
Chapter 33

Best Practices in Digitisation: Planning and Workflow Processes

Shekar Bandi¹, Mallikarjun Angadi² and J. Shivarama³

*¹Professional Assistant; ²Deputy Librarian; ³Assistant Professor,
Tata Institute of Social Sciences, Mumbai, Maharashtra, India*

¹E-mail: shekar.bandit@tiss.edu

²E-mail: mallikarjun@tiss.edu

³E-mail: j.shivarama@tiss.edu

The study has given an introduction to the fundamental principles involved in the digitization process and outlined some key concepts such as, defining digitization, examining pathways, introducing the notion of 'fit for purpose', and assessing archival concerns and dissemination compression techniques. The paper has also discussed characteristics as well as on some issues and preservation challenges. Although there are drastic changes in digital technology, finance, staff training, manpower, infrastructure etc are serious problems to be tackled before attempt for digitization.

Keywords: *Digitization, Digitization process, Digital processing, Digital preservation, Digitization issues and Preservation challenges.*

Introduction

The information contained in traditional print materials like books, journals, reports, published works, minutes of the important meetings, manuscripts, cannot be preserved forever for a number of reasons. As years pass by, the information contained in them gets faded out; the medium becomes brittle and finally becomes unusable. Unless we have alternatives arrangement for recapturing and reproducing it another format important will be lost forever. Fortunately, technological advances have provided us with suitable alternatives for preserving such valuable information;

information technology has brought tremendous changes in the way of the life of the human being. Publication industry is no exception to it. To have wider access and long-term preservation policy of scholarly human knowledge many of the professional organisations and publication houses are moving towards the electronic publication of their print resources.

Koganuramath, Muttayya M. and Angad, Mallikarjun. (2010) discussed capabilities of digital technology digitization its importance and various steps involved in the digitization process and efforts to preserve, manage, and provide access to scholarly information digitization prerequisites and also discusses the practical experience of digitization of two major projects carried out by TISS library. First one is that of digitization of 55 volumes (1952-2006) Sociological Bulletin and Indian Journal of Social Work 67 volumes (1940-2006).

Preservation Committee of the Canadian Council of Archives (2002) discussed technological approaches digitization encourages preservation by limiting the handling of original records, access strategy and impact of a digitization program on the institution's other public service activities. The committee also discussed costs and complexities inherent in the development of a digitization program, institutions should try to share resources (financial, material, human) and also highlights collaborate with others, where possible.

IFLA (2014), Report has been discussed the design of the new digital collection will be determined by the goals of the institution, its functions, and intended users. As digital collections and projects grow over time, it is useful to contemplate the future development and interaction with other collections from the same or other institutions, and also discussed workflow for creating the digital collection and Metadata, preservation of the digital collection and also recommends some useful recommendations regarding digitization and preservation and checklists.

Digitization

Digitization is a process to capture an analog signal into digital form. This paper, the term 'digitization' is a shorthand phrase that describes the process of making an electronic version of a 'real world' object or event, enabling the object to be stored, displayed and manipulated on a computer, and disseminated over networks and/or the World Wide Web. Image may be captured using a scanner or a digital camera and to optimize the clarity, OCR software may be employed to the electronic image. The numerical system used by computers is called binary and is made up of a series of ones and zeros. These ones and zeros are commonly referred to as 'bits' of information. A fundamental point to note from any digitization process is that the binary or digital channels are relatively narrow, and only a partial representation of an analogue object can ever be rendered in digital form. In other words, the digital object can ever only be a version of the real thing. The digitizer therefore has to make informed decisions about what level of detail is required in the digital version of an object, for that digital version to serve its intended purpose.

Aim of the digitization is to enhance access and improve preservation. By digitizing their collections, such as libraries can make information accessible that

was previously only available to a select group of researchers. Digital projects allow users to search collections rapidly and comprehensively from anywhere at any time.

Benefits of Digitization

- ☆ **Access:** Digitized content provides the advantage of search over print media.
- ☆ **Preservation:** Digital information can be copied and does not depend on having a permanent object and keeping under guard, but on the ability to make multiple copies, presuming at least one of them survives
- ☆ **Reduced costs of Handling:** Digitization reduces the costs of handling, storing and duplicating paper documents and in some case can reproduce the lost documents.
- ☆ **Organization and dissemination:** Digital or electronic images can be indexed and stored in a document retrieval system

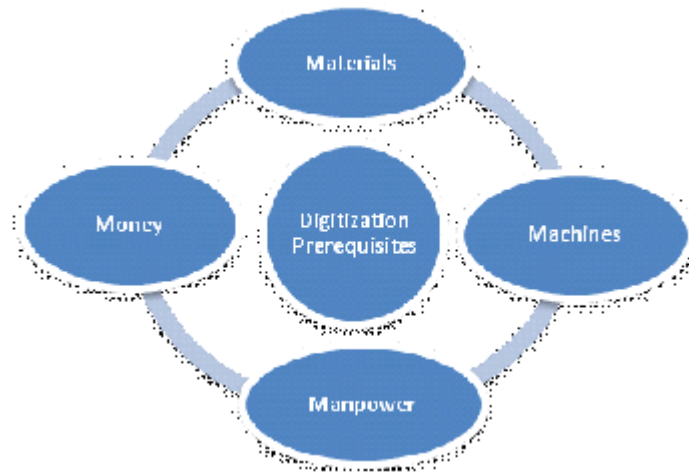


Figure 33.1: Pre-requisites for Digitisation.

