POLICY & PROCEDURE FOR DOCUMENT DIGITISATION

VERSION 1.0

The Jute Corporation of India Limited
15N, Nellie Sengupta Sarani
7th Floor
Kolkata – 700 087

SL No	Policy No	Version	Created By	Revised By	Approved By	Effective Date
1	JCI/IT/2020/02	1.0	AM (IT & Tech)	-	CMD	10/11/2020

1. Digitisation Process

Digitisation Process comprises of following steps:

1.1 Collect and prepare the documents
1.2 Mode for Scanning
1.3 Scanning specifications & format
1.4 Storage of the scanned documents
1.5 Post-Scanning

1.1 Collect and prepare the documents

The documents identified by individual units are collected. The involvement of the Section concerned is considered important as the knowledge of the files, documents will be required. The necessary signoff needs to be provided by the Digitisation Coordinator on the number of Files and Documents received. The documents are:

- Documents must be carefully separated. If stapled, pins need to be carefully extracted. Page separators or post-it notes can be placed for separating the categories.
- Any external annexure like photographs is to be fixed carefully and neatly, if required with glue.
- Proper Dusting of the document Dust may affect the quality of the image.
- Ironing and Smoothening of the document this is done in the case of documents that are folded at the edges.
- Page Numbering-Each page of the document is numbered on the corner. This helps in maintaining the count and also in avoiding any misplaced sheets of any document during the course of file movement.
- Clip the document set.

Important Points:

- ✓ Make sure that the scanned documents are readable.
- ✓ Collect the originals from the scanner, collate and clip the documents once again.
- ✓ All the physical records (per connection) shall be bound in a file. The file should have the details of the content and a unique identifier on the top of the file to help faster retrieval.
- ✓ If DPI (Dots per Inch) is more, the photo is sharp but the size of the file is more.
- ✓ If DPI is less, the photo is dull but the size of the file is less.

1.2 Mode for Scanning

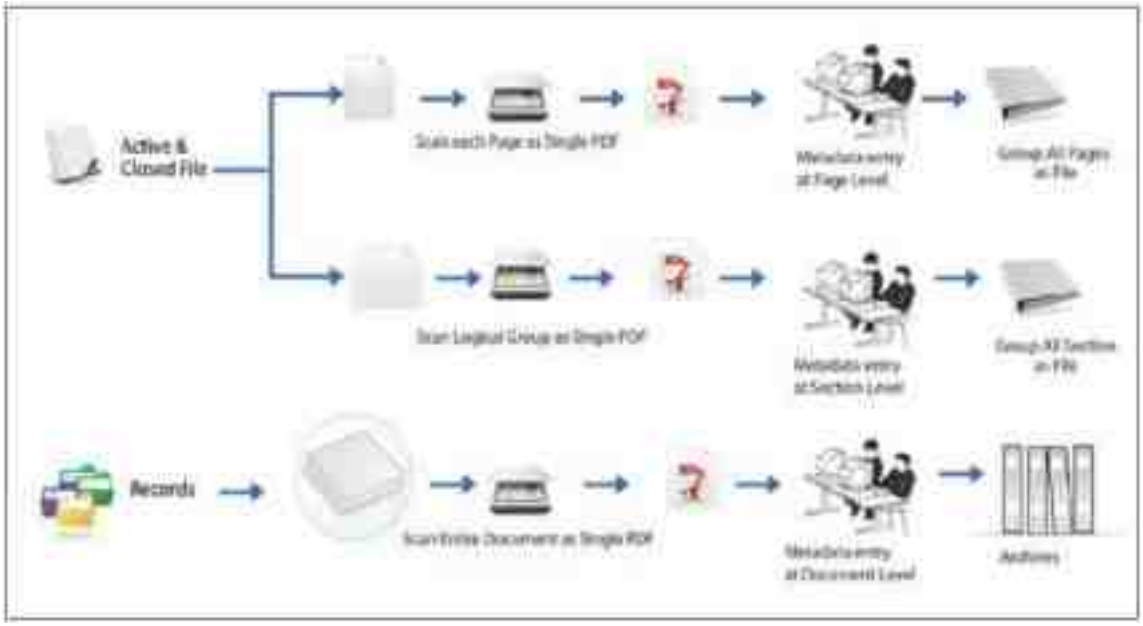
There are various modes and options are available for scanning files and records. Depending on the contents and complexity of the file, it can be scanned as a single pdf or each page can be scanned in the file as a separate pdf. The different methods can be applied to different components of the same file as well. For example, all correspondences may be broken into logical sections and scanned such that there

is one pdf for each section. At the same time, notes in the file can be scanned page by page as each note is a significant component of a file.

