Processes of Digital Content Development in the University Libraries of West Bengal

Sukumar Mondal
Former Student, Dept. of Library and Information Science, Jadavpur University, Kolkata, India
Saroj Mondal
Former Student, Dept. of Library and Information Science, Jadavpur University, Kolkata, India

Abstract
Today's world information is going to be a digital information .To achieves Digital content needs to follow some specific procedure which are discussed in this paper. This paper deals with digital content development process in four university library. For this purpose four university libraries in the west Bengal are surveyed. The finding of this study is major university library scanning the document through scanner and processing the scanned image through OCR software, such as ABBY Fine reader.

Key word*: Digital Content, Digital Content development, Digital content development software, digital resources of university library, Scanning, OCR.*

1. Introduction: Digital content is the important tool of digital library. Generally content is information and experiences that may provide value for and end user/audience in specific context. Digital content refers to information available for download or distribution through computer resident networked and it is not tangible to its user but can be use seamlessly. But before the distribution, should be followed some procedure, where documents or contents are creating, organizing, archiving and storing in the any type of storage medium. Such as database, CD-ROM, hard disk of computer, server files storage or digital preservation repository system which also called Institutional repository system. And this whole process is called digital content development system. . Now this time, like other development country Indian's much university taken digital content development projects. But different country, different state, different institutions followed different policy in the fields of digital content development system. In this study to find out that the trend and status of digital content development process in the university libraries of West Bengal.

2. Objectives of the Study: The purpose of this research is to find out/investigation of the 'trends and status of digital content development in the university libraries of west Bengal'. In order to achieve the goal of study will precisely attempt:

1. To investigate current situation of digital content Development process in the university libraries   of West Bengal.
2. Which planning and funding policy they are used for develop of the digital content?
3. To find out which equipment they are used for developing digital content.
4. To investigate which software they are used for developing digital content.
5. To investigate that which problems they facing to the work.

3. Methodology: Document research will be applied for making related and relevant literature review and investigation or study the standard component or tools (such as scanner, digital camera, software etc.) for creating a digital content and building a digital library or an institutional repository system infrastructure. Descriptive or survey method use for investigate current status, find out content development process. Observation, questionnaire and interview system is the important process including survey method. We apply participant observation method for directly saw the process and questionnaire method follow for getting some information about their policy, infrastructure, strategy, collection etc. and interview method are apply for known their satisfaction level.

4. Literature Review: The process of digital content development involves converting the existing library materials or content into digital format. The physical or analog object is 'captured, by some device such as a scanner, digital camera or recorder, which converts the analog features of the project to numerical values, enabling them to be read electronically (**EADIE,** 2005). Development of existing text can be carried out through two main methods-transcription and optical character recognition (OCR). Scanning using OCR software is another method which is used to digitize text. **KENNEY AND RIEGER,** in their digital imaging tutorial (2000), discuss these issued in greater detail and suggest that some questions should influence the selection of a scanner. The RLG, GLIR and DLF guides to 'quality in visual resource imaging', particularly guide number 2, selecting scanner (WILLIAMS, 1998) .The software that runs the scanner is also important. It should be straight forward to use and incorporate an ability to run batch scans to save capture time, where applicable (**NOF-DIGITIZE,** 2002). Scanning software comes with the scanner and helps to create image files in formats. Such as TIFF, JPG, GIF etc. OCR software allows the option of maintaining text and graphics in their original layout as well as plain ASCII and word processing formats. Some of the commonly used OCR software are care's omni page and Xerox text bridge, ABBYY Fine Reader (**SAHOO,** 2003).

The primary function of metadata, therefore are facilitate the identification, location, retrieval, manipulation and use of digitized objects in the networked electronic environment. OAIS identifies the type of metadata to support digital resources (**CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM,** 1999). CEDARS has now develop an outline specification describing the complete set of metadata elements that attempts to reflected the multiple needs and functions that need to be addressed. Dublin core metadata element set applicable to a variety of digital object types.