Digitization is not an a easy task to done its very short time its needs there are 4M's prerequisites Materials/Content selection of the content is one of the tuff job in this process because of content is king in digital era. Machines are includes Hardware *i.e.* system, scanner and storage devices, and Software *i.e.* OCR software (Abby Fine Reader). Skilled Manpower is most important to complete any projects on certain boundaries like mater of cost and time. Money is always playing in major role in this process because of other 3M's are depending on this.

Before start the scanning its needs to know the which type document you going to scan and what is the conditions of the document and how much resolution is required minimum 200 dpi is required for the OCR (300 dpi is recommended), which mode of scanning is required *i.e.* Color/Grayscale/Block-and-white, brightness of the scanned Image, page size etc.

Digitization Workflow

Shifting from the printed product to the digital library involves progressing through many processing stages, large and small. Only with efficient processes and

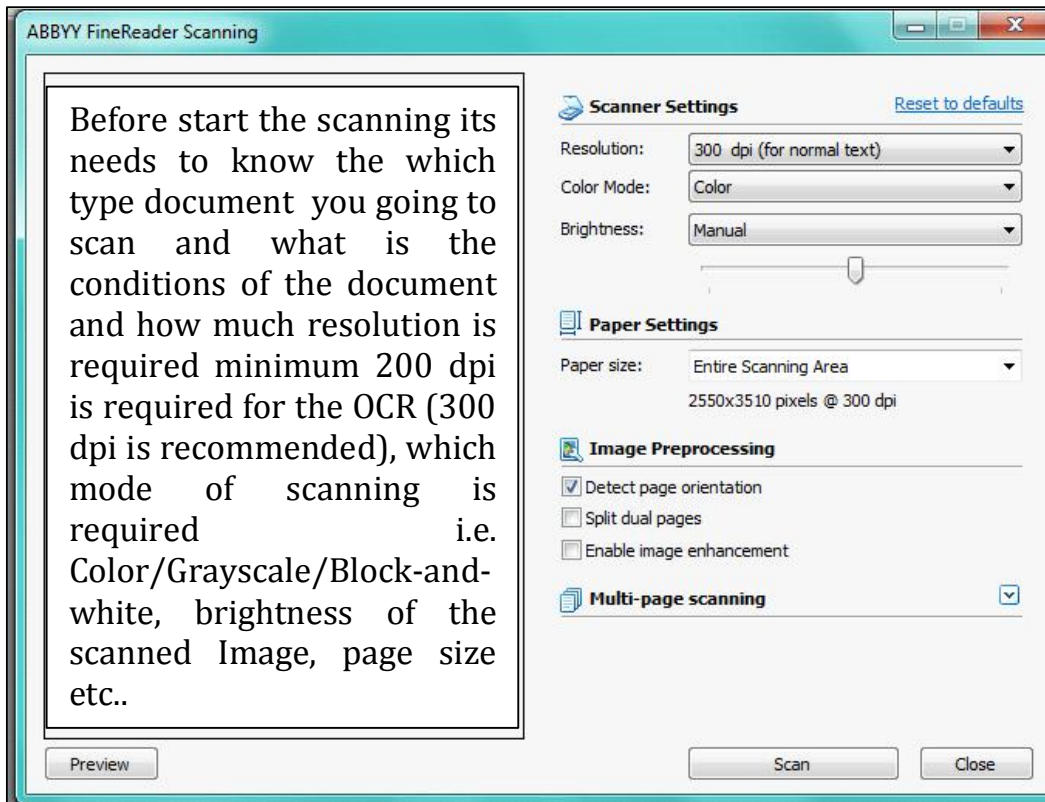


Figure 33.2: Resolution Settings in Abby Finereader.

a well thought-through work scheme can the (usually) considerable amount of documents and data be controlled.

