eOffice

A DIGITAL WORK PLACE SOLUTION

DIGITISATION FRAMEWORK

About the Document

eOffice is aimed to bring in more transparency, efficiency and accountability in Government transactions leading to increased productivity. One of the key areas of importance in eOffice is to bring in the required transition to electronic mode of working on files and documents. Moving from a manual document, file and paper based functioning to an "electronic" environment requires effective management to implement the transition. This involves digitising and storing existing physical files and records and also in managing the newly created electronic files and records.

The following are the key areas that need to be considered in the Digitisation Framework:

- Approach to be followed for digitising the existing documents and files.
- Identification of files and records and their storage after the digitisation.
- Roles and responsibilities of the various stakeholders involved in the transition process.
- Setting Standard Procedures and Guidelines to be followed.
- Establishing a long term retention policy.

It is important that the approach adopted ensures the integrity of official records in order to provide authentic, complete and accessible information. Decisions need to be made about the mode and process for the transformation, about the equipment needed, and about the technical standards in storage adopted. Quality assurance measures also need to be defined and implemented. Effective storage of physical files is important for retrieval of these files for any reference purposes. The required training need to be imparted so as to have the department officials meet the responsibilities as per the guidelines laid down. The Digitisation Strategy Framework document includes all the above factors covering the decision points, guidelines and principles, storage policies. It is considered as a guiding document for enabling a successful transition to eOffice.

Intended users

The content in this document applies to all user departments/ministries moving to an electronic file and knowledge management system. The various recommendations and guidelines would apply to all files and documents that are used in the government. The document can be used as a comprehensive guide on the steps and the decisions one needs to take before starting the transition.

Feedback and Suggestions may be sent to:

Head (eOffice Project Division)

National Informatics Centre (NIC)

A Block, CGO Complex, Lodhi Road

New Delhi- 110003

Phone: 011-24365515

Email:eoffice@nic.in

Executive Summary

The eOffice product has been designed with the objective of bringing in a simplified, responsive, effective and transparent working of all government offices. Digitisation of files and records play a very important part in the implementation of eOffice and in enabling an effective transition to the electronic mode of functioning.

This document provides a comprehensive framework for making the transition to an electronic office in which the strategy for digitisation of files and records plays a very important part. It includes the decisions to be taken by management, roles and responsibilities, illustrations of the various implementation models and a summary of various policies and standards that need to be considered for effective long term digitisation.

This document is expected to serve as a guide for Ministries and Departments in their transition to eOffice mode.

Abbreviations

- ❖ ADF .... Automatic Document Feeder
- ❖ DC .... Digitisation Coordinator
- ❖ DFA...Draft for Approval
- ❖ DNO ....Department Nodal Officer
- ❖ DPI..Dots Per Inch
- ❖ DSC .... Digital Signature Certificate
- ❖ FR.. Fresh Receipt
- ❖ ICT .... Information Communication Technology
- ❖ MMP ....Mission Mode Project
- ❖ NC .... Nodal Coordinator
- ❖ NeGP... National eGovernance Programme
- ❖ PIC.. Programme Implementation Committee
- ❖ PPI..Pixels Per Inch
- ❖ PSC ....Programme Steering Committee
- ❖ SOPs..Standard Operating Procedures

Table of Contents

1. Introduction to eOffice

The eOffice is a Mission Mode Project under the National eGovernance Programme (NeGP). It is aimed to help conduct the office procedures electronically to achieve a simplified, responsive, effective and transparent working in all government offices. The product is a standard reusable one amenable for replication across governments and local bodies. When implemented across the Government departments, the eOffice is expected to usher in 'A Single Collaborative Workspace for the Government'.

The eOffice product has been designed with the aim of:

- Establishing a paperless environment in government offices.
- Transforming the existing manual, paper-driven processes into electronic workflows.
- Standardise a single product across Government.
- Establish a single authoritative Government directory.
- Develop on an open architecture framework for scalability.

It is expected to facilitate the following:

- Enhance transparency – files can be tracked and their status is known to all at all times.
- Increase accountability – the responsibility of quality and speed of decision making is easier to monitor.
- Assure data security and data integrity.
- Provide a platform for re-inventing and re-engineering the government.
- Promote innovation by releasing staff energy and time from unproductive procedures.
- Transform the government work culture and ethics.
- Promote greater collaboration in the work place and effective knowledge management.