The best way to ensure long-term access to digital data is to use standard formats and open systems (such as OAIS) wherever possible, and to have a preservation and sustainability strategy for the project (**DEEGAN AND TANNER,** 2002). There are presently several technical approaches to managing long-term access to digital data. These includes: Refreshing, migration and emulation (**LAWRENCE ET AL.,** 2000). BESSER (2000) outlines three way in which metadata, if properly used, can assist in the long-term preservation of digital data.In a survey of risk factors for technology projects, the following reasons were identified as the cause of project failure.

    32% - inadequate project management and control

    20% - lack of communication

    17% - failure to define objects

    17% - lack of familiarity with project scope and complexity

    14% - incorrect technology, project size & other.

<div align="center">(TANNER, 2001).</div>

However above this discussion we cannot know the current situation of digital content development in the university libraries of west Bengal. Development process supported two

important things: technology and another is economy. In these fields west Bengal is a poor state of India. Generally want to know that 'what is the current situation'?

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

_ Hardware (Scanners, computers, data storage and data output peripherals)
_ Software (image capturing and image editing)
_ Network (data transmission)

The whole process of digital content development consist some of following activities:
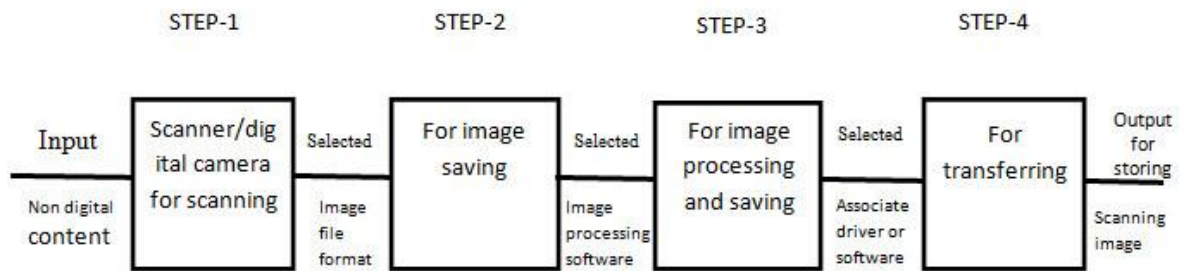
6. Data Analysis and Findings: The proposed study has been intended to keep restricted within the four university libraries near Kolkata. Which are Jadavpur University, University Of Calcutta, Rabindra Bharati University and West Bengal University of Technology. Rabindra Bharati University and West Bengal University of Technology are not taken development project but they are stored digital content, which they are collected in digital format (Such as soft copy of project report, research report and other).

6.1. SCANNING OF THE NON-DIGITAL CONTENT: Scanning is the important steps of digital content development. In this section we wanted to know that 'which equipment they are used for scanning' and 'which software related these scanning processes'? We also want to know that 'which image file format they are used for saving scanning images'?

Using scanning equipment in the four university central library:

Type of scanning equipment	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL
scanner	✓		Not taken digital content creation project	Not taken digital content creation project
Digital camera		✓		
other	Not found	Not found		

From the above table it is seen that mainly two typeS of scanning equipment use for scanning the non digital materials. JU central library use the scanner, naming Minolta ps7000 over head scanner and CU central library use the digital camera, naming Cannon Eos600D digital camera. Other two universities not took digitization project. So they are not using any type of equipments. Using good (quality) scanner better than digital camera. Because there are not any setting problems. Other hand

one time setting digital camera if replace for any fact; again do this setting. I don't know, maybe have another various technological problems of both equipment, but they are not share with me.

6.2. Saving for the Image Processing: Now I inform that which image file format they are used for saving scanning image? Because it is a very important step of the digital content development process. If you want to see the scanning image or transfer it to other space, must be needed save the images, other than it is not possible.