The challenge is best faced with the assistance of what are called workflow systems. Assigning and systematizing file names is another aspect of workflow. The following diagram how the digitization work has been carried out and how to make it simple;

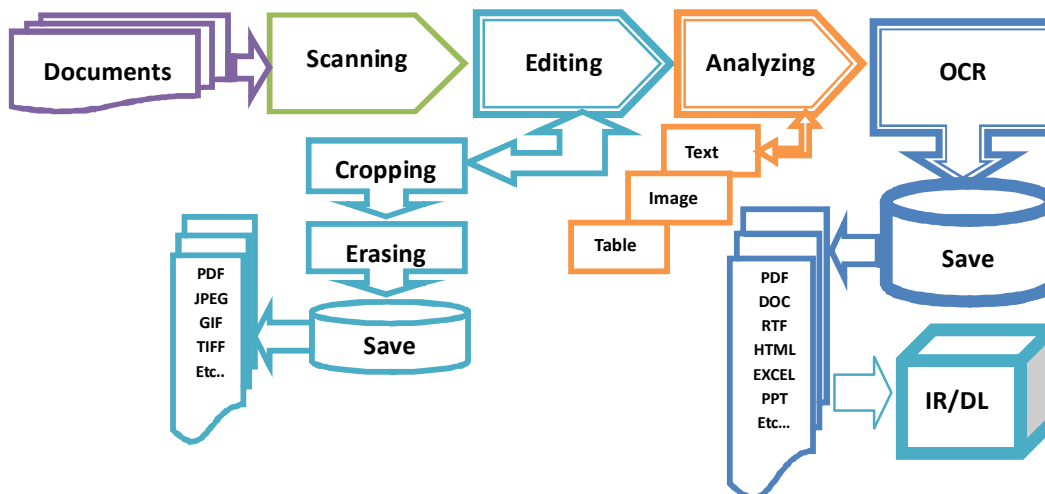
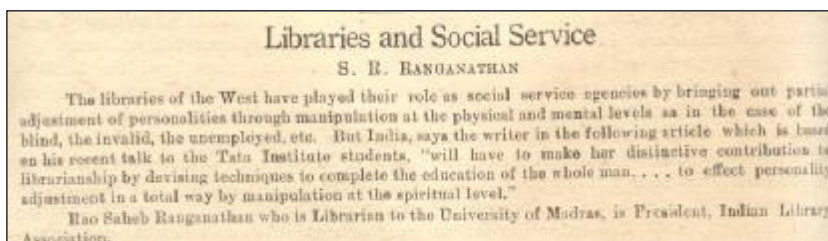


Figure 33.3: Digitiation Workflow Processes.

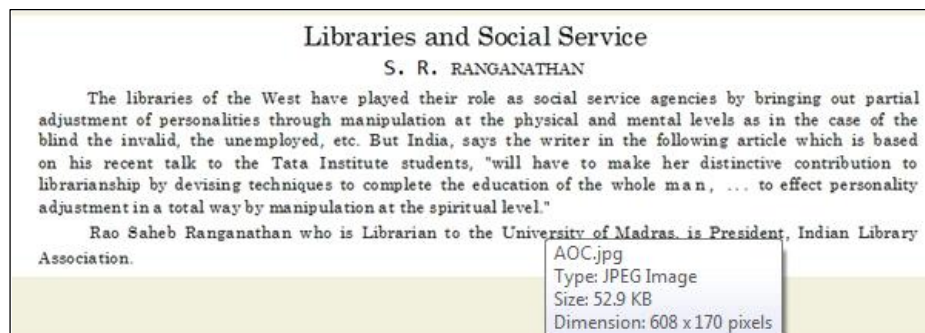
- ☆ **Documents:** The selection of documents is an essential task in the development of a digital collection. Collection, works, editions, and copies are to be studied and checked against the scope of the new digital collection.
- ☆ **Scanning:** Before going to scanning its needs to answer the questions *i.e.* what kind of documents you want digitize, what type of hardware (scanner) and software (OCR software) are required.
- ☆ **Editing:** Once your scanning job is over it menace you done only 50 per cent of work rest of the 50 per cent job in editing in this process crop the unwanted area and erase the unwanted dots and line using erase tool and after editing if you want you can save the objects in PDF (image-pdf not text-pdf), JPEG, GIF, TIFF etc.
- ☆ **Analyzing:** Once editing job is over if you want OCR your document its needs to analyze the document have only text or includes image and tables. There are OCR software supports to auto analyzing or if you want you can manually also do.
- ☆ **OCR:** it's playing vital role in the digitization, OCR helps to reduce the size of the document and helps to word and phrase searching without OCRed documents are not supports to search in any word and also size of the document is more than double. After OCR you can save the documents as per your requirements (PDF format is recommended for Preservation).
- ☆ **IR/DL:** Once digitization process has been successfully completed final and per-most step is preservation and access add required metadata on each items and upload the IR/DL.

Examples of Before OCR and After OCR

Before –OCR



After- OCR



Digitization Issues and Preservation Challenges

Even though libraries and librarians all over the world are marching towards digitization, there exist some constraints in the process and their maintenance. The problems facing digitization are

Data Size

Many of the storage media praised by people all over the world may become less useful only long after they become unreadable. Thus documents digitized and stored in such media become outdated and their maintenance will be more difficult than print media. The changes and improvements of storage medium put serious questions about the future of digitized materials and their alteration.

