

# Securing the Future of Information: Digitisation and Preservation of Documents in e-Format

Alka Bansal, Vinod Kumari, Ashok Kumar, and  
Dr Mohinder Singh

## Abstract

Computers waved a (magic) wand and the whole process creation, access, delivery and preservation of information changed for the literary, scholarly, artistic, scientific and domestic community. Computers transformed the analogue world in to the digital world of 0 and 1. Digitisation came as a most wonderful and welcome tool in hands of libraries, museums, archives, societies, publishers, and others for preserving billions of paper/analogue documents in digital format for retaining their original look as well as with a point of view of their future relevance and potential. Preservation is the activity of protecting something from loss or danger. Digital preservation includes the preservation of print and non print material in digitised form for effective, efficient and purposeful use.

Any library can't digitise all its resources. What should it do? What should it preserve? Why should and for whom should it preserve? How this ever changing digitisation scenario is putting a pressure on librarians/archivists, etc. This paper is making an attempt to seek answers to these questions. This paper also explores the concept of preservation by digitisation, the need of digital preservation, and the problems of digital preservation. How digital preservation activities are being carried out by DESIDOC for its publications: *Defence Science Journal*, *DESIDOC Bulletin of Information Technology*, *PST Series*, etc., are described. A detailed case study of development of database of negatives of photographs taken at various events, being developed by DESIDOC by use of various software like MS Access, Corel Draw, PhotoShop, etc., is discussed.

## 1. INTRODUCTION

History is the conscience of our present society. Those who fail to understand the past may repeat its errors. These are good justifications for preserving historic sites and monuments. As far as documentation and desktop publishing is concerned the computers make it easier to create, edit, preserve and finally retrieve records of our intellectual.

Preservation is the activity of protecting something from loss or danger. Preservation can be done for betterment, improvement,

advancement - a change for the better; and for progress in development. Digital preservation includes the preservation of print and non print material in digitised form for effective, efficient and purposeful use.

We have come a long way from 'preservation or retrieval' to 'preservation for retrieval'. 'Preservation or retrieval' was a phase when books were chained and kept in locked shelves to preserve them from loss, damage, theft, fire, etc. Then came the time when the concept of 'use by all' came in. Books were no longer chained in. But they

were not all that secure, or fully utilised, because the libraries still have not many copies of those books which could satisfy their readers or were theft-proof. Twenty years ago, paper filing and capturing documents on film were common preservation methods. Computers came and suddenly changed the world of libraries, not only in terms of information access by means of computerised cataloguing but also full text was being made available on the desktop. Desktop computers have enabled the capture, processing and dissemination of much more information than was ever possible. In this form the documents were considered not only secure as their original version could not be changed, but also the retrieval was easy, fast and efficient. It was called 'preservation for retrieval'.

Now the question arises why should a librarian digitally preserve material. What will happen if every library starts conversion and preservation of all its material in digitised form. Is it possible? If yes, will it be worth the effort, the labour, the cost and the time? Why should a library preserve something that is accessible at some other place with 24X7 availability, even if it owns those resources (for example should it digitise the back volumes of journals). In our view it should be the responsibility of the publishing agency to take care of it, i.e., digitise them and make them available at a lower cost or at no cost to libraries because they will have the necessary infrastructure for conversion and it will be used continuously. That way they can maintain standards in their products, hence uniformity will be maintained. In a way, libraries' efforts to digitise will be saved and publishers who by and large depend on libraries for their sale will have libraries dependent on them even for future.

## **2. NEED OF DIGITAL PRESERVATION**

- To maintain the look and feel of the original record
- To update the document
- To make use of the record later on

- To store cultural/intellectual heritage
- To make it accessible throughout the world
- To retain the content and format of the document
- To store from the point of view of their contemporary relevance and potential
- Capability to digitise the analogue records is there
- Media can stay longer
- Retrieval is easier.

## **3. PROBLEMS WITH DIGITAL PRESERVATION**

- Different digitisation groups use different formats for preservation
- Different documents are dependent on software used to create them
- A platform independent version is required for retrieval by all
- Sharing of these documents at different platforms is difficult
- It is useful to store documents in a format that can download from the internet for large use, hence browsing through internet capability should be there.
- Technology used today may become obsolete tomorrow.

