
Library Automation: The Oxford English Dictionary (Simpson & Weiner, 1989) defines automation as “application of automatic control to any branch of industry or science by extension, the use of electronic or mechanical devices to replace human labour”.

History of Library Automation: The initial work on library automation began in 1930's when punch card equipment was implemented for circulation and acquisition in libraries. During the 1930's and early 1940's progress on computer systems was slow because of depression and World War II. The library automation progressed along with the developments in computer and communication technology.

- The ENIAC (Electronic Numerical Integrator and Calculator) computer was developed by John Mauchly and J. Presper Eckert at the University of Pennsylvania.
- A major breakthrough occurred in 1947 when Bell Laboratories replaced vacuum tubes with the invention of the transistor.
- The UNIVAC I (Universal Automatic Computer) became the first computer using transistors and was used at the U.S. Bureau of the Census from 1951 until 1963.
- Invention of integrated circuit by Robert Noyce of Intel and Jack Kirby of Texas Instruments in 1960s can be considered as yet another landmark.
- Development of a new indexing technique called "keyword in context" (KWIC) by H.P. Luhn, in 1961 for articles appearing in Chemical Abstracts.
- Between 1965 and 1968, LoC began the MARC I project, followed quickly by MARC II. MARC was designed as way of "tagging" bibliographic records using 3-digit numbers to identify fields.
- The MARC II format became the basis of a standard incorporated by NISO (National Information Standards Organization) in 1974.
- ARPANET, a network established by the Defense Advanced Research Projects Agency in 1969 brought into existence the use of e-mail, telnet and ftp.
- The use of commercial systems for searching reference databases (such as DIALOG) began in 1970s. BALLOTS (Bibliographical Automation of Large Library Operations) in the late 1970's was one of the first and later became the foundation for RLIN (the Research Libraries Information Network).
- A sub-net of ARPANET made MELVYL, the University of California online public access catalogue, available on a national level in 1980.
- During 1980s, the size of computers decreased, at the same time, technology provided faster chips, additional RAM and greater storage capacity.
- The UNESCO started distributing Micro CDS / ISIS in 1980s through its distribution entre in every developed country.
- Several integrated library package started appearing in the market place. The LibSys in India was launched towards the end of 1980s.
The introduction of CD-ROMs in the late 80s changed the way libraries operate.

Connections to "outside" databases such as OCLC, DIALOG, and RLIN continued, however, in the early 90's the databases that were previously available on-line became available on CD-ROM, either in parts or in their entirety.

The Internet gave rise to yet another era in library automation. The use of networks for e-mail, ftp, telnet, Internet, and connections to on-line commercial systems grew.

The World Wide Web developed in 1993 became the fastest growing media of information delivery of all kinds.

Expert systems and knowledge systems became available in the 90s with improvement in software and hardware capabilities.

**Planning for Library Automation:** Planning library automation involves a number of key decisions with regard to purchase of integrated library software, computer hardware (server and PCs), building library databases (books and patrons), setting-up network infrastructure, manpower requirement and training, etc.

**Process of planning library automation:**

**Hardware Requirement:** The process of automating library requires hardware both for hosting the library software and databases, i.e. servers, as well as computers to access it, i.e. clients. It is important that the server is scalable so that additional storage, processing power or networking capabilities can be added, whenever required. Clients are the machines that reside on the user's end or kept in library for users to access library OPAC.

**Local Area Network (LAN):** A Local Area Network (LAN) is a required to facilitate interaction between server and clients.

**Creation of Databases of Records:** Three important databases that are required as pre-requisite to the library automation are:

- Database of Library Books
- Database of Journals
- Patron Database

**Retro-conversion of Bibliographic Records:** There are a wide range of options available to libraries for retrospective conversion of bibliographic records available in a library.

**Manpower and their Training:** Trained and skilled manpower is a crucial requirement for successful automation of library.

**Standards and Protocols:** Standards and protocols are of permanent importance in the process of implementation of library automation. Standards and protocols are instrumental in facilitating the operability, data transfer and data change.

- **AACR-II:** The first edition of AACR appeared in 1967, a cooperative effort of catalogers in the U.S., U.K., and Canada. The second edition, published in 1978 and revised in 1988, was developed by the Joint Steering Committee for Revision of AACR with Michael Gorman and Paul Winkler as editors.
Machine Readable Catalogue (MARC): The Library of Congress developed MARC in the 1960’s. “Machine readable” means that the computer can read and interpret information found in the cataloguing record.

Z39.50 or OAI-PMH: Z39.50 is an ANSI / NISO standard for information storage and retrieval. Z39.50 protocol is used for searching and retrieving bibliographic records across more than one library system. Z39.50 has been extended to allow system feedback and inter-system dialogue.

FRBR (Functional Requirements for Bibliographic Records): FRBR was developed by an IFLA Study Group (1992-1997). IFLA continues to monitor the application of FRBR and promotes its use. As Patrick LeBoeuf put it, FRBR is "a framework for commonly shared understanding".

Dublin Core: The Dublin Core refers to a set of metadata elements that can be assigned to web pages so as to facilitate discovery of electronic resources. A set of 17 core elements in Dublin Core include: Title, Creator, Subject and Keywords, Description, Publisher, Contributor, Date, Resource Type, Format, Resource Identifier, Source, Language, Relation, Coverage, Rights Management, Audience, Rights Holder. Dublin Core is being expanded with "qualifiers" for each core elements. For example, core element “creator” can further be qualified as "creator.author" or "creator.compiler" or "creator.editor" to specify that creator is an author, a compiler or an editor.

ILL Standards: The ISO Inter Library Loan Protocol (ISO ILL) was developed to provide uniform procedures when accessing a library across a network to order copy or loan material, and for carrying out the administrative tasks involved in loan management.

Open URL: Open URL is an emerging standard for transporting information within a URL to a ‘resolution server’ that can accept the URL syntax and provide context-sensitive services based on the information in the URL.

Search/Retrieve Web Service (SRW) and Search/Retrieve URL Service (SRU): SRW and SRU are intended to define a standard form for Internet search queries as well as the structure of the responses. SRW/U was developed mainly with the aim of simplifying some of the complexities involved with the Z39.50 protocol, while keeping the useful parts of the protocol, such as CQL3 query syntax.

Automation of In-house Operations: Most integrated library systems consists of five modules, namely acquisition, cataloguing, Library OPAC, circulation and serials control, that are used to automate various in-house operations of the library.