Documents Type

In an age of information explosion and information pollution, librarians are in a dilemma about 'what type of records is to be digitized' and 'what type of records not to be digitized'. The documents in high demand today may become obsolete even tomorrow because of the vast developments in the subject and printing and publishing industry. A digitized document deselected from the collection is lost forever. To overcome the problem, librarians should seek the advice of subject experts in each field and users of the library about the importance of each and every record and from this list selection of records for digitization can be done.

Multilingual Text Support

Digital library system should prove support to multilingual content to various functions of library in order to facilitate activities such as acquisition, storage, organization, and access to the digital collection.

Technology Obsolescence

The technology behind digitization is undergoing drastic changes continuously. The computer hardware, software, storage media and content formats are undergoing great revolution. The digitized materials become unreadable if the background devices become obsolete as time passes by which ultimately results in the loss of data. Like print media, digital media is also affected by light, heat, moisture, insects, acid content and air pollution. Digital storage media are always under the threat of above factors. While selecting the storage medium, technological obsolescence should be taken into consideration.

Copyrights

The issues regarding copyright raise serious matters before librarians in digitization. Research scholars usually include graphs, data from books and journals without prior permission of the author. In a digital library users are always demanding back issues of journals and rare historical archives for which the library has no copyright. This may lead to serious dissatisfaction about digitization among users. As a final solution to this matter, librarians must be given permission to digitize copyright works in connection with digitization.

Conclusion

The study has given an introduction to the fundamental principles involved in the digitization process. It has outlined some key concepts and themes, such as, defining digitization, examining pathways, introducing the notion of 'fit for purpose', and assessing archival concerns and dissemination compression techniques. The paper has also discussed characteristics as well as on some issues and preservation challenges. Although there are drastic changes in digital technology, finance, staff training, manpower, infrastructure etc are serious problems to be tackled before attempt for digitization. In a country like India having great history in traditional medicine, ancient art, culture, architecture, etc, the information that our great ancestors gave us through inscriptions, archives, and through rare books is to be digitized for our future generation.

References

- Koganuramath, Muttayya M. and Angad, Mallikarjun. Digitisation in an Academic Library: A Success Story at Tata Institute of Social Sciences, *DESIDOC Journal of Library and Information Technology*, Vol. 30, No. 1, January 2010, pp. 38-43.
- Eadie, Mick. (2005). The Digitisation Process: an introduction to some key themes. Retrieved from http://www.ahds.ac.uk/__print__/creating/information-papers/digitisation-process/index.htm
- Detusche Forschungspemeinschaft. (2013). DFG Practical Guidelines on Digitization. Retrieved from http://www.dfg.de/formulare/12_151/
- Watanabe, Eiji., Ozeki, Takashi. and Koham, Takeshi. Digitization of Hand-Written Notes Using a Wearable Camera, *International Journal of Information and Education Technology*, Vol. 5, No. 10, October 2015.
- Yvonne, Inden. and Nicole, Graf. Digitization – Best Practices. Retrieved from http://www.digitalisierung.ethz.ch/index_e.html
- Illinois State Library's Digital Resource Guide. Retrieved from http://www.cyberdriveillinois.com/library/digital/di_res.htm
- OY Rieger. (2008). Preservation in the Age of Large-Scale Digitization A White Paper, Council on Library and Information Resources Washington, D.C.
- Declaration of Principles Concerning the Relationship of Digitization to Preservation of Archival Records. Retrieved from http://www.cdncouncilarchives.ca/digitization_en.pdf
- FLA Rare Book and Manuscripts Section, (2014), Guidelines for Planning the Digitization of Rare Book and Manuscript Collections. Retrieved from http://www.ifla.org/files/assets/rare-books-and-manuscripts/rbms-guidelines/ifla_guidelines_for_planning_the_digitization_of_rare_book_and_manuscripts_collections_september_2014.pdf

American Library Association, (2013). Association for Library Collections and Technical Services, Preservation and Reformatting Section. *Minimum Digitization Capture Recommendations*. Retrieved from

<http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>

University of British Columbia. (2012). Digital preservation remains a pressing concern, as evidenced by this UNESCO declaration: UNESCO, *UNESCO/UBC Vancouver Declaration: The Memory of the World in the Digital Age: Digitization and Preservation*. Retrieved from

http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_en.pdf

Sabbagh, Karim., El-Darwiche, Bahjat., Roman Friedrich and Singh, Milind. (2012). Digitization: ICT's next evolution. Retrieved from

http://www.strategyand.pwc.com/media/file/Strategyand_Maximizing-the-Impact-of-Digitization.pdf