## **4. BEGINNING A DIGITISATION PROJECT**

Any digitisation project will require careful planning, designing, implementation, and evaluation. Paul Conway<sup>1</sup> describes what decisions are to be taken while preserving of documents.

### **4.1 Why Digitise and Preserve**

There must be a clarity regarding why libraries want to digitise and preserve material. May be accessibility to other sources/services is a long term solution than investing its own resources which are manpower, machine, money, time, etc.

The library has to consider the aspect if it wants to market the digitised product or just for free use by all or by its own users. The

library has to consider the purpose of conversion of its collection in digitised format such as—social cause i.e., providing information free of cost; for making money by marketing; to save duplication of efforts; to help others to create their own products; or to up date the product.

## **4.2 For Whom to Digitise and Preserve?**

Library has to decide who is the target audience? For example DESIDOC digitises and maintains a full text database of defence related newspaper clippings which is operational on DRONA (an intranet of DRDO) for use of its scientists. News items and editorial from 17 Indian newspapers published in English and Hindi language are covered in this database. Clippings database is updated daily online. A total of 2049 clippings were added in the year 2004.

Library also has to take a consideration regarding patrons' change in preferences and will the preserved documents be useful for those other than target audience?

## **4.3 What to Digitise and Preserve?**

Not all the information/documents are needed to be digitised or can be digitised which are available in the library. Selection regarding the right content for the right user at the right time is must and it includes defining value of information to be digitised, recognizing it, and then deciding to address its preservation needs in most appropriate way justifying the cost of a migration/conversion.<sup>2</sup>

Librarians and archivists have to choose which books, articles, photographs, films, and other materials are to be converted from paper or film into digital form. Decisions regarding the continuing value of digital files, with future storage and access systems are to be taken by project leaders. For example DESIDOC has taken an initiative to convert the negative of photographs of scientific and technical tests and experiments done by DRDO, technology products, award ceremonies of DRDO, and various other events/ceremonies in to digitised images

which will lead to direct visualisation and recognition of photographs. A detailed description has been given in detail later on.

## **4.4 How to Digitise and Make Access?**

Careful studies, and experience is required for conversion and preservation of material. The following points are to be considered while digitising and preservation:

- What equipments are to be used?
- Which software is needed?
- What standards are to be followed?
- What is life expectancy of storage media and access media.

As the storage media and access software can become obsolete with time and advancement in storage and communication technology, decisions regarding the longevity, and capability of the media is to be taken care of.

## **4.5 How Long to Preserve?**

The project team has to decide the duration of use of the digitised information whether it is for long-term (access to digital information indefinitely) or medium-term (access to digital materials beyond changes in technology for a defined period) or short-term (access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology).

As we know a chain is as strong as its weakest link. Quality control in reproduction from printed to digitised media largely depends on the limitations of capture-and-display technology. The primary goal of preservation quality is to capture as much intellectual and visual content as is technically possible and then present that content to readers in ways most appropriate to their needs. The organisation of the content in a digital document must take full advantage of its capabilities and make the search of information for users easier. During

conversion it is must to make sure that files are not altered accidentally or intentionally. Responsibility to create and maintain structural indexes and bibliographic linkages within well developed and well understood database standards to be shared across platforms, systems, and international boundaries is on the shoulders of the project team which ensure access to a preserved item or collection.

#### **4.6 Who Will Preserve?**

Many people in an institution fulfill many roles. An organisation has to identifying who can control which facet of digital technology, and which trends can be influenced and by whom.

### **5. EFFORTS OF DESIDOC TOWARDS DIGITISATION**

Development of digital libraries has changed the way the traditional libraries operate. The main focus of the libraries has shifted to making the full text of information available at the desktop of the users. DESIDOC has initiated the process of developing 'Digital Library' and undertaken a number of activities. Digitised versions of full-text of *Defence Science Journal* (1950-2004); Newspaper Clippings Service *Patrika* (2000-2004), *DRDO Science Spectrum* 2004; Popular Science & Technology Series (13 issues) were developed.

DESIDOC, being a central publication agency of DRDO brings out large number of regular and special publications to disseminate current information on various research and development activities being carried out by various DRDO labs/estts to the outside world.

The regular publications, such as *Defence Science Journal*, *Technology Focus*, *DESIDOC Bulletin of Information Technology*, *DRDO Newsletter* and *DRDO Samachar* have undergone various changes in terms of coverage, quality and timeliness. Large number of special publications were also brought out to meet occasional needs of the labs/estts and DRDO HQrs. Digital printing is

an important activity of DESIDOC for publishing these regular as well as special publications of DRDO.