Selected image file format for saving scanning image:

Name of the format	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL
TIFF	✓	Information not provide Not found	Not taken digital content creation project	Not taken digital content creation project
JPEG	✓			
other	Not found			

From this table it is seen that JU CL use TIFF and JPEG image file formats for and CUCL use JPEG formats for saving the scanning image after image processing. I wanted to know that 'why they are use of this format' and 'why they are not save directly pdf format'? They answer that if you saving directly pdf format, then they are not doing image processing works. So overall it is clear that if you control the images, scanning image must be saving to the image file format.

Now I inform that which software they are used for transferring the scanning image from scanner to image processing software for control the quality of scanning image.

Selected software for transferring scanning image:

Name of the software	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL
ABBY Fine Reader	✓		Not taken digital content creation project	Not taken digital content creation project
ACDsee9		✓		
Other	Not found	Not found		

From the above table it is seen that JU central library use ABBY Fine Reader software, which is a one type of commercial OCR software and it is done the whole work after scanning to before archive. CU central library use ACDsee9 software for transferring the scanning image into image processing software. Other two universities not took digitization process. So overall discussion I think that ABBY Fine Reader is better than ACDsee9 software.

After scanning, image processing is the next and important step of the digital content creation or development system for controlling the image quality. In this section we wanted to know that which image processing software they are use. Which image processing parameters they are use? Which type of color they are adjusting? Also known that how much dpi resolution they are use for controlling the image brightness?

6.3. Image Processing: Now I inform that which image processing software they are use for control the image quality.

Selected image processing software:

Name of the software	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL

ABBY Fine Reader	✓		Not taken digital content creation project	Not taken digital content creation project
Adobe Photoshop CS		✓		
other	-	-		

From the above table it is seen that JU central library use ABBY Fine Reader software for image processing work and which is a one type commercial OCR software and it is done the whole work after scanning to before archive. CU central library use Adobe PhotoshopCS software for image processing. Other two universities not took digitization process. So overall I think that ABBY Fine Reader is better than other software.

Now we see that which parameters they are use for controlling the scanning image.

Using parameters for control the images:

Name of the parameters	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL
Noising	✓	✓	Not taken digital content creation project	Not taken digital content creation project
Image enhancement	✓	✓		
Intelligent Crop	✓	✓		
De-Skew	✓	✓		
Color adjustments	✓	✓		
reformatting	✓			
Resolution setting	✓	✓		
other	-	-		

For controlling the image page, there is available various types of parameters into a image processing software. But bring in better quality, needed some basic parameters, such as Noising, Image enhancement, Intelligent Crop, De-Skew, Color adjustments, Resolution setting, reformatting etc. JU central university use all (on the table) parameters. As well as CU central library use all parameters except reformatting. And other two universities not have taken digitization project. So overall discussion it is seen that JU and CU both central library generally maintain the standard in the field of image processing works.

Color adjustment is another stage of this workflow. Bitonal or binary scanning represents one bit per pixel (either "0" (black) or "1" (white). Gray scale scanning is used for reliable reproduction of intermediate or continuous tones found in black & white photographs to represent shades of grey. Selected color:

Type of color	JUCL	CUCL	RBUCL	WBUTCL
Black & white	✓	✓	Not taken digital content creation project	Not taken digital content creation project
Grey scale	✓	-		
other	Not found	Not found		

In the international level mainly using two types of color formats. One is black & white and another is grey scale. Sometimes many organizations have taken other color option. However JU and CU central library follow the standard. JU central library use both black & white and grey scale and CU central library use only black & white color image format. So I calculate the overall percentage for knowing which color use in maximum fields? coming following the figure.