The contents of an active file, i.e., Correspondences, Notes, Issues, DFAs, etc, may be scanned in either of the two ways:

i. **Page-by-page:** Each correspondence is scanned as a separate image and a separate pdf file. This is more time consuming but leads to easier indexing of the images.

ii. **Bulk Scanning:** All documents are scanned into the same image and the same file. In this case each page needs to be identified and indexed separately as per its type and use.



Mode of Scanning

Closed files can be digitised in the same way that active files are digitised with some differences in scanning and indexing.

1.3 Scanning Specifications & Format

Document Type/ Condition	Color & DPI
Regular text	100 dpi B/W
Text with images	300 dpi Grayscale
Very damaged /Tarnished/Clouded	450 dpi B/W
Seriously damaged / Tarnished/ Clouded	600 dpi B/W
Documents with Photograph	600 dpi Grayscale/Color

The choice of storage format for electronic documents can have significant and far-reaching consequences.

1.4 Entry of the Metadata

The required metadata is entered against each document. The indicative metadata is as follows:

Active Files	Closed Files	Recorded files
• File no.	• File no.	• File no.

• Subject Description • Subject Category • Subject Sub Category • Language • Remarks • Previous Reference • Next Reference • Created On	• Subject Description • Subject Category • Subject Sub Category • Language • Remarks • Previous Reference • Next Reference • Next Reference • Created On • Closed by • Closed On • Closing Remarks	• Subject Description • Subject Category • Subject Sub Category • Language • Remarks • Previous Reference • Next Reference • Next Reference • Created On • Closed by • Closed On • Closing Remarks • Recorded by • Recorded On • Recording Remarks

1.5 Storing the scanned documents

Temporary

- Scanned copies of Notings are saved in the Notings folder under the eFile folder.
- Scanned copies of correspondences are saved in the Correspondence folder under the eFile folder.

Permanent

Once the physical files have successfully been scanned and indexed, their storage must be done effectively so that it is linked with the corresponding eFile. The eFile number must be clearly mentioned on the physical file at the right-hand corner and the physical file can be stored in the following manner:

- A record room must be assigned in the section where the file is to be stored.
- Staff must be clearly assigned responsible for lock & key to the record room. In case of confidential files, separate lockers must be assigned for each file (these lockers could be placed in the office of the concerned officers) whose lock & key will be assigned to the officer to whom the file is addressed.

- Care must be taken to clearly label the physical files in serial order or through some other customized labelling convention so that they can be clearly linked to the eFiles.
- A register must be maintained for tracking the issuance and return of these physical files. Each time the physical file is issued and returned; an entry must be logged in the register.

Accurate labelling of stored files is critical for the success of the digitisation process as active physical files should be readily available for reference from time to time.

1.6 Post-Scanning

Processing & Quality Control

The scanned images need to be processed for removal of punch holes and shadows, checking of pixilation, sharpness and skew etc. A combination of automatic/ manual processing using image enhancement software and experts can be used for processing and Quality Assurance of the scanned image.

Indexing

Indexing is one of the most critical steps of the digitisation process as it establishes the identity of the scanned files so that it can be referenced/ linked and retrieved with ease at a later stage. Indexing provides each document with a unique identity. Indexing can be of two types:

- Metadata Indexing: Indexing if done only on the metadata parameters.
- OCR (Optical Character Recognition) indexing.

It is a comprehensive content-based indexing method which reads and indexes the contents of each page.

Final QC and Security

The scanned images along with the indexed data are put through a final Quality Check. All scanned files must conform to the digitisation guidelines being followed in the process and should contain the minimum amount of metadata defined for the set of documents.

- Check the first and last page of the scanned output.
- Total Number of pages of scanned output to be verified against the number of pages in the original document.
- Clarity in output and completeness with respect to the contents in the scanned output.