#### **5.1 Defence Science Journal**

A full text database of the articles published in *Defence Science Journal* has been developed which is the first in house developed full text database of a premier DRDO publication which provides value added searches like author, title, keyword, etc. The database covers areas of Defence, Military technology, Electronics, Food Science, Physiology, Nuclear Medicine, Psychology, etc. The articles can be accessed by a large number of users via DRONA (DRDO Rapid Online Network Access). The database is also available on CD along with user friendly retrieval software.

#### **5.2 Popular Science and Technology (PST) Series**

Full text of thirteen volumes of PST series published by DESIDOC, are digitised, edited and redesigned to convert it into hypertext. The complete series is transformed into e-book form and is available on CD.

#### **5.3 Digitisation of DESIDOC Bulletin of Information Technology**

DESIDOC has made efforts to digitise the full text of the papers included in each issue of DESIDOC Bulletin of Information Technology (DBIT) for uploading on DRONA (Intranet of DRDO) and on internet site (as per policy). The typesetting of the text of DBIT is being done using the Corel Ventura Software. It supports the html formatting. Following steps are taken for each chapter of each issue to make html file or pdf file and upload on internet and DRONA.

##### **5.3.1 How to publish to HTML**

Corel VENTURA lets automatically map its tags to HTML tags. To publish to HTML the following steps are to be followed:

1. Open ventura file, Click File, Publish As, HTML.

2. Type location where to save the HTML file and name it in Main HTML Document text box.
3. Specify the range of the publication for conversion:
  - Publication
  - Current Chapter
4. Enable the Launch Browser check box to open browser
5. Click the Publish button.

We can create a hyperlink within publication to link to another part of publication or to link to another document on the web. If we are creating a hyperlink to another part of the same ventura publication, then we need to create a marker. A marker acts as a target for the hyperlink. We can insert the same marker in different locations in our publication.

### 5.3.2 Creating a pdf File of Full Text of DBIT

PDF files can be downloaded, viewed, and printed from several different platforms with the page layout and typography of the original document intact. First of all a prn file of a paper (figure 1) is created using ventura software (figure 2) since we are using ventura for publishing the DBIT.

Then adobe distiller software (figure 3) is used to create the pdf file. Figure 4 shows the pdf view of the paper. Then the files are uploaded on internet and DRONA.



Figure 1. Paper of DBIT in Ventura software

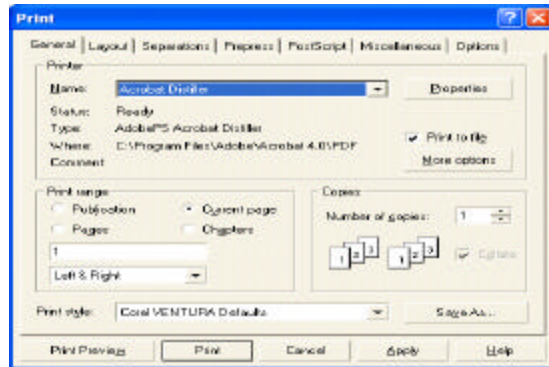


Figure 2. Creating a prn file

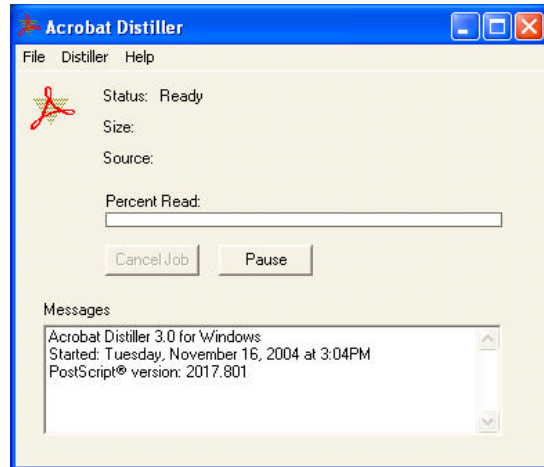


Figure 3. Creating a pdf file



Figure 4. pdf file of the paper

### 5.3.3 Creating an HTML File from PageMaker File

PageMaker creates HTML pages using a subset of the style tags available in the 3.2 version of the HTML specification; it exports some but not all browser-specific tags such as Netscape Navigator extensions to HTML.