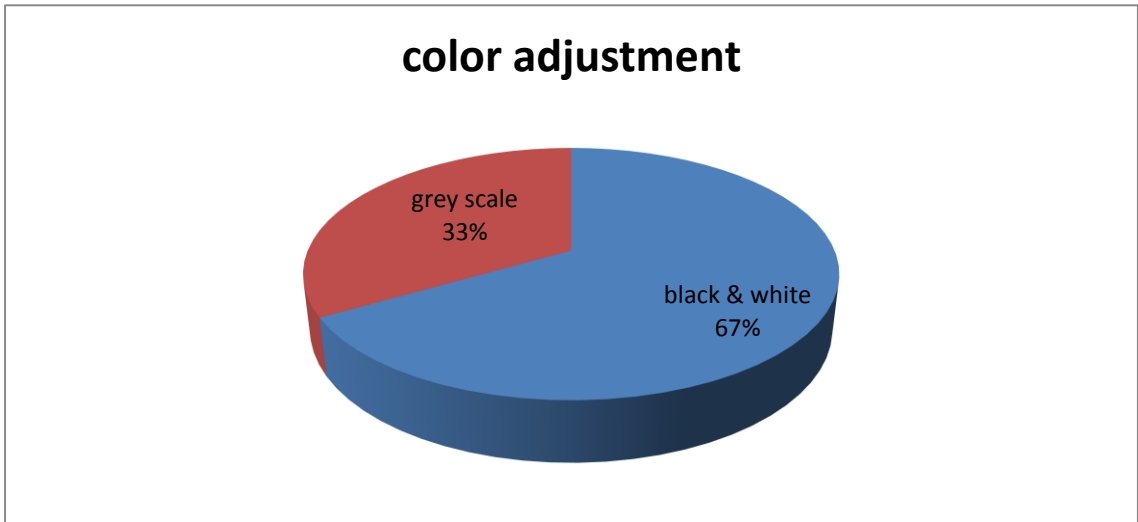## color adjustment

grey scale 33%

black & white 67%

Figure- Color adjustment.

Calculate this figure mainly based on those university libraries, which are taken digital content creation project. So it is seen that in case of color adjustment 67% university choose black and white color and 33% university choose grey scale color.

Now we are see that how much dpi resolution they are use for controlling the image brightness?

Resolution is use for digitization:

Nature of dpi	JUCL	CUCL	RBUCL	WBUTCL
various	-	✓	Not taken digital content creation project	Not taken digital content creation project
300	-	-		
400-500	-	-		
600-1500	✓	-		
other	-	-		

JU central library use 600dpi and CU central library use dpi according to various image page.

OCR or Optical Character Recognition is the process of recognizing characters from scanned document image. In this process, an image with textual content is read and the characters present in the text are output by the OCR software. However now I inform that which OCR software they are use and what is the purpose of the use.

Using OCR software to recognize the character from the digitized or scanned image:

Name of the software	Name of the university			
	JUCL	CUCL	RBUCL	WBUTCL
ABBY Fine Reader	✓		Not taken digital content creation	Not taken digital content

ACDsee9		✓	project	creation project
other	Not found	Not found		

purpose of the OCR processing software in four university library.

University libraries	Purpose of the OCR processing software			
	Automatic indexing	Computer assisted reading	Image transferring	other
JUCL	✓	✓	✓	✓
CUCL	✓	✓	✓	-
RBUCL	-	-	-	-
WBUTCL	-	-	-	-

From the above tables it is seen that JU central library use ABBY Fine Reader OCR software and they are use the software for the whole work after scanning to before archive. CU central library use ACDsee9 OCR software for purpose of Automatic indexing, Computer assisted reading and Image transferring. Other two university central library's not taken digitization project.

Maintaining metadata is a crucial part of the digital content development project for searching and disseminating the digital content in online or offline both. So it is very important part of selection of proper metadata schema. There is various type of standard Meta data schema available in the world, Dublin core metadata set is the most popular schema in the market. However now we see that what type of metadata schema they are use.

Following standard metadata to archive the digital content in four university library.