Integrated Library Software and Criteria for Selection of a Software: Some of the important library automation software packages are:
SOUL: Software for University Libraries (SOUL) is state-of-the-art integrated library management software designed and developed by the INFLIBNET Centre, based on requirements of college and university libraries. The latest version of the software i.e. SOUL 2.0 was released in January 2009

Koha: It was originally build up and developed in New Zealand by the company called as Katipo Communications Limited. The Integrated Library Software Solution was first deployed in the year January, 2000 for Horowhenua Library Trust. The name of the ILS become koha from a term called Māori which means a "gift" or "donation". The powerful Zebra indexing engine was introduced in 2005 with the release of Koha 3.0 version.

LibSys was started with Info-Tek Consultants Pvt. Ltd, a New Delhi-based software company with a mission to retain as a market leader in the field of library automation. The company is engaged in providing software solutions since the year 1984. The company is now known as LIBSYS Ltd.

Libsys Versions

- LSEase – LSEase is the library management software with basic features and modules of Libsys giving prospects a low cost and high value based proposition.
- LSAcademia - LSAcademia is a complete Enterprise resource planning (ERP), is business management software Solution to manage an Academic/ Institution Campus whether it is a school, college or an institute.
- Libsys4 – Libsys4 is one of the very popular versions of Libsys which is provided on various platforms as per the need of the client.
- Libsys7 - LIBSYS7 is a true realization of Lib 2.0, providing end to end library business functionality through its complete, inclusive seven modules. The Libsys7 provides platform independent Web based solution.
- LSDigital - LSDigital is document digitization software provided by Libsys Ltd.

Barcode, RFID, QR Code, Biometric, Smartcard: Features and Applications.

Automatic Identification: Automatic identification of documents in a library and patrons is an important component of effective automation of circulation as well as several other processes. Barcode technology and RFID are two automatic identification technology.
Barcode technology: Barcode is not a new technology, it was introduced in 1940 although it was first applied commercially in 1960’s as a method for tracking rail road cars. Since then, it has been used extensively in consumer industry, material handling industries and libraries.

Radio Frequency Identification (RFID): RFID (Radio Frequency Identification) is a term used for a radio-enabled device that communicates with or interrogates a tag or smart label, which is embedded with a single microchip processor and an antenna.

Radio-frequency identification (RFID) is part of family of AIDC (Automatic Identification and Data Capture) technologies that includes barcodes and smart cards, it is non-contact (wireless) use of radio-frequency electromagnetic fields to transfer data, for the purposes of automatic identification and tracking of tags attached to objects. The tags contain electronically stored information which can be read, added or modified. These tags doesn’t need to be in line of sight of the reader and can be embedded in variety of objects. A tag is constructed of a silicon microchip and is etched with an antenna.

Discovered in 1935 by Scottish physicist Sir Robert Alexander Watson-Watt, the first known application of RFID was done in World War II where in RFID was used to identify the planes (enemy or own) using radar.

Components of RFID
Generally RFID system consist of an object (i.e. an item which needs to be identified or tracked) with a tag, a RFID reader to read data from the tag. A server or a docking station with software that interface with a system that makes use of information from the tag.

RFID Tags mainly are of two types: active and passive.
- Active tags can be powered and read at short range via electromagnetic induction (magnetic field).
- Passive type of tags can use local power source (a battery) or can collect energy from the surrounding electromagnetic field and act as passive transponder to emit microwaves or UHF radio waves.

US based National Information Standards Organization (NISO) in its document “RFID in US Libraries” (NISO-RP-6-2012) released in March 2012 has made following recommendations for RFID tags to be used in libraries.
- In Libraries, 13.56MHz High Frequency tags should be used;
- RFID tags for library use should be passive;
- The typical read range of tags for library applications should not be increased substantially beyond the present range of 8-20 inches for smaller tags in future;
- Only tags including standardized AFI feature should be used in libraries;
- The system will cause no interference with other applications;
- The system will utilize ISO/IEC 18000-3 Mode 1 tags programmed so that they should work for identification of items in other libraries;
- The system will use tags that will not interfere with the operation of security systems in other libraries;
- Security implementations for RFID in libraries should not lock a compliant system into any one security possibility (EAS, AFI, Virtual deactivation), but rather leave security as a place for differentiation between vendors;
- RFID tags should be reprogrammable for migration purposes; and
- Data on RFID tags should be encoded according to the recommended data model, using encoding described in ISO/IEC 15962 and using relative object IDs specified in anticipated standard ISO/NP 28560 for RFID in libraries.

RFID Standards For Library
The first edition of the NCIP (NISO Circulation Interchange Protocol) standard (version 1.00) was published as ANSI/NISO Z39.83 in 2002. It was then revised in 2008 (version 2.00).

ISO 180003- Mode 1: This is an ISO standard for parameters for air interface communications at 13.56 MHz (High Frequency)
ISO 28560-1:2011: This is a set of Information and documentation — Data model for use of radiofrequency identifier (RFID) in libraries.

SIP2 Protocol is a communication protocol that provides a standard interface between Library Management System (LMS) and library automation devices.

RFID tag life is said to be 100,000 transactions or at least 10 years.

**QR code**
QR is short for "Quick Response". The QR code system was invented in 1994 by the Japanese company Denso Wave. Its purpose was to track vehicles during manufacturing; it was designed to allow high-speed component scanning.

A QR Code is a barcode that contains data that can be read by a phone's camera. These codes, once scanned by your phone, can provide you with a URL, contact information, SMS, or similar links to information right on your phone.

Unlike other barcodes which have to be sent to a database to retrieve information, QR codes are self-contained; hence the "Quick Response."

Most smartphones have a tool that allows them to download a QR Code reader. Once you download a reader, the phone's camera acts like a scanner, allowing it to "read" the barcode.

**What types of QR codes can be made?**
Directing a user to a website is only the most basic option for a QR code. Other options include:

- **Contact information**: A scanned QR code functions like a virtual business card, including your name, phone number, email, address and company details. These are automatically stored in a phone’s contacts when scanned.

- **SMS**: Scanning creates a text message with a predetermined recipient. All the scanner needs to do is hit send.

- **Email**: Scanning stores an email complete with message, subject line and recipient. All the user needs to do is hit send.

- **Phone call**: Scanning automatically loads or starts a phone call to an embedded number.

- **Text**: Scanning reveals a small amount of text stored in the QR code.