HTML uses a limited set of named text formats which are conceptually similar to paragraph styles. PageMaker automatically converts copies of imported graphics (whether inline or independent) to GIF or JPEG. Shapes drawn with PageMaker drawing tools are not exported, with the exception of horizontal lines which become horizontal rules in HTML.

Page Layout: Using HTML tables, PageMaker can approximate multi-column page design, including elements outside margins, text and graphics that span columns, and text wrapping around graphics.

To assign a page as an HTML page:

- 1 Choose File > Export > HTML.
- 2 In the HTML Export dialog box, click New.
- 3 Enter a document title
- 4 Select Assign PageMaker Pages to define a page or range of pages to be exported as HTML
- 5 Click Done and return to the HTML Export dialog box
- 6 Click Document, and do the following:
  - Navigate to the location where to store the HTML document to and Name the document.
  - Select Save Images into This Folder to include images from the publication
- 7 Click OK. At this point we are only defining a target location, not actually saving the HTML document.
- 8 To open the HTML document (once the export process is complete) in a web browser or HTML editing application such as Adobe PageMill, click View HTML. Then click the Application icon to open a dialog box to specify which browser or HTML editor to use to open the exported HTML file. The application selected will open automatically when the export process is complete.
- 9 To generate the HTML, click Export HTML; otherwise, click Done.

### 5.3.4 Creating pdf File from PageMaker Document

PDF files can be viewed in web browsers using the Acrobat Reader 3.0 plug-in. To change PDF options:

1 In the Export Adobe PDF dialog box, click PDF Options.

2 If a table of contents, an index, or hyperlinks etc., are included in a PageMaker document, we can set options in the Hyperlinks section as follows:

- Select Preserve Table of Contents links to add a hypertext link between each entry in the table of contents and the page to which the entry refers.
- Select Preserve Index Links to add a hypertext link between each entry in the index and the page to which the entry refers.

(For the Destination Magnification option, select an option from each pop-up menu to tell Acrobat at which magnification and position it should show a page when opened by a hypertext link from a table of contents, index entry, or internal hyperlink)

- Select Preserve External Links to include the hyperlinks you specified to external URLs—that is, to destinations outside the current publication or booked publications. This option is unavailable if you did not define this kind of hyperlink in the current publication or booked publication.

3 If we have created a TOC or index in PageMaker document, we can set options of Bookmarks.

4 Click OK and pdf file will be created.

### 5.4 Digitisation of Negatives of Photographs

As we have earlier discussed the negatives of the photographs taken at different occasions at various DRDO Labs and hqrs are being digitised and a database is created and maintained at DESIDOC. The procedure being followed is discussed below:

- (a) First of all the negatives are scanned with the help of Umax scanner and Magicscan software (figure 5 & 6)



Figure 5. Scanning of negatives



Figure 6. Setting for scanning of negatives

- (b) Each negative gets converted into a digital image which is visible on screen and a unique classification number is given to each of these digital images which has an identification with the corresponding negative of that photograph. Each image is improved/edited using Adobe Photoshop software.
- (c) A database has been created using MS Access software (figure 7). Various fields such as identification number, name of the laboratory, event, category (function, technology, tech test, person, etc.), thumbnail image, negative number, placement of negative, type of photograph (digital, 35 mm, 70 mm), name and rank of person in the photograph, date on which photograph is taken, venue where the photograph is clicked, album no., file name, CD number, key words, etc., are created in the database. Each record contains a thumbnail image which by clicking gives the original size of the photograph (a hyperlinking facility is provided) alongwith complete information regarding the event through the fields mentioned above. Key word searching facility is included in the

technology, tech test, person, etc.), thumbnail image, negative number, placement of negative, type of photograph (digital, 35 mm, 70 mm), name and rank of person in the photograph, date on which photograph is taken, venue where the photograph is clicked, album no., file name, CD number, key words, etc., are created in the database. Each record contains a thumbnail image which by clicking gives the original size of the photograph (a hyperlinking facility is provided) alongwith complete information regarding the event through the fields mentioned above. Key word searching facility is included in the