2. Digitisation Framework

In the overall transition to eOffice, the digitisation of documents and files play a key role. Hence establishing a framework for digitisation is critical for the transition to an electronic mode of functioning. The framework helps to build an integrated approach, connecting the areas that are considered essential in preparing the user department/ ministry for making the transition. The framework will include the Standards, Policies, Guidelines and Best Practices that will help any organization intending to implement eOffice.

The Digitisation Framework has been designed to be in sync with the overall eOffice Implementation The overall framework illustrated in the figure below, depicts the areas that include the Governance Structure, Approach and Methodology, Implementation of the Digitisation to the final integration into eOffice. Each area depicted has further been elaborated in the sections which follow:

Framework for Digitisation

2.1 Governance Structure

For the successful implementation of eOffice in any organization, the commitment of the top management is essential. This needs to be combined with the distribution of responsibilities across the organization by designating clear roles and responsibilities. The implementation of eOffice, must be driven and monitored by a well-defined Governance Structure.

The Governance structure and institutional mechanisms considered essential for the eOffice implementation are detailed in Annexure 1. Some of the important functions are highlighted below:

- Finalization of the Digitisation Strategy based on the Framework.
- Identification of documents for Digitisation and their categorisation into those which should be done in-house and those to be Out-sourced: These inputs will go towards the estimation of the volume of documents for Digitisation.
- Digitisation Time Table and Mode of Digitisation (in-house, Outsourced or mixed).
- Provide Project oversight, direction and guidance as needed for the overall implementation of the project.
- Allocation of the required resources for the smooth implementation: Manpower, ICT Infrastructure.
- Notify the required Policies, Guidelines, Best practices and Standard Operating Procedures(SoPs) for eOffice.
- Ensure meeting the Legal requirements of eOffice: Compliance to Open Storage Formats, Prescribed eGovernance Standards , IT Act 2000 and its amendments.
- Ensure that the Standard Scanning Guidelines are followed with Quality Control.
- Ensure that the digitised records are stored in open standard format.
- Adherence to the Policies and Guidelines of the "Digitisation Framework Document".

2.2 Digitisation Approach

The phased approach to be adopted is detailed in the following sections:

- Phase 1: Plan
- Phase 2: Prepare
- Phase 3: Implement

PHASE 1: PLAN	
Step 1	Identification of Department Nodal Officer and Nodal Coordinators as per the Governance Structure at Annexure 1
Step 2	Identification of documents which are candidates for Digitisation
Step 3	Selection of the Data within the identified documents
Step 4	Decision on the phasing or sequencing
Step 5	Current State Assessment of Volume & Efforts
Step 6	Decision on the model of Digitisation: In-house, outsourced, mixed mode

PHASE 2: PREPARE	
Step 7	Preparation of digitisation time table for the department
Step 8	Procurement of resources estimated in Phase 1
Step 9	User education and training

PHASE 3: IMPLEMENT	
Step 10	Implementation as per the : ▶ Digitisation Time Table ▶ Digitisation Process ▶ Guidelines for Scanning & Storage

Phases of Digitisation Process

2.2.1 Phase 1: Plan

This is the first step to be carried out in each unit of the department by the designated Digitisation Coordinator as per the Governance Structure at **Annexure 1.**

i. **Documents**

- Acts, Office Memorandums, Circulars, Annual reports, and Policies.

ii. **Files**

- All the files that are likely to continue beyond six months in the open status should be selected.

- For the files that are likely to be closed within six months, the files should be selected on the following criteria:

  - The files that may end up as permanent records.

  - The files that are involved in issuing or likely to be issued in future important instructions and directions.

  - The files that have gone or likely to go before the Cabinet.

  - The files that are likely to be added to the "Precedent Book".

  - Any other files identified as important for reference by the concerned Digitisation Coordinator.

  - Disposed of files and the files which will be disposed of in the next three to four months time period will not be scanned unless they are candidates for the record room.

iii. **Records**

- The records that are currently referred to in the "Precedent Book".

- The records that have gone or likely to go before the "Cabinet".

- Records that are frequently accessed are identified from the respective record rooms or section registers.

- All the records that may end up as permanent records.

- Any other records identified as important for reference by the Digitisation Coordinator concerned.

Step 3: Selection of the data that need to be digitised within the selected documents.

It is important to ensure that the digitisation activity is undertaken in a meaningful manner and not merely as a scanning exercise. It is not necessary to digitise all the pages or the content in the identified files and records. There are several pages in a file that will not have significance for future reference. Therefore it is necessary to identify only those pages which have value for future references and these need to be digitised. Papers in Closed or Recorded files need to be selected more selectively under supervision.

