writer) is used to convert a file into PDF format. The Acrobat Reader program available free from the Adobe site.

PostScript : A page description language developed and marketed by Adobe Systems. PostScript can be used by a wide variety of computers and printers, and is the dominant format used for desktop publishing. Documents in PostScript format are able to use the full resolution of any PostScript printer, because they describe the page to be printed in terms of primitive shapes which are interpreted by the printer's own controller. PostScript is often used to share documents on the Internet because of this ability to work on many different platforms and printers.

Primary Colour : Colours that are basis for all other colour combinations. Primary colours are red, green and blue (RGB).

RGB : Red, Green and Blue (RGB) are the colours that are basis for all other colour combinations. They are called primary colours.

Resolution : Resolution refers to the number of pixels contained on a display monitor, printers and bit-mapped graphic images. In the case of printers and images, it is indicated the number of dots per inch. For graphics monitors, the screen resolution signifies the number of dots (pixels) on the entire screen. Resolution generally refers to the sharpness and clarity of an image.

Scanner : An optical input device that uses light-sensing equipment to capture an image on paper or some other subject. The image is translated into digital signals that can then be manipulated by optical character recognition (OCR) software or graphics software. Scanners come in a number of types, including flatbed (scan head passes over a stationary subject), feed (subject is pulled across a stationary scan head), drum (subject is rotated around a stationary scan head), and hand-held (user passes device over a stationary subject).

TIFF : Tag Image File Format (TIFF) is a common format for exchanging raster (bitmapped) images between application programs. Usually identified with the ".tiff" or ".tif" filename extension, the format was developed in 1986 by an industry committee chaired by the Aldus Corporation (now part of Adobe). One of the more common image formats, TIFFs are common in desktop publishing, faxing, and medical imaging applications.

Threshold : The minimum level at which a signal of any kind can be detected, either by the human senses or by using any electronic instrumentation. In image processing, threshold is a specified grey level used for producing a binary image.

7.17 REFERENCES AND FURTHER READING

- Arms, William Y. (2000). *Digital Libraries*. The MIT Press: Cambridge, MA.
- Arms, W.Y. (1995). Key Concepts in the Architecture of the Digital Library. *D-lib Magazine*.
- Haigh, S. (1996). Optical Character Recognition (OCR) as a Digitization Technology. *Network Notes*. no. 37.
- IMLS: A Framework of Guidance for Building Good Digital Collections*.
(<http://www.ims.gov/pubs/forumframework.htm>)
- Jantz, Ronald. (2001). Technological Discontinuities in the Library: Digital Projects That Illustrate New Opportunities for the Librarian and the Library. *IFLA Journal* 27, 74-77.
- Kenney, Anne R. and Stephen Chapman. (1996). *Digital Imaging for Libraries and Archives*. Ithaca: Dept. of Preservation and Conservation, Cornell University Library.
- Kessler, Jack. (1996). *Internet Digital Libraries: The International Dimension*. Boston: Artech House Publishers.
- Lesk, Michael. (1997). *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco: Morgan Kaufmann Publishers.
- NorthEast Document Conservation Center. *NEDCC Handbook for Digital Projects: A Management Tool for Preservation and Access*.
- Noerr, Peter (2000). *Digital Library Tool Kit*. U.S.A.: Sun Microsystems.
(<http://www.sun.com/products-n-solutions/edu/libraries/digitaltoolkit.html>)
- Ostrow, Stephen. Digitizing Historical Pictorial Collections for the Internet. *CLIR* February 1998. (<http://www.clir.org/pubs/reports/ostrow/pub71.html>)
- Rosenfeld, Louis and Morville, Peter (1998). *Information Architecture*. Cambridge: O'Reilly.
- Tyson, Jeff. (2003). *How Scanners Work: How Stuff Works*. (<http://www.howstuffworks.com>)
- Townsend, Sean, [et. al.]. *Digitising History*. (http://hds.essex.ac.uk/g2gp/digitising_history/index.html)

SOSS-IGNOU/P.O. 2T/Dec,2008(Reprint)



ignou
THE PEOPLE'S
UNIVERSITY