Special Paper

Digitisation: Preservation and Challenges

N.S. Mani

National Archives of India, Janpath, New Delhi-110 001
E-mail: archives@nic.in; nsmani53@hotmail.com

ABSTRACT

Digitisation involves conversion from analog to digital. With the boom in the information technologies and also with the changes in the retrieval of information need arises to convert the heritage into digital form. This paper discusses Digitisation, its need and how to preserve the digitised documents.

Keywords: Digitisation, preservation, conversion, resolution

1. INTRODUCTION

Digitisation is the process of converting an analog signal into a digital signal. Digitisation converts an image into a series of picture elements or pixels (little squares that are either black or white), a specific shade of gray (grayscale) or colour. Each pixel is represented by a single or series of binary digits, either 1s or 0s. Many computer programs like drawing, illustration, 3D model, computer-aided design/computer-aided manufacturing, and architectural design use vectors—arrows of direction, points, and lines that define shapes, as compared to the individual picture elements used to represent a raster image. One important part of digitising is the type of imaging sensor used. These image sensors/detectors can be compared to the film used in photography. The common digital imaging sensors are:

(i) Charged Coupled Devices, or CCDs, used in flatbed scanners and digital cameras.

(ii) Photomultiplier tubes, or PMTs, used in drum scanners.

(iii) Complementary metal oxide semiconductors, or CMOS chips, used in low-end flatbed scanners and low-end digital cameras.

The most common sensor used in all types of scanners and digital cameras is CCD. The photo multiplier tube, or PMT, is used only in drum scanners for the graphic arts or prepress market, i.e., printing and publishing. CMOS chips do not produce the same image quality as compared to CCDs and cannot match the resolution of CCDs. Now-a-days, CMOS chips are used only in low-end digital cameras and scanners. In scanners and digital cameras, a lens or set of lenses are used to focus an image onto the sensor—a CCD, PMT, or CMOS chip. Sometimes people refer to digital imaging as lens-less. This is not true. Without a lens to focus the light, digital images would be blurry, just as in photography. All scanners have lenses. In most cases, the design of the scanner hides the lens (es) within the body of the scanner. Scanners also have built-in light sources to illuminate the items being scanned. The light is either reflected (as with documents or photographic prints) or transmitted (as with microfilm, photographic negatives, or colour transparencies) by the item being scanned, and the image is focused by the lens (es) onto the imaging sensor. In the case of CCDs, the light falls onto the little light sensors or photo diodes on the CCD. These sites generate an electrical current, or voltage.

2. ANALOG-TO-DIGITAL CONVERSION

Once a digital image has been created and stored in any media, there is a corresponding digital to analog conversion that allows the computer to present the image in a human readable form on either a display or printer. Displaying an image on a computer monitor or printing the image on a printer needs analog representation of a digital image.

2.1 Basic Image Measures

There are three important measures of every static digital image:

Resolution: The number of dots, or pixels (picture element), used to represent an image. This is always given as a measure of linear or area density (e.g., 300 dots/inch). Resolution, or spatial frequency, is the number of times an image is sampled during the scanning process. Resolution—the number of pixels in an image can be described in a number of ways like dots per inch (dpi) or pixels per inch (ppi).

Pixel Bit Depth: This measure defines the number of shades that can actually be represented by the amount of information saved for each pixel. These can range from 1 bit/pixel for binary (fax type) images to 24 bits per pixel in high quality colour images.

Colour: There are many ways to represent, compress, and distribute color images. Suffice it to say that the smaller the image files size, the less accurately it renders the original image.

The key to a successful digital conversion program is the relationships among three concepts:

(i)   The purpose digitisation is going to serve.

(ii) Original document's condition and status.

(iii) Technology tolerance for the analog-to-digital conversion.

Information in digital form is more fragile. The recent survey of the Web shows that the overall proportion of Internet content is commercial and only six percent is educational or informational. Digital images have become commonplace in libraries and archives. With serious effort to ensure long-term access to today's digital image files, organisations like library, archives, historical society, or museum are earmarking a certain percentage of their budget for digital preservation without which the risk of loss is tremendous.

3. PRESERVATION

Digital preservation refers to the various methods of keeping digital material alive for the future. Digital preservation typically centres on the choice of interim storage media, the life expectancy of a digital imaging system, and the expectation to migrate the digital files to future systems while maintaining both. Digital imaging technologies create an entirely new form of information from traditional documents. It is not simply another reformatting option in the preservation tool kit. Digital imaging involves transforming the very concept of format, not simply creating a faithful reproduction of a book, document, photograph, or map on a different medium. The power of digital enhancement, the possibilities for structured indexes, and the mathematics of compression and communication together fundamentally alter the concept of preservation in the digital world. These transformations, along with the new possibilities they place on information professionals, force us to transform library and archival services and programs in turn.