Type of standard	University libraries			
	JUCL	CUCL	RBUCL	WBUTCL
ANSI/NISO Z39.85-2007 Dublin Core Metadata Set:	✓	Information not provide	Information not provide	Information not provide
Encoded Archival Description [EAD]:	-			
Metadata Implementation Strategies [PREMIS]:	-			
Metadata Encoding & Transmission Service [METS]:	-			
Encoded Archival Context [EAC]:	-			
other	-			

In the field selection of standard metadata schema only JU central library provide their information. They use ☐ANSI/NISO Z39.85-2007 Dublin Core Metadata Set. But other libraries not given information about this topic.

6.4 STORING: After image processing work, need to be transferred from it to the hard disc of scanning workstation to external large capacity storage devices such as an optical disc, CD ROM/DVD ROM disc, snap servers etc. or Stored as part of a digital preservation repository system etc.. However some important topic related this section, such as which preservation strategy they are

use? What format they are use for storing? Which software they are use for displaying? Where they are store etc. however first we are see that which preservation strategy they are use?

Following strategy to managing digital holdings into four university libraries:

University library	Type of strategy			
	Migration policy	Refreshing policy	Emulation policy	other
JUCL	-	-	-	✓
CUCL	-	-	-	✓
RBUCL	-	-	-	✓
WBUTCL	-	-	-	✓

From the above table it is seen that every university central library's not follow the standard of preservation strategy. They are preserve their digital content using by various system, but they are not share their system information.

Now I inform that which format they are use for storing digital content.

Using format for storing digital content into four university libraries:

University libraries	Type of format			
	image	pdf	Standard office document	other
JUCL	✓	✓	-	-
CUCL	✓	✓	✓	-
RBUCL	-	✓	-	-
WBUTCL	✓	✓	✓	✓
Total %	75	100	0	25

From the above table it is seen that JU central library use image and pdf format. CU central library use image, pdf and standard office document file format. RBU central library only pdf format and WBUT central library use image, pdf, standard office document and also multimedia format. So if you calculate according to percentage of overall aspect, coming the following figure:
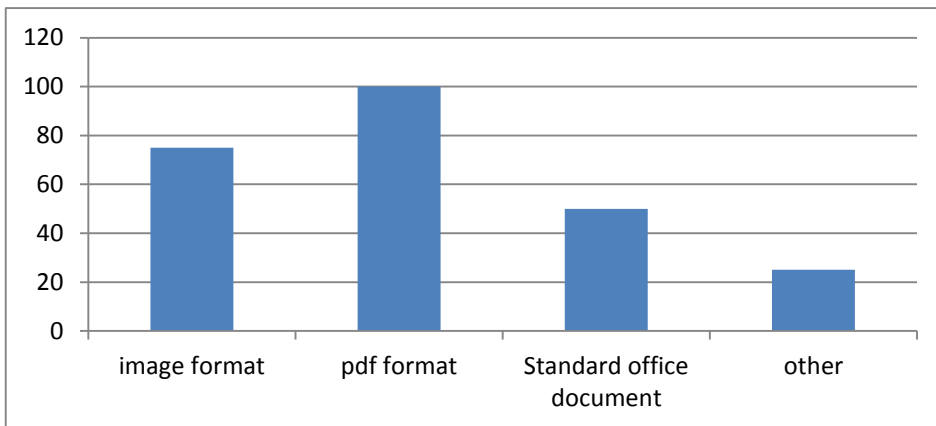

Figure- Storing formats.

According to figure all organization use pdf format and case of other format 75% organization use image file format with pdf format, 50% organization use standard office document with pdf and other format and 25% organization use other format, specially multimedia with other format.

Now we see that which image file format they are used for storing digital content.