Physical files can go through multiple iterations at multiple locations/users. In the absence of appropriate safeguards, it is relatively easy to alter or delete them, intentionally or unintentionally. Alterations to electronic files can be virtually undetectable, undermining their evidential value as files.

2. Guidelines, Standards & Procedures

Guidelines, Standards and Procedures for managing digital records are an important element of the Strategy framework. It aims to define the organization's approach to managing digital records and provide the necessary senior management authority for the implementation of the framework. Procedures outline how the guidelines and Standards will be implemented.

2.1 Guidelines for Metadata

Metadata is the data describing the context, content and structure of records and their management over time. It allows users to control, manage, understand and search records over time through a set of pre-defined parameters that can be captured for every scanned document. Structured and comprehensive metadata captured can help to:

- Identify, authenticate, and categorise the records.
- Enabling topic-based references and retrieval of specific files/ records.
- Ensure accessibility and accountability of files/ records by defining the access privilege metadata.

Metadata can be of various types depending on the level of information to be captured and the intended use and can be classified as follows:
Regular metadata – contains information about the contents of the file

- the title of a record.
- the subject it covers.
- its format.
- the date of Creation of record.
- history.
- details of its disposal.
- keyword must also be captured while preparing metadata.

Administrative metadata contains information such as sender address and location of scanning

Structural metadata contains information pertaining to the structure of the file/ document such as size of page, type of document etc.

For appropriate creation and maintenance of metadata, it is important to consider the various types of files/records that are used in the departments/ministries.

Guidelines and procedures that may be employed are:

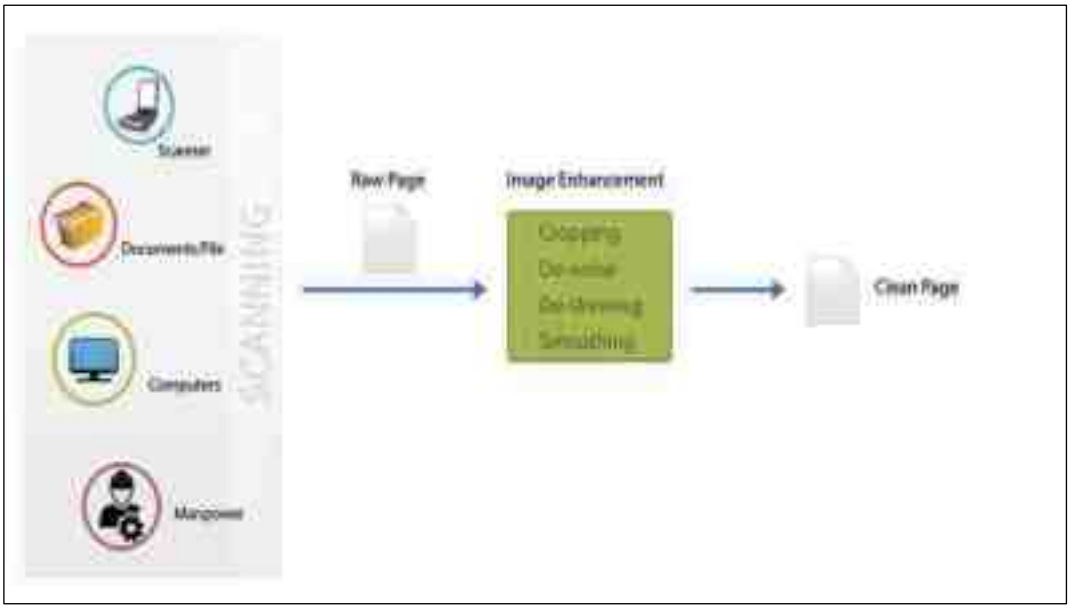
- identify the metadata elements to be captured.
- establish when and how metadata is to be captured.
- ensure that the metadata needs to be retained.
- detail the need to store metadata along with the scanned record to maintain the linkages.
- assign roles and responsibilities for capturing and managing metadata.

2.2 Guidelines for Storage and Storage format

Guidelines for storing the scanned files need to be evolved. For active files, the storage can be in the respective client system of the user who is performing the scanning as these are normally migrated using the Migration Module. For files that are identified to be outsourced, the storage can be on a Central File Server identified by the department so that systematic and periodic back up can be taken. Accurate labelling of stored files is critical for the success of the digitisation process as active physical files should be readily available for reference from time to time.