- **Location coordinates**: Scanning sends location coordinates to a geolocation app.

- **Calendar event**: Scanning automatically adds an event to the scanner’s calendar with a reminder.

- **Wi-Fi network**: Scanning stores Wi-Fi network credentials for automatic network connection and authentication.

- **Social media profile**: Scanning follows a specific profile (on Facebook, Twitter, Instagram, etc.) using the scanner’s personal profile.

- **App store**: Scanning links directly to a page on an app store, making it easy to download.
**Dynamic QR codes:** These codes are generated once, but the data stored on them can be edited. They include embedded analytics and password protection so creators can see how often they are scanned. They can even add basic logic such as device-based redirection. For instance, they can detect what device (Android vs. iPhone) is being used and redirect to the appropriate app store (Google Play vs. Apple App Store) or music library (Spotify vs. iTunes).

**Biometrics** is a technological and scientific authentication method based on biology and used in information assurance (IA). Biometric identification authenticates secure entry, data or access via human biological information such as DNA or fingerprints. Biometric systems include several linked components for effective functionality.

A biometric system can be either an 'identification' system or a 'verification' (authentication) system, which are defined below.

**Identification - One to Many:** Biometrics can be used to determine a person's identity even without his knowledge or consent. For example, scanning a crowd with a camera and using face recognition technology, one can determine matches against a known database.

**Verification - One to One:** Biometrics can also be used to verify a person's identity. For example, one can grant physical access to a secure area in a building by using finger scans or can grant access to a bank account at an ATM by using retinal scan.

Biometric authentication requires to compare a registered or enrolled biometric sample (biometric template or identifier) against a newly captured biometric sample (for example, the one captured during a login). This is a three-step process (Capture, Process, Enroll) followed by a Verification or Identification process.

Biometrics is used for security systems and replacement systems for ID cards, tokens or PINs. A key difference between biometrics and other systems is that biometric verification of physical information requires a person to be present, which adds a layer of security because other ID types can be stolen, lost or forged.

A biometric system includes the following components and features:

- A sensor that grabs data and changes it into a usable, digital format via software. This data may be from human behavioral or physical characteristics, such as a fingerprint or retinal scan. An acquisition device, such as a microphone or scanner, captures the data.
- A biometric template developed via the biometric system's signal processing algorithms. These templates are compared to the biometric system’s data storage, and data is usually encrypted for added security. A matching algorithm compares new templates with others held in the biometric system’s data storage facility.
- A decision process uses matching event results.

**A smart card,** chip card, or integrated circuit card (ICC) is a physical electronic authorization device, used to control access to a resource. It is typically a plastic credit card sized card with an embedded integrated circuit

In 1968 and 1969 Helmut Gröttrup and Jürgen Dethloff jointly filed patents for the automated chip card. Roland Moreno patented the memory card concept in 1974. An important patent for smart cards with a microprocessor and memory as used today was filed by Jürgen Dethloff in 1976 and granted as USP 4105156 in 1978
In 1977, Michel Ugon from Honeywell Bull invented the first microprocessor smart card with two chips: one microprocessor and one memory, and in 1978, he patented the self-programmable one-chip microcomputer (SPOM) that defines the necessary architecture to program the chip.

The first mass use of the cards was as a telephone card for payment in French payphones, starting in 1983.

Digitization refers to the process of translating a piece of information such as a book, journal articles, sound recordings, pictures, audio tapes or video recordings, etc. into bits. Bits are the fundamental units of information in a computer system. Converting information into these binary digits is called Digitization.

The Harvard University Library Digitization Initiative provides the following guidelines for Digitization of images and text materials:

- Determine whether page images, full text, or both need to be produced to meet project requirements;
- Access source materials and plan appropriate preparation, transfer, handling and disposition procedures;
- Create archival versions of page images and / or full text for long-term storage and production of deliverables as needed; and
- Create deliverables for distribution as page images and / or full text;

Needs of Digitization: Digitizing 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 fulltext...
search on a document that is digitized and OCR'd. It is possible to create hyperlinks to lead a reader to related items within the text itself as well as to external resources. Ultimately, Digitization does not mean replacing the traditional library collections and services; rather, it serves to enhance them.

The **process of selection** of material for Digitization involves identification, selection and prioritization of documents that are to be digitized. The IPR issues must be addressed early in the selection process.

The following four steps are involved in the **process of Digitization**.

**Scanning, Indexing, Store, Retrieve**

**Scanning**
- 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)

**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 image involves linking of database of scanned image 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.

**Store**: The most tenacious problem of a document image relates to its file size and, therefore, to its storage. The file size varies directly with scanning resolution, the size of the area being digitized 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.

**Retrieve**: 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.

There are four basic approaches that can be adapted to translate from print to digital:

1. Scanned as Image Only
2. OCR and Retaining Page Layout
3. Retaining Page Layout using Acrobat Capture; and
4. Re-keying the Data

Digital images, also called “bit-mapped page image” are “electronic photographs” composed of set of bits or pixels (picture elements) represented by “0” and “1”. The quality of digital image can be monitored at the time of capture by the following factors:

1. Bit depth / dynamic range
2. Resolution
3. Threshold
4. Image enhancement

- 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.
- 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).
- The threshold setting in bitonal scanning defines the point on a scale, usually ranging from 0 - 255, at which grey values will be interpreted as black or white pixels.
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 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.

Lossless compression is primarily used in bitonal images.

Lossy compression is used with grey-scale / colour scanning

TIFF-G4 International Telecommunication Union (ITU Group 4) is considered as de facto standard compression scheme for black & white bitonal images

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

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).

OCR (Optical Character Recognition) programs are software tools used to transform a scanned textual page images into a word processing file.

There are following types of Scanners
1. Flatbed Scanners – right angle, prism and overhead flatbed
2. Sheet-Feed Scanners
3. Drum Scanners
4. Digital Cameras
5. Slide Scanners
6. Microfilm Scanner
7. Video Frame Grabber
8. Hand-held scanners.

Digital Library: Genesis, Characteristics, Types, Architecture; Standards, Formats and Protocols, DOI.

The World Wide Web (WWW) or the Web is a collection of thousands and thousands of documents and is considered as a digital library by many people. The web is means by which most digital libraries are accessed, but it is not a digital library itself although it has several features of a digital library.