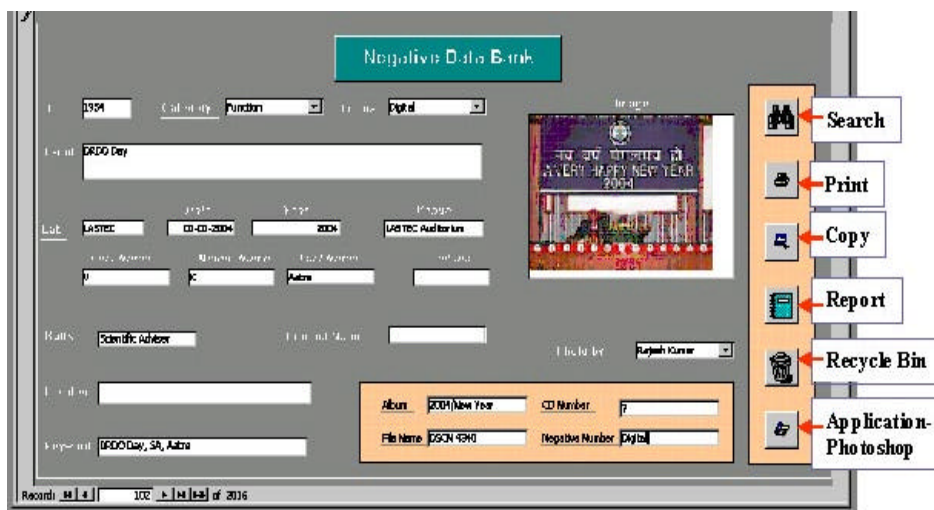


Figure 7. Database of negatives images

database. Search can be performed from any of the fields.

- (d) Whenever a demand is received, a search is performed then the original size of the photograph(s) can be seen and photographs can be printed either digitally on photographic paper using printer or through negatives in traditional way.

## 5.5 Newspaper Clippings Database

To keep the top management abreast of latest developments that have relevance to Defence R&D, defence technologies, defence policies, DESIDOC provides a daily newspaper clipping service. This service includes the scanning of Indian newspapers and magazines. Clippings are selected, marked and arranged on the following topics of interest: Ministry of Defence, DRDO, National security and defence policy, science & technology, International news, and other news items on defence science and technology.

The clippings are also scanned and hosted on DRONA for online access by other DRDO labs/estts. This database contains 11,927 records at present. Special issues are also brought out.

## 6. CONCLUSION

Digital products have made the life of librarians easier as searching is faster and they require less storage area. Definitely the digital image conversion requires a lot of efforts, understanding, and long standing institutional commitment with the excellence

in use of technology. Not only conversion but preservation is must which again requires deep knowledge and decision making regarding storage media, life expectancy of medium, file type for future use with such changes in digital scenario and integrity of original documents. DESIDOC has taken an initiative in this direction which is providing fruitful results by serving its users.

Though it has put lot of new responsibilities on shoulders of us as library and information professionals, it has transformed and improved our status and improved our services and sources to better serve our patrons.

## REFERENCE

1. Conway, Paul. The relevance of preservation in a digital world. <http://www.nedcc.org/plam3/tleaf55.htm>
2. Lynch, Clifford A. The integrity of digital information: Mechanics and definitional issues. *Journal of the American Society for Information Science*, 1994, **45**, 737-44.
3. [Amol.org.au/craft/employ/capture/course/course/glossary.html](http://Amol.org.au/craft/employ/capture/course/course/glossary.html)
4. Buckland, Michael K. Information as thing. *Journal of the American Society for Information Science*, 1991, **42**, 351-60.
5. <http://www.ciks.org/need.html>
6. Digital presentation: An introduction to the standards issues surrounding the deposit on non-print publications. [www.bic.org.uk/digpres.doc](http://www.bic.org.uk/digpres.doc)

<b>Contributors:</b> <b>Smt. Alka Bansal</b> , Sc 'C', Defence Scientific Informatgion and Documentation Centre (DESIDOC), Metcalfe House, Delhi - 110 054 <b>Smt. Vinod Kumari Sharma</b> , Sc 'E', DESIDOC, Metcalfe House, Delhi - 110 054 <b>Sh. Ashok Kumar</b> , Sc 'E', DESIDOC, Metcalfe House, Delhi - 110 054 <b>Dr Mohinder Singh</b> , Director, DESIDOC, Metcalfe House, Delhi - 110 054
--