The following are not recommended for Digitisation:

- Any extra copies of Government Orders / Memorandums /D.O letters issued in the file.
- Drafts for Approval/Papers under consideration where the final decision and issue has been generated.
- Copies of vouchers, tickets which have been billed already.
- Any correspondence/noting/issue for which the soft copy is available.

Apply the Record Retention Schedule and Policies for old and legacy files to minimise digitisation efforts.

The final decisions will need to be taken by the designated Digitisation Coordinator in consultation with the Unit.

Step 4: Decision on the Phasing and sequencing manner to be adopted.

It is important to decide the phasing digitisation of the identified documents, files and records, especially when there is a large number and volume. The outcome of this decision will become an input for the Step 8 of Phase 2: Digitisation Timetable.

The following approaches are recommended to select the files to be digitised and in which order:

- Digitise all files which are not in circulation.
- Closed and Recorded files from the current to the older could be digitised first.
- Closed Files that are referenced in the Active or Current Files.

Active files should be done as per the transition decision (Refer Annexure 2). The active files belonging to the category decided to be taken up for converting to electronic mode in the first phase of the Transition Time Table. This can be done by the Sections themselves and do not need any advance preparation.

Step 5: Current State Assessment of Volume & Efforts

Volume Estimation: Based on the selection done in Steps 1 and 2, the expected volume of documents is calculated.

Efforts Estimation: This is done based on the following activities:

- Pre-Processing activities

- Scanning

- Entry of metadata which must have keywords of the document content and keywords for searching.

- Storage of the digitised documents

- Uploading to eOffice

Step 6: Decision on the model of Digitisation: In-house, outsourced, mixed mode

The model of digitisation is the next step to be decided. The recommended approach is as follows:

(i) Active and Current files

All active and current files will not be converted to electronic file from day one. The active/current files that will be taken up for transition to electronic file will depend on the Transition Approach adopted and finalized by the organization and the finalised Transition Time Table.

The Digitisation activity of Active/Current files is normally best under taken by the Section concerned.

(ii) Record room and Legacy Files: Out- sourced to external empanelled agencies who can take up the activity.

Scanners

Prior to implementation of the digitisation process, there should be a high level of understanding within the organization of the technical aspects of scanning. There are many types of scanners available in the market with different features and most models include the in-built software or drivers for Optical Character recognition (OCR), image manipulation, support for multiple formats. It is recommended to choose the scanner based on the requirements.

A scanner can be chosen depending on the various parameters such as:

- Volume of documents to be scanned.
- Quality of documents and subsequent scanned images.
- Speed of scanning required.
- Type of documents (file, book, letter, etc.).

Parameter	Description	Type
Volume of Documents	High	High end scanner that can scan up to 100 pages/ minute
Type	• Good Quality • Laminated documents • Books	• Automatic Document Feeder (ADF) scanner • Flat Bed Scanner • Book Scanner

2.2.2   Phase 2: Prepare

Step 7: Preparation of digitisation time table for the department

The Digitisation Time Table provides the detailed calendar of digitisation activity that need to be carried out, both for in-house as well as outsourced.

Step 8: Procurement of resources estimated in Phase 1

The resources can be procured following due procedures.

Step 9: User education and training
- Training for Doc identification, indexing, Preparing metadata, etc. at first stage
- Implementation and Execution Plan (scanning) with Handholding support
- Continued Training for different activities

2.2.3   Phase 3: Implement

The implementation follows the steps and methodology outlined in the following Section 4.3 – Digitisation Process.

Step 10: Implementation as per the suggested points.

- Digitisation Time Table
- Digitisation Process
- Guidelines for Scanning, storage

3. Digitisation Timetable

Serial	Actions	Activities	By whom	Timeline
1	Phase 1-Plan	• Identification of Department Nodal Officer, Nodal Coordinators, and Digitisation Coordinator as per the Governance Structure at Annexure1 • Identification of Documents that are candidates for Digitisation • Selection of the data that need to be digitised within the selected documents • Decision on the Phasing and sequencing manner to be adopted • Current State Assessment of Volume & Efforts • Decision on the model of Digitisation: In-house, outsourced, mixed mode • Resource requirement	Project Steering Committee, Project Implementation Committee, Departmental Nodal Officer, Nodal Coordinators, Digitisation Coordinator	To be decided by the Concerned Department
2	Phase 2- Prepare	• Preparation of digitisation time table for the department • Procurement of resources estimated in Phase 1 • User education and training	Project Implementation Committee, Departmental Nodal Officer, Nodal Coordinators, Digitisation Coordinator	To be decided by the Concerned Department

3	Phase 3- Implement	• Implementation as per the suggested points
		  o Digitisation Time Table
		  o Digitisation Process
		  o Guidelines for Scanning, storage
			Departmental Nodal Officer, Nodal Coordinators, Digitisation Coordinator	To be decided by the Concerned Department

3.1 Digitisation Process

DIGITISATION PROCESS

START
Collect & Prepare
Scanning
Processing & QC
Indexing
Final QC & Security

HOW TO

Collect and prepare the documents.