Preservation is not just only for paper/palm leaves. In the digital imaging environment it is indeed a matter of challenge to define what preservation means? The digital world poses significant challenges to, but does not eliminate the need for responsible, effective preservation activity. Digitisation offers many advantages to organisations with large collection of documents/maps/microforms/etc., Information from these can be accessed by users directly without the human intervention. Information content in digital form can be retrieved by users remotely.

Digital image quality is improving day by day with more advancement in research on this field. The brittle, old, and fragile originals need not be consulted again and again for research and the digital form of these can be accessed by many at a single point of time for research. The power of full-text searching and sophisticated, cross-collection indexing affords users the opportunity to make new uses of digitised resources.

Digitisation of documents/books or cultural resources requires a deep and longstanding Organisational commitment in terms of investment in equipment and software as compared to the traditional preservation. Wherever large scale Digitisation programme are being undertaken it should be a prerequisite to decide definitions and standards for digital preservation and policy for migration.

In the digital world, preservation is the creation of digital products worth maintaining over time. The persistence of digital products requires careful attention to the maintenance of content (the bits and bytes) functionality (how the bits work in a system). Preservation in the digital world is not absolute, but depends on the continuing transformative impact of the digital product on the information work of end-users. Acquiring an imaging system to enhance access to library and archives materials is as simple as choosing the combination of off-the-shelf scanners, computers, and monitors that meets immediate functional specifications.

With time, most of these small-scale, pilot projects will fade away quietly and the initial investment will be lost. The costs of maintaining these systems become apparent, as vendors go out of business, and as patrons become more accustomed to remote-access image databases and the latest developments in the field. This commitment is a continuing one, decisions about preservation cannot be deferred in the hope that technological solutions will emerge like a medieval knight

in shining armor. An appraisal of the present value of a book, a manuscript collection, or a series of photographs in its original format is the necessary point of departure for making a judgment about preservation of the digital image version.

Before initiating a digitisation programme, the following should be kept in mind

✦ The condition and status of the document being converted.

✦ The capabilities of the present technology for digital conversion, and

✦ The method by which the digital images will be utilised.

The common application of Digitisation in an archive or library is digital copies that can be used for ready reference in lieu of consulting the original documents. For Digitisation the use of an intermediate film or paper copy to facilitate the scanning process may or may not be necessary or advisable. Ultimately, the purpose of digital image products is determined by the uses to which they will be put, while preservation of original source documents must be determined by their specific preservation needs.

A major challenge in choosing methods from analog to digital conversion is obtaining an in-depth understanding of the source material and the techniques being used for the same. The following should be taken in to account before undertaking the job:

✦ Format of the source (including size of object, its structure, and its physical condition).

✦ Physical condition and its impact on the ability of the item to be handled during the conversion process.

✦ Level of detail (including the colour of the paper ink used etc.).

All conversion tools have limitations in terms of the size of source documents they can handle with a given level of digital resolution. "You get what you pay for" typically applies in the selection of scanning equipment. The principles of preservation of digital documents are the same as those of the analog world, i.e., longevity, choice, quality, integrity, and accessibility. When we talk of preservation, archivists not only insist for alkaline paper standards, archival quality microfilm, mass deacidification, and more rugged magnetic media, but also issues surrounding the life expectancy of storage media, the very concept of permanence. The commitment

to quality standards, do it once, do it right, permeates all preservation activity, including library binding standards, archival microfilm creation guidelines, conservation treatment procedures, the choice of supplies and materials, and a low tolerance for error. The evolution of preservation microfilming as a central strategy for the bulk of brittle library materials has placed the quality of the medium and the quality of the visual image on an equal plane. In the pursuit of quality microfilm, compromise on visual truth and archival stability is dictated largely by the characteristics of the item/material chosen for the preservation.

The control of digital quality standards is possible now, just as it is for microfilm. In the digital world, preservation is the action and access is the thing—the act of preserving access. A more accurate construct simply states "preserve accessibility". Librarians and archivists can influence vendors and manufacturers to provide new equipment that is backward compatible with existing systems. This capability assists image file system migration in the same way that today's word processing software allows access to documents created with earlier versions. Procedures manuals (or instruction sheets) for selection, handling, scanning, metadata creation, and quality control flowcharts or other workflow diagrams, data element list the contents.

4. CONCLUSION

Since libraries and archives are actively integrating digital technologies into acquisition, cataloging, systems, and even preservation departments, perhaps it is legitimate to consider these activities as costs that lie outside of a conversion project. Copyright does not apply to all works, and it does not last forever. It is intended as a limited monopoly permitting authors to profit from their creative efforts and, eventually, for the public to be able to freely use creative works with the idea that such fluid use will inspire new creation benefiting society at large. When a creative work is not subject to copyright protection, it is referred to as being in the public domain. If a use is a fair use, then one may copy an otherwise protected work without permission from the copyright owner and without the use being deemed an infringement. Other countries have a similar concept often called "fair dealing," but this is generally more restrictive and laden with administrative requirements.

In view of the multifocussed challenges being posed by Digitisation, the Government of India has very recently formed a forum–National Digitisation Preservation Programme [NDPP] to formulate different programmes and polices for the same.