The earliest application of digital library concepts involved character-coded storage and full-text indexing of legal and scientific documents. The Legal Information through Electronics (LITE) System was first implemented by the US Air Force in 1967. DIALOG became the first commercial online service in 1972.

Some of the better known text storage and retrieval packages included: IBM’s Storage and Information Retrieval System (STAIRS), Battelle Automated Search Information System (BASIS), INQUIRE, BRS/SEARCH, DOCU/MASTER, ASSASSIN,STATUS, CAIRS, etc. Micro-CDS/ISIS, one of such advanced nonnumerical information storage and retrieval software developed by UNESCO in 1985, was used extensively by libraries especially in developing countries. Micro CDS/ISIS is currently available in different flavours including CDS/ISIS for Windows, GenISIS, JavaISIS, WEBLIS, WWW-ISIS, etc.

Availability of a wide range of Database Management System (DBMS) such as Ingres, Microsoft Access, MS-SQL Server, MS FoxPro, MySQL, NoSQL, Oracle, Postgres, SQLite and MongoDB in late 1990s and early 2000 also contributed to evolution of digital libraries.
The first digital image was produced in 1920, however, the invention of the CCD (chargecoupled device) in 1969 at AT&T Bell Labs by Willard Boyle and George E. Smith led to its application of imaging technology in consumer products like digital scanners and digital camera.

There are essentially three methods of building digital collections:
1. digitization, converting paper and other media in existing collections to digital form
2. acquisition of original digital works created by publishers and scholars. Example items would be electronic books, journals, and datasets.
3. access to external materials not held in-house by providing pointers to Web sites, other library collections, or publishers’ servers.

Some of most important standards used in the digital libraries are listed below:
• **User Interface**
  - Common web browser compatible to all platforms
• **Data Handling and Interchange**
  - Graphic Formats - JPEG, TIFF, GIF, PNG, Group 4 Fax, CGM
  - Structured Documents - HTML, XML, PDF
  - Moving Pictures/3-D - MPEG, AVI, GIF89A, QuickTime, Real Video, ViviActive, VRML  ● **Metadata**
  - Resource Description - Dublin Core, METS, MODS,
  - MARC, TEI Headers, Other Open Source and Domain Specific Standards,
  - PREMIS (Preservation Metadata: Implementation Strategies)
  - Resource Identification - URN, PURL, DOI, SICI
• **Search and Retrieval**
  - Federation and Harvesting: FTP-enabled, OAI-PMH for intermittently transfer data from one system to another
  - Federated search: Z39.50 protocol, SRW Protocol
• **Security, Authentication and Payment Services**
  - Emerging e-Commerce Standards.
**PURLS.** PURLS are persistent URLs. They are a scheme developed by OCLC in an attempt to separate a document name from its location and therefore increase the probability that it will always be found. PURLS work through a mapping of a unique, never-changing PURL to an actual URL. If a document moves, the URL is updated, but the PURL stays the same. In operation, a user requests a document through a PURL, a PURL server looks up the corresponding URL in a database, and then the URL is used to pass the document to the user. Because PURLS also confound a name with an access method, like URLs, they are not true

**Uniform Resource Name (URN).** URNs are a development of the Internet Engineering Task Force (IETF). A URN is not a naming scheme in itself, but a framework for defining identifiers (Lynch, 1998). They contain a naming authority identifier (a central authority given the task of assigning identifiers) and an object identifier (assigned by the central authority). Like PURLs, URNs must be resolved, through a database or other such system, into actual URLs. Unlike PURLs, however, a URN can be resolved into more than one URL, such as one for each of several different formats. There is currently no working URN system.

**Digital Object Identifier (DOI) System.** DOI is an initiative by the Association of American Publishers and the (American) Corporation for National Research Initiatives designed to provide a method by which digital objects can be reliably identified and accessed. The CNRI Handle system, which underlies DOI, is a system that resolves digital identifiers into the information required to locate and access a digital object. The main impetus of the DOI system is to provide publishers with a method by which the intellectual property right issues associated with their materials can be managed

**Digital Preservation - Need, Purpose, Standards, Methods, Techniques, Projects (National and International)**

Kirchhoff (2008) defines digital preservation as “series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the very long-term”.

**Fundamental needs for digital preservation include:**
- Exponential growth in digital information available in libraries and its ephemeral nature;
- Increased complexity of digital objects (incorporating text, images, audio, video, GIS, formats, etc.) and their increasing dependency on the software required to read and use them;
- Rapid flux of technology, standards and formats;
- Multiplicity of standards and formats;
- Absence of widely-accepted standards that will assure access overtime;
- Need to ensure usability, durability and intellectual integrity of the digital information; and
- Rapid changes and obsolescence of storage media (e.g., Limited life span of storage media).

- Digital contents are machine-dependent. It may not be possible to access the information unless there is appropriate hardware, and associated software.
UNESCO’s Guidelines for the Preservation of Digital Heritage (2003) group these strategies under the following four categories:

1 **Short-term Strategies**
   - Bit-stream Copying
   - Refreshing
   - Replication
   - Technology Preservation or Computer Museum
   - Backwards Compatibility and Version Migration

2 **Medium- to Long-term Strategies**
   - Migration
   - Viewers and Migration at the Point of Access
   - Emulation
   - Canonicalization

3 **Investment Strategies**
   - Restricting Range of Formats and Standards
   - Reliance on Standards
   - Data Abstraction and Structuring
   - Encapsulation
   - Software Re-engineering
   - Universal Virtual Computer

4 **Alternative strategies**
   - Analogue Backups
   - Digital Archaeology or Data Recovery

- Short-term digital preservation strategies are likely to work for a short period of time only.
- Bit-stream copying, commonly known as “backing up data” refers to the process of making an exact duplicate of a digital object.
- Refreshing essentially means copying digital information from one long-term storage medium to another of the same type, with no change whatsoever in the bit-stream (e.g. from an older CD-RW to a new CD-RW).
Replication is used to represent multiple digital preservation strategies. Bit-stream copying is a form of replication. LOCKSS (Lots of Copies Keeps Stuff Safe) is a consortia form of replication, while peer to peer data trading is an open, free-market form of replication.

Technological preservation is based on keeping and maintaining the technical environment that is used for creation of contents including operating systems, original application software, media drives, etc.

Backwards Compatibility and Version Migration strategy relies on the ability of current versions of software to interpret and present digital material created with previous versions of the same software and to save them in current format.

Medium to Long-term Preservation Strategies proposed for medium and long-term preservation are likely to work for a long period of time.