Open Standards Storage format is to be adopted for storage. Technology Standards prescribed for eGovernance as per the eGovernance Standards and Digital preservation policy is to be followed.

2.3 Guidelines for Maintaining Quality Control



Quality Control Process

Most software and hardware that will be used in a digitisation program will provide a range of variable parameters such as image resolution and output file format, and informed choices need to be made on each of these.

Bit depth

The nature of the documents being digitised should be the main factor for the bit depth used in the images produced. Capturing a document at a bit depth which is lower than recommended will possibly result in an image that is visibly different from the original record. In some situations, this visible difference and loss of information will be acceptable. For example: when digitising a document with black and white content, but a coloured letterhead, the loss of colour in the letterhead may be acceptable. Choosing a higher than recommended colour depth, such as 24-bit colour for a black and white document, will not provide any benefits, but will result in an increase in the file size of the image produced and may even introduce small areas of extra colours not present in the original document.

As is the case when determining the resolution to use, the mode of use of the digital images should be considered when deciding upon an appropriate bit-depth. If imaged pages will most often be viewed on computer screens, the use of a higher than normal bit-depth may be warranted. Capturing a document that contains a watermark, highlighting, or hand written annotations into an image with low bit depth may cause text to be obscured leading to a loss of information.

Capture Resolution

As the resolution of an image cannot be increased beyond that at which it was originally digitised without recapture, it is crucial that an appropriate resolution is selected prior to image capture. Higher Pixels per Inch (PPI) settings will result in images which are able to contain more detail per inch while increasing the file size of the resultant image. It should also be noted increasing resolution beyond certain thresholds will not provide a more useful image with existing viewing and printing technology.

The time and effort required to locate a paper record, prepare it for scanning, and return it to storage need to be considered when determining what resolution will be used. Setting the resolution of the capture device at its highest level for the best possible quality is recommended to avoid the need to rescan the paper record.

Compression and file size

Some form of compression should be applied to digitised records to enable storage and access in an efficient manner.

- **Lossless compression** provides file size reduction while being able to reproduce an exact, true and accurate digital copy of the image created at time of digitisation. Wherever possible, lossless compression should be employed.
- **Lossy compression** is not suitable when original paper records are authorized for early disposal as the accuracy of the image may be called into question. However, when originals are being retained, the additional file size reduction that lossy compression provides can mean that a small, perhaps indistinguishable, loss of data may be acceptable for some file types. When employing lossy compression techniques, the resulting image should not appear noticeably different from the original paper record.

Colour Management

Colour management must begin with correct scanner operation and the time of capture so that the original scan is as accurate representation as possible. The aim point adjustments and white neutralization provide the starting point for good colour management. However additional steps must be taken to capture colour images that represent the original document well. When possible, the original document should be compared to the scanned image under controlled viewing conditions on a colour-controlled monitor. Using graphics processing software such as Photoshop, the image may be adjusted so that the colours are a close match. During production, the required adjustment may be noted and run as a batch process.

2.4 Security Guidelines

The security guidelines of the digitised documents before they are uploaded and migrated to eOffice are mentioned in the section. Once the digitised documents are on the eOffice, the security of eOffice as applicable prevails.

Systems and procedures need to be evolved to protect the documents that are digitised and placed in temporary locations and await migration to eOffice. For large digitisation programs that may be widely distributed throughout an organization and where several staff need to add to and modify the collection of digitised records, a system may be used to manage access to the information and to provide an audit of system access and modification. In small scale digitizing implementations, security and access control may be provided through the use of a password protected system by a single operator, with other authorized staff given read-only access. This could be accomplished using the built-in security features of most current computer operating systems.

- Safe custody and handling of the original paper files and documents need to be ensured.
- Responsibility of the documents need to be allocated at the time of document collection is done.
- Access controls should be applied and maintained within a file management system. This includes specifying who can read or alter a document, preventing copying or printing, or setting an expiration date.

2.5 User Training

Training on how to convert the physical files to electronic files and its migration to eOffice needs to be imparted to the section users. This is normally undertaken by the Digitisation Coordinator of the section.

Need for adherence to the following factors need to be emphasized to the users performing the conversion:

- Scanning guidelines.
- Storage formats.
- Security.