Using image file format for storing digital content in four university library:

University library	Type of format			
	jpeg	tiff	gif	other
JUCL	✓	✓	-	-
CUCL	✓	-	-	✓
RBUCL	Not found			
WBUTCL	Not found			

From the above table it is seen that JU central library use jpeg and Tiff image file format and CU central library use jpeg image file format. CU central library's information provider said that they use another image file format, but he cannot provide details information. Other two universities not provide information about this topic.

Now we are see that which software they are used for display the digital content

Using software for display digital content into four university library:

University library	Type of software				
	Microsoft image viewer	Netscape	Pdf reader	Adobe reader	other
JUCL	✓	-	-	✓	✓
CUCL	-	-	✓	-	✓
RBUCL	-	-	-	✓	✓
WBUTCL	✓	-	-	✓	✓
Total%	50	-	25	75	100

So it is seen that display of pdf file all university use adobe reader except CU central library, but they use similar another software, naming pdf reader. But display the image file format it is not clear which software are mostly uses. JU and WBUT central library use Microsoft image viewer software but in the field of other university, they are not given details information about this topic. However if you calculate the percentage according to overall data, becomes the following figure:
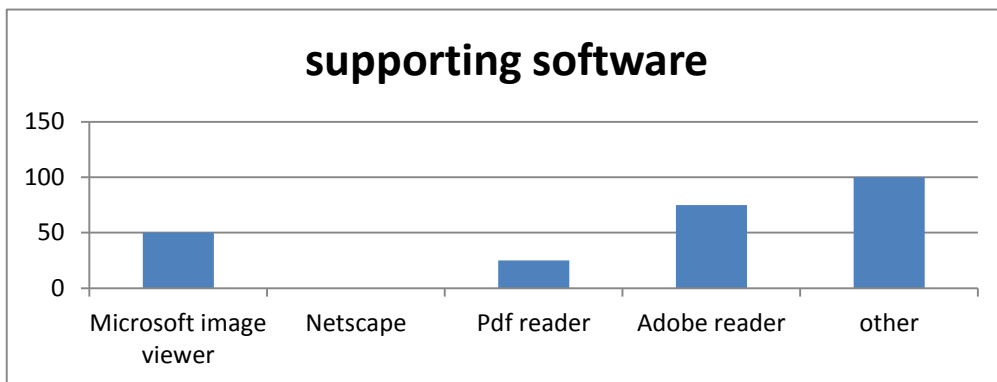

Figure- Supporting software for display.

From the above figure it is seen that 75% organization use adobe reader and10% organization use pdf reader for display pdf file and in the field of image 50% organization use   Microsoft image viewer software and other software use all organization, but they not provide details information. So according to overall discussion I think that if all libraries provide their details information, and then change the figure.

Now focused on the storage medium, which is use for store and dissemination of digital content. Stored the digital content for provide service into four university libraries.

medium	University library				
	JUCL	CUCL	RBUCL	WBUTCL	Total %
Stored on CD-Rom or DVD:	-		-	-	0
Stored on tape drive or hard disk of computer, with backup	-	✓		-	25
Stored on server file storage, with back up			✓		25
Stored as part of a digital preservation repository system:	✓			✓	50
other	-	-	-	-	-

 From the above table it is seen that JU and WBUT central library use institutional repository system for storing their internal and external both content in digital format. CU central library Stored on tape drive or hard disk of computer, with back up and RBUCL Stored on server file storage, with backup system.. However it is clearly present by using following figure:
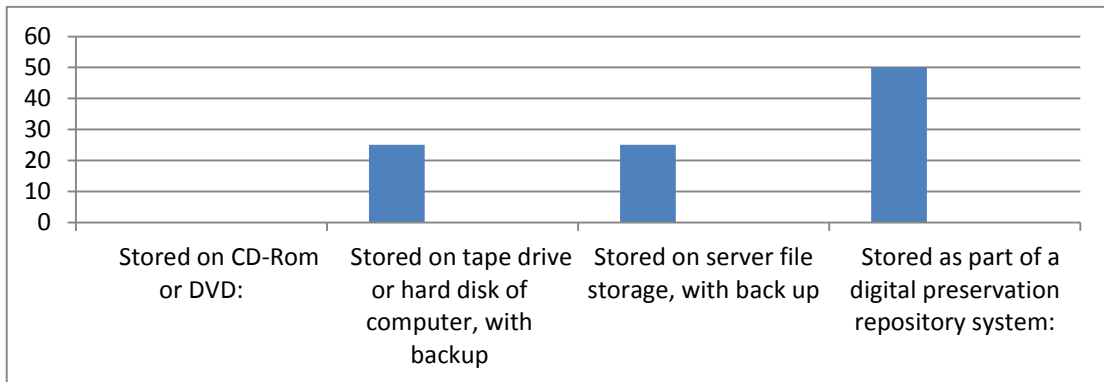