Migration is a broader and richer concept of digital preservation than “refreshing”. Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.

Viewers and Migration at the Point of Access has been proposed as an alternative to recurring and incremental migration.

Cannibalisation is a technique designed to allow determination of whether the essential characteristics of a document have remained intact through a conversion from one format to another.

Emulation uses a special type of software, called an emulator, to translate instructions from original software to execute on new platforms. The old software is said to run “in emulation” on newer platforms.

Investment preservation strategies involve investment of efforts at the time of archiving digital materials. Such strategies include: Restricting Formats and Standards, Reliance on Standards, Data Abstraction and Structuring, Encapsulation, Software Re-engineering and Universal Virtual Computer.

Restricting Formats and Standards Preservation programmes may decide to only store data in a limited range of formats and standards. This can be achieved either by only accepting material in specified formats or by converting material from other formats before storage.

Reliance on Standards preservation strategy involves the use of open, widely available and supported standards and file formats that are likely to stable for a longer period of time discarding proprietary or less-supported standards.

Data abstraction, sometimes also called normalization, involves analyzing and tagging data so that the functions, relationships and structure of specific elements can be described.

Encapsulation may be seen as a technique of grouping together digital objects and metadata necessary to describe and provide access to that object.

Software reengineering may offer a number of strategies for transforming software as technologies change, similar to transformation of data formats.

Universal Virtual Computer is a form of emulation. It requires development of a computer program independent of any existing hardware or software that could simulate the basic architecture of every computer since the beginning, including memory, a sequence of registers, and rules for how to move information among them.

Alternative strategies to digital preservation include taking analogue backup of document (print or microfilm) or recovering data from obsolete digital media.

Analogue backups combine the conversion of digital objects into analogue form with the use of durable analogue media, e.g., taking high-quality printouts or the creation of silver halide microfilm from digital images.

Digital archaeology includes methods and procedures to rescue content from damaged media or from obsolete or damaged hardware and software environments.

Combination Strategies, no single strategy is appropriate for all data types, situations, or institutions.

IPR issues are not simple in the digital preservation world, where migration copies, archival copies, derivative versions, and other states of an object exist, over a period of time.

US Copyright Law section 108 (Limitations on exclusive rights: Reproduction by libraries and archives): The Section 108 of US Copyright Law as modified by Digital Millennium Copyright Act of 1998 has a
provision that allows libraries and archives to copy, digitise and make accessible published documents in their collections.

- **US Copyright Law section 107 (Limitations on exclusive rights: Fair use):** Another exemption libraries and archives can use for their digital preservation programs is Section 107, Fair Use. Fair use is a judicially interpreted doctrine decided on a case-by-case basis.

- **Portico** is the largest community-supported digital archives in the world, committed to the preservation of digital publications such as e-journals, e-books, and other digital content.

- **LOCKSS Programme**, based at Stanford University Libraries, provides libraries and publishers with award-winning, low-cost, open source digital preservation tools to preserve and provide access to persistent and authoritative digital content.

- **CLOCKSS (Controlled LOCKSS)** is a not-for-profit joint venture between the world’s leading academic publishers and research libraries whose mission is to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of Webbased scholarly publications for the benefit of the greater global research community.

**Digital Library Initiatives – National and International**

Government institutions of national importance, national level institutions, research organization, universities, state government institutions, financial institutions, private institutions are keyplayer for initiating digital libraries in India.

Some of the important digital library initiatives and programmes initiated across the country are as follow:

**Digital Library of Books**

- **Digital Library of India** (http://www.dli.ernet.in/)
  
  **Year of Establishment:** 2003
  
  **Implemented By:** Indian Institute of Science (IISc), Bangalore
  
  **Participating Institutions:** Indian Institute of Information Technology Hyderabad; ERNET (Education and Research Network) India; Centre for Development of Advanced Computing (CDAC) Supported By: Ministry of Communications and Information Technology, Government of India, National Science Foundation, USA

- **Vigyan Prasar Digital Library** (http://www.vigyanprasar.gov.in/digilib/) Year of **Establishment:** 1989
  
  **Implemented By:** Vigyan Prasar, Noida Participating Institutions: National Council for Science & Technology Communication (NCSTC), New Delhi; NCSTC Network, Delhi; National Children's Science Congress
  
  **Supported By:** Department of Science & Technology, Government of India

- **NCERT Online Text Books** (http://www.ncert.nic.in/textbooks/testing/Index.htm)
  
  **Year of Establishment:** 1961
  
  **Implemented By:** National Council of Educational Research and Training (NCERT), New Delhi
  
  **Supported By:** Ministry of Human Resource Development, Government of India

**Digital Library of Manuscripts**

- **Kalasampada: Digital Library Resources for Indian Cultural Heritage** (http://www.ignca.nic.in/dlrich.html)
  
  **Implemented By:** Cultural Informatics Laboratory, Indira Gandhi National Centre for the Arts (IGNCA), New Delhi, India
  
  **Supported By:** Ministry of Communications and Information Technology, Government of India
- **National Databank on Indian Art and Culture (NDBIAC)** ([http://ignca.nic.in/ndb_0001.htm](http://ignca.nic.in/ndb_0001.htm))
  **Implemented By:** Indira Gandhi National Centre for the Arts (IGNCA), New Delhi, India
  **Supported By:** MCIT and Archeological Survey of India

- **National Mission for Manuscripts** ([http://www.namami.org/index.htm](http://www.namami.org/index.htm))
  **Year of Establishment:** 2003
  **Implemented By:** National Mission for Manuscript
  **Supported By:** Ministry of Tourism and Culture, Government of India

- **Muktabodha: Digital Library and Archiving Project** ([http://www.muktabodhalib.org/digital_library.htm](http://www.muktabodhalib.org/digital_library.htm))
  **Year of Establishment:** 2003
  **Implemented By:** Muktabodha Indological Research Institute, New Delhi
  **Participating Institution:** French Institute of Pondicherry (IFP), Ecole française d'ExtrêmeOrient (EFEO)
  **Supporting Agency:** SYDA Foundation, USA

**Electronic Theses & Dissertation (ETD)**

- **ShodhGanga: Indian ETD Repository** ([http://shodhganga.inflibnet.ac.in/](http://shodhganga.inflibnet.ac.in/))
  **Year of Establishment:** 2010
  **Implemented By:** INFLIBNET Centre
  **Participating Institutions:** More than 160 Universities
  **Supported By:** University Grants Commission

- **Vidyainidhi Digital Library** ([http://www.vidyanidhi.org.in/](http://www.vidyanidhi.org.in/))
  **Year of Establishment:** 2000
  **Implemented By:** Department of Library Science, University of Mysore
  **Supported By:** NISSAT, DSIR, Government of India, Ford Foundation and Microsoft India

**Digital Library of Journals - Initiatives by Scientific Society and Publishers**

- **Indian Academy of Sciences** ([www.ias.ac.in/pubs/journals/](http://www.ias.ac.in/pubs/journals/))
  **Implemented By:** Indian Academy of Sciences (IAS), Bangalore
  **Partner Institutions:** Current Science Association, Bangalore; Indian Institute of Sciences, Bangalore; SpringerLink, Germany
  **Supported By:** Ministry of Science and Technology, Government of India

- **Indian National Science Academy** ([www.insa.ac.in](http://www.insa.ac.in))
  **Implemented By:** Indian National Science Academy (INSA), New Delhi
  **Supported By:** National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research (DSIR), Government of India

- **NISCAIR Research Journals** ([http://nopr.niscair.res.in/](http://nopr.niscair.res.in/))
  **Implemented By:** NISCAIR

- **IndMed** ([IndMED@NIC: http://indmed.nic.in](http://indmed.nic.in)) ([OpenMED@NIC: http://openmed.nic.in](http://openmed.nic.in))
  **Implemented By:** MEDLARS Centre (IMC), Bibliographic Informatics Division, National Informatics Centre (NIC), New Delhi
  **Supporting Agency:** Indian Council of Medical Research (ICMR), Ministry of Health and Family Welfare, Government of India
Digital Library Project in USA

Early Digital Library Projects
Project Gutenberg (http://www.gutenberg.org/)
Project Gutenberg was launched in the year 1971

Early Experiments Digitization of Journal Articles


- **Chemistry Online Retrieval Experiment (CORE)**: CORE was a joint project by Bellcore, Cornell University, OCLC, and the American Chemical Society that ran from 1991 to 1995.

- **Association for Computing Machinery (ACM)**: In 1993, the ACM decided that its future production process would use a computer system that creates a database of journal articles, conference proceedings, magazines and newsletters, all marked up in SGML.

- **American Memory**: The Library of Congress, which is the world's biggest library, has a huge number of special collections of unique or unpublished materials.

- **Digital Library Initiative - Phase-1 (1994-1998)**: The Digital Libraries Initiative (DLI) was the result of a community-based process which began in the late 1980s with informal discussions between researchers and funding agencies in USA.
  - Six major universities participated in the first phase of DLI-1 (1994-1998)
    - University of Michigan
    - University of California, Berkeley (The Environmental Electronic Library)
    - University of California, Santa Barbara (The Alexandria Digital Library)
    - Carnegie Mellon University (The Informedia Digital Video Library)
    - University of Illinois, Urbana Champaign (Federated Repositories of Scientific Literature)
    - Stanford University (Infobus)

- **Digital Library Initiative Phase-2 (DLI-2) 1999-2004**
  - National Library of Medicine (NLM),
  - the Library of Congress (LOC),
  - the National Endowment for the Humanities (NEH),
  - the National Archives and Records Administration (NARA),
  - the Smithsonian Institution (SI), and
  - the Institute of Museum and Library Services (IMLS).
Major Digital Library Projects
- The Networked Digital Library of Theses and Dissertations (NDLTD) (http://thumper.vtls.com:6090/)
- National Science Digital Library System (NSDL) (https://nsdl.org/)
- Digital Library for Earth System Education (DLESE) (http://www.dlese.org/library/index.jsp)
- ArXiv (http://arxiv.org/)
- CiteSeer (ResearchIndex) (http://citeseerx.ist.psu.edu/index)
- Networked Computer Science Technical Reference Library (NCSTRL) (http://www.ncstlr.org/)
- OAIster (http://oaister.worldcat.org/)

Digital Library initiatives in UK

Early Digital Library Project
- **ELINAR (Electronic Library Information Online Retrieval):** The ELINAR project was started in 1993 with the financial support of De Montfort University, the British Library and IBM UK.

Major Funding Agencies in UK
- Joint Information Systems Committee (JISC) (https://www.jisc.ac.uk/)
- UK Office for Library and Information Networking (UKOLN) (http://www.ukoln.ac.uk/)

Major Digital Library Projects
- eLib: Electronic Libraries Programme (eLib) (http://www.ukoln.ac.uk/services/elib/)
- The Resource Discovery Network (RDN) (http://www.jisc.ac.uk/whatwe/programmes/x4l/rdnfe.aspx)
- Intute (http://www.intute.ac.uk/)
- JISC IE Metadata Schema Registry (http://www.ukoln.ac.uk/projects/lemsr/)
- eBank UK (http://www.ukoln.ac.uk/projects/ebank-uk/)
- Focus on Access to Institutional Resources (FAIR) (https://www.jisc.ac.uk/rd/projects/focus-on-access-to-institutional-resources-fair)
- ePrints UK (http://www.ukoln.ac.uk/projects/eprints-uk/)
- British Library Digital Library Programme
  - Archival Sound Recordings
  - The 19th Century British Newspapers
  - ETHOS (Electronic Theses Online Services)
- BOPCRIS (British Official Publications Collaborative Reader Information Service)
- Medical Journals Back Files
- Newsfilm Online
- Oxford Digital Library
- The Cambridge Digital Library (http://cudl.lib.cam.ac.uk/)

Institutional Repositories - Need, Purpose, Types and Tools; Institutional Repositories in India; ROAR, DOAR, SHERPA-ROMEO.

- Institutional repositories are a web-based archive of scholarly material designed to organize and provide seamless access to scholarly publications produced by faculty or researchers of an institution in all subject disciplines.

Benefits of Institutional repositories:
Institutional Repositories by capturing, preserving and disseminating collective intellectual capital, serves as meaningful indicators of an institution’s academic quality. An Institutional Repository concentrates the institutional product credited by an academic or other institutions researchers, making it easier to demonstrate its scientific, social and financial values.