Figure- Storage medium.

From the above figure it is seen that 50% organization use institutional repository system, 25% organization stored on tape drive or hard disk of computer, with backup system and 25% stored on server file storage, with backup system.

7. Conclusion: A university library's digital section or institutional repositories system expected to provide access to the digital information collection and university's academic user want to access quickly and quality information. Successfully fulfil this expectation the university libraries needed taken the project of digital content development. This dissertation helps those university libraries

which are taken a new project. This study will be helpful to fill up knowledge gap on the area. The finding of this study major university library scanning the document through scanner and processing the scanned image through OCR software, such as ABBY Fine reader and when ready the content for archive, then store the intellectual assets in the institutional repositories software, such as DSpace. Proper planning, implementing, monitoring and evaluation of content development programmed may be done based on the findings of this study.

Bibliographical Reference:

1. Besser, H. (2000). Digital longevity. In sits, M.K.(ed.), Handbook for digital projects : a management tool for preservation and access, Northeast Document Conservation Center(NEDCC).
2. CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS) (2001). *Reference model for an open archival information system (OAIS). Red Book. Issue 2* (No. CCSDS 650.0- R-2). Washington, DC, National Aeronautics and Space Administration. Retrieved from http://www. ccsds.org/documents/ pdf/CCSDS-650.0-R-2.pdf
3. Deegan, M., & Tanner, S.(2002). Digital futures : strategies for the information age. In Hughes, L.M (author), digitizing collections (p.145-162). London : Facet Publishing.
4. Eadie, Mick (2005).The Digitisation Process: an introduction to some key themes.AHDS. Retrieved form: http://ahds.ac.uk/creating/information-papers /digitisation-process/
5. Kenney, A.R., & Rieger, O.Y. (2000). Moving theory into practice : digital imaging for libraries and archives. Mountain view, CA : Research Libraries Group.
6. Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H., & Kenney, A. R. (2001). Risk management of digital information: A file format investigation. Journal of Academic Librarianship, 27(5), 417-418. Retrieved from http://search.proquest.com/docview/57508623?accountid =16284
7. NOF- digitize technical advisory service.(2002). The digitization process. In Hughes, L.M (author), digitizing collections (p.145-162). London : Facet Publishing.
8. Sahoo, Bibhuti Bhusan. (2003). Digitisation of Print Materials, Audio and Video. Retrieved from : https://drtc.isibang.ac.in/bitstream/1849/ 47/2/M_OCR demo_ bibhuti.pdf
9. Tanner, S.(2001). From vision to implementation : strategies and management issues for digitizing collections. Retrieved from : http://heds.herts.ac.uk /resources/papers/Lboro2000.pdf.
10. Williams, D. (1998). Selecting a scanner, guide to quality in visual resource imaging, guide no.2, RLG/CLIR/DLF. Retrieved from : www.rlg.org /visguides/visguide2.html.

Abbreviations:
 JUCL: Jadavpur University Central Library
 CUCL: University of Calcutta central library
 RBUCL: Rabindra Bharati University Central Library
 WBUTCL: West Bengal University of Technology Central Library