Essential Elements of Institutional Repository

- As the digital Institutional Repository can be any collection of digital material hosted, owned or controlled and disseminate by any institution irrespective of purpose of provenance.
  - Institutionally defined
  - Scholarly contents
  - Cumulative and perpetual
  - Interoperability and Open Access
  - *Cumulative and perpetual*
  - *Interoperability and Open Access*

Digital Repository Software

Digital Repository Softwares: Open Source

There are number of software’s available for creating/developing institutional digitals repositories; the brief of the some are given below;

- **DSpace**: DSpace (http://www.dspace.org) was developed jointly by the MIT library and HP. DSpace modestly describes itself as a groundbreaking digital repository system.

- **Eprints**: Eprints (http://www.eprints.org) is the original digital repository software developed by the University of Southampton to manage an open archive. Eprints was the Open Archives Initiative (OAI) –Complaint repository software.

- **Fedora**: Fedora (Flexible Extensible Digital Object and Repository Architecture) is a digital repository system developed jointly by Cornell University Information Science and University of Virginia Library as project.

- **Greenstone**: Greenstone (http://www.greenstone.org) is software for building and distributing digital library collections. This software is produced by the New Zealand Digital Library Project at University of Waikato and developed and distributed in cooperation with UNESCO

Digital Repository Softwares: Commercial

**CONTENTdm®**- Developed at the University of Washington and distributed by OCLC. The software has tools for acquiring or creating collections; tools for storage of the content and a set of tools for display and retrieval of objects.

**Digi Tool**- This is a ‘enterprise solution for the management of digital assets in Libraries and academic environment.

**EN Compass**- It is a one module of EN Compass or suite of software for managing and accessing digital content. EN Compass has many modules for various purposes.
Hyperion- It provides organisation, storage and access to digital files by searching both associated metadata and full text of text files.

Meta Source- Meta source is a suite of tools used to manage digital collections, including, digital object storage, crawling external collections and support for Metadata schemes.

VITAL- VITAL is a institutional repository software developed by VTLS. VITAL is a set of workflow extensions, management utilities and enhanced searching capabilities build on Fedora Repository Architecture.

Institutional Repositories: Examples

Indian Institute of Science
Indian Institute of Technology Bombay
Indian Institute of Technology Delhi
Indian Institute of Technology Roorkee
Massachusetts Institute of Technology
National Aerospace Laboratories
Raman Research Institute
University of California eScholarship Repository
University of Michigan Deep Blue
University of North Texas Digital Library
University of Queensland UQ eSpace
University of Southampton ePrints

The Bethesda and Berlin statements say that for a work to be Open Access, users must be able to ‘copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship.’ These are referred to as the BBB definition by Suber.

To achieve Open Access to scholarly journal literature, BOAI (2002) recommend two complementary strategies.

Self Archiving (Green OA): First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving

Open Access Journals (Gold OA) These are journals that are freely available to scholars online for downloads and use

Open Access Tools: DOAJ, DOAR, ROAR, SHERPA-ROMEO, and SPARC

The Directory of Open Access Journals (DOAJ) DOAJ: The Directory of Open Access Journals (http://www.doaj.org/) is a website that lists Open Access journals and is maintained by Infrastructure Services for Open Access (IS4OA). Until January 2013, the DOAJ was maintained by Lund University.

OpenDOAR is an authoritative directory of academic Open Access repositories.OpenDOAR is one of the SHERPA Services including RoMEO and JULIET, run by the Centre for Research Communications (CRC).

SPARC(r), the Scholarly Publishing and Academic Resources Coalition, (URL: http:/ /www.sparc.arl.org ) is an international alliance of academic and research libraries working to create a more open system of scholarly communication. SPARC was developed by the Association of Research Libraries in 1998

Registry of Open Access Repositories (ROAR) The aim of ROAR is to promote the development of Open Access by providing timely information about the growth and status of repositories throughout the world. ROAR is hosted at the University of Southampton, UK and is made possible by funding from the JISC. ROAR is part of the EPrints.org network.
SHERPA-RoMEO  The original SHERPA partnership was formed for the SHERPA project (2002-2006) and drew from research-led universities with an active interest in establishing an example of a then-new concept - an Open Access institutional repository. (Website: http://www.sherpa.ac.uk/)

SHERPA services and the SHERPA Partnership are both based at the Centre for Research Communications at the University of Nottingham. SHERPA services include:

- RoMEO - Publisher's copyright & archiving policies
- JULIET - Research funder's archiving mandates and guidelines
- OpenDOAR worldwide Directory of Open Access Repositories
- SHERPA Search - Simple full-text search of UK repositories

RoMEO-(Publishers' Copyright and Archiving Policies) Website: (http://www.sherpa.ac.uk/romeo/)

RoMEO is part of SHERPA Services based at the University of Nottingham. RoMEO is a searchable database of publisher’s policies regarding the self-archiving of journal articles on the web and in Open Access repositories.

<table>
<thead>
<tr>
<th>ROMEO colour</th>
<th>Archiving policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>can archive pre-print and post-print or publisher's version/PDF</td>
</tr>
<tr>
<td>blue</td>
<td>can archive post-print (ie final draft post-refereeing) or publisher's version/PDF</td>
</tr>
<tr>
<td>yellow</td>
<td>can archive pre-print (ie pre-refereeing)</td>
</tr>
<tr>
<td>white</td>
<td>archiving not formally supported</td>
</tr>
</tbody>
</table>

**Content Management Systems – Architecture, Data Integration, CMS Software – Selection, Implementation and Evaluation.**

Content management can be defined as a process of creating, collecting, organizing, categorizing and structuring information resources of any type or format so that they can be saved, retrieved, published, updated and re-purposed in any way desirable (Yu, Holly, 2004)

Content Management System (CMS) contains the terms content and management (administration) that imprecisely refer only to a system that administers content.

A CMS is a software system used for content management. It offers a way to manage large amount of web-based information that escapes the burden of coding all the information into each page in HTML by hand.

The “big four” of content management might be identified as:

**Web content management (WCM)**

The management of content primarily intended for mass delivery via a website. WCM excels at separating content from presentation and publishing to multiple channels.

**Enterprise content management (ECM)**

The management of general business content, not necessarily intended for mass delivery or consumption (e.g., employee resumes, incident reports, memos, etc.). This flavor was more traditionally known as “document management,” but the label has been generalized over the years. ECM excels in collaboration, access control, and file management.
Digital asset management (DAM)
The management and manipulation of rich digital assets such as images, audio, and video for usage in other media. DAM excels at metadata and renditioning.

Records management (RM)
The management of transactional information and other records that are created as a byproduct of business operations (e.g., sales records, access records, contracts, etc.). RM excels at retention and access control.

Famous Content Management Softwares

Drupal “Drupal is software that allows an individual or a community of users to easily publish, manage and organize a great variety of content on a website. Tens of thousands of people and organizations have used Drupal to set up scores of different kinds of web sites. Drupal is open source software licensed under the GPL, and is maintained and developed by a community of thousands of users and developers”

Joomla is one of the most popular and commonly used CMS today, the software is released under GNU General public license, so anybody can use the software and make modifications to cater for their own use

WordPress is a free and open-source content management system based on PHP & MySQL. Features include a plugin architecture and a template system. WordPress started in 2003 when Mike Little and Matt Mullenweg created a fork of b2/cafelog. The need for an elegant, well-architected personal publishing system was clear even then. Today, WordPress is built on PHP and MySQL, and licensed under the GPLv2. It is also the platform of choice for over 33% of all sites across the web.

Application of Artificial Intelligence, Expert Systems and Robotics in Libraries; Social Mobile Analytics Cloud (SMAC); Cloud Computing.

Artificial Intelligence (AI) encompasses the following general areas of research: (1) automatic programming, (2) computer vision, (3) expert systems, (4) intelligent computer-assisted instruction, (5) natural language processing, (6) planning and decision support, (7) robotics, and (8) speech recognition

The study of AI itself serves as means of understanding human intelligence. The reasons stated are as follows:
1. The use of computers demands a clear statement of the problem and clear strategy for solutions and this requires a clear understanding about the human thinking process.
2. Computer models force precision. Implementing a theory uncovers conceptual mistakes and oversights that ordinarily escape the notice of even the most meticulous researcher.
3. Computer implementations qualify task requirements. Once a program performs a task, upper bound statements can be made about how much information processing the task requires.
4. It is usually simple to deprive a computer program of some piece of knowledge in order to test how important that information really is

In general, expert systems are composed of basic components such as:
i) a user interface: To facilitate user interaction
ii) a knowledge base: The facts or knowledge based upon which the ES makes decisions
iii) an inference mechanism: the reasoning engines built according to heuristics reasoning or facts.

Decision support systems (DSS) are human–computer decision-making systems to support managerial judgments, and intuitions to solve managerial problems by providing necessary information, generating, evaluating and suggesting decision alternatives. Most organisational problems need a combination of quantitative and qualitative data processing. Decision support systems (DSS) are a subset of computer-based
information systems (CBIS). DSS maybe part Management support systems or also expert systems and executive information systems.

**Inference Engine:** A component of expert systems that makes decisions based upon reasoning built into the systems through factual or ruled based methods.

The International Data Corporation (IDC) refers to SMAC as "the third platform." The first platform was the mainframe, which began in the late 1950s and continues today. The second platform was the client/server model, a concept central to the role of networking where one program requests a service or resource from another program. The third platform is SMAC, a combination of "technology enablers that allow businesses to accelerate their digital transformation."

The term Cloud refers to a Network or Internet. In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over network, i.e., on public networks or on private networks, i.e., WAN, LAN or VPN.

Applications such as e-mail, web conferencing, customer relationship management (CRM), all run in cloud.

Cloud Computing refers to manipulating, configuring, and accessing the applications online. It offers online data storage, infrastructure and application.

Cloud can have any of the four types of access: Public, Private, Hybrid and Community.

**PUBLIC CLOUD** The Public Cloud allows systems and services to be easily accessible to the general public. Public cloud may be less secure because of its openness, e.g., e-mail.

**PRIVATE CLOUD** The Private Cloud allows systems and services to be accessible within an organization. It offers increased security because of its private nature.

**COMMUNITY CLOUD** The Community Cloud allows systems and services to be accessible by group of organizations.

**HYBRID CLOUD** The Hybrid Cloud is mixture of public and private cloud. However, the critical activities are performed using private cloud while the non-critical activities are performed using public cloud.

Service Models are the reference models on which the Cloud Computing is based. These can be categorized into three basic service models as listed below:
1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

INFRASTRUCTURE AS A SERVICE (IAAS)
IaaS provides access to fundamental resources such as physical machines, virtual machines, virtual storage, etc.

PLATFORM AS A SERVICE (PAAS)
PaaS provides the runtime environment for applications, development & deployment tools, etc.

SOFTWARE AS A SERVICE (SAAS)
SaaS model allows to use software applications as a service to end users.

Ontology – Tools (RDF, RDFS, Protege); Semantic Web, Linked Data, Big Data, Data Mining, Data Harvesting.

Architecture of Semantic Web

In Semantic Web (SW), Ontology is an important layer architecture that is develop with the help of web ontology language (OWL) and makes the web semantic. OWL is designs and creates metadata that required for developing SW applications. It is an essential part of SW

Resource Description Framework (RDF)
It is a language for representing information about resources in the web. RDF identifies things by using URIs. It uses simple statements (Triples) to describe things. It is a domain dependent technology providing a way to build an object model from which actual data is referred. Development of RDF started with the initiation of
PICS (Platform for Internet Content Selection) project in 1995. The extension of PICS project was PICSNG (PICS Next Generation), which was later called as RDF (Resource Description Framework).

Protégé was developed by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. It is a free, open-source ontology editor and framework for building intelligent systems.

RDF describes resources with classes, properties, and values. In addition, RDF also need a way to define application-specific classes and properties. Application-specific classes and properties must be defined using extensions to RDF.

RDF Schema (RDFS)
RDF Schema does not provide actual application-specific classes and properties. Instead RDF Schema provides the framework to describe application-specific classes and properties.

Classes in RDF Schema is much like classes in object oriented programming languages. This allows resources to be defined as instances of classes, and subclasses of classes.

Linked data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the internet to become a global database. Tim Berners-Lee, director of the World Wide Web Consortium (W3C), coined the term in a 2006 design note about the Semantic Web project.

Big Data is a phrase used to mean a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques. In most enterprise scenarios the volume of data is too big or it moves too fast or it exceeds current processing capacity.

Characteristics Of Big Data
(i) **Volume** – The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data.

(ii) **Variety** – Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.

(iii) **Velocity** – The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.

(iv) **Variability** – This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Data mining is the process of automatic extraction of interesting (non trivial, implicit, previously unknown and potentially useful) information or patterns from the data in large databases.

Data harvesting is similar process to data mining or data extraction. It is a process to scrape information form website using an automated bot.