Mode for Scanning.

Scanning specifications & format.

Storage of the scanned documents.

Post-Scanning.

PRE-REQUISITES

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 shall be bound in a file.

The physical file should have a unique Identifier affixed on top to help faster retrieval.

Digitisation Process comprises of following steps:

3.1.1 Collect and prepare the documents.

3.1.2 Mode for Scanning.

3.1.3 Scanning specifications & format.

3.1.4 Storage of the scanned documents.

3.1.5 Post-Scanning.

Each step is described in the following section.

3.1.1 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.

3.1.2 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.
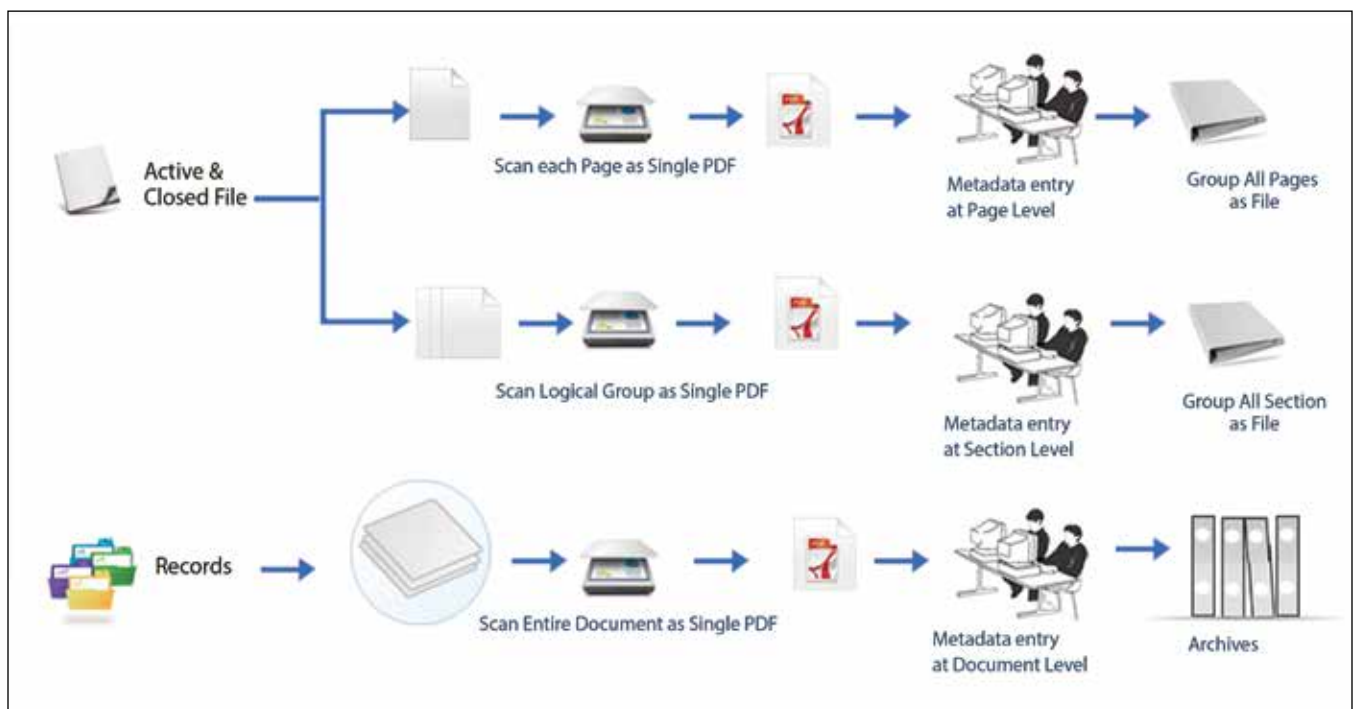
Mode of Scanning

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

3.1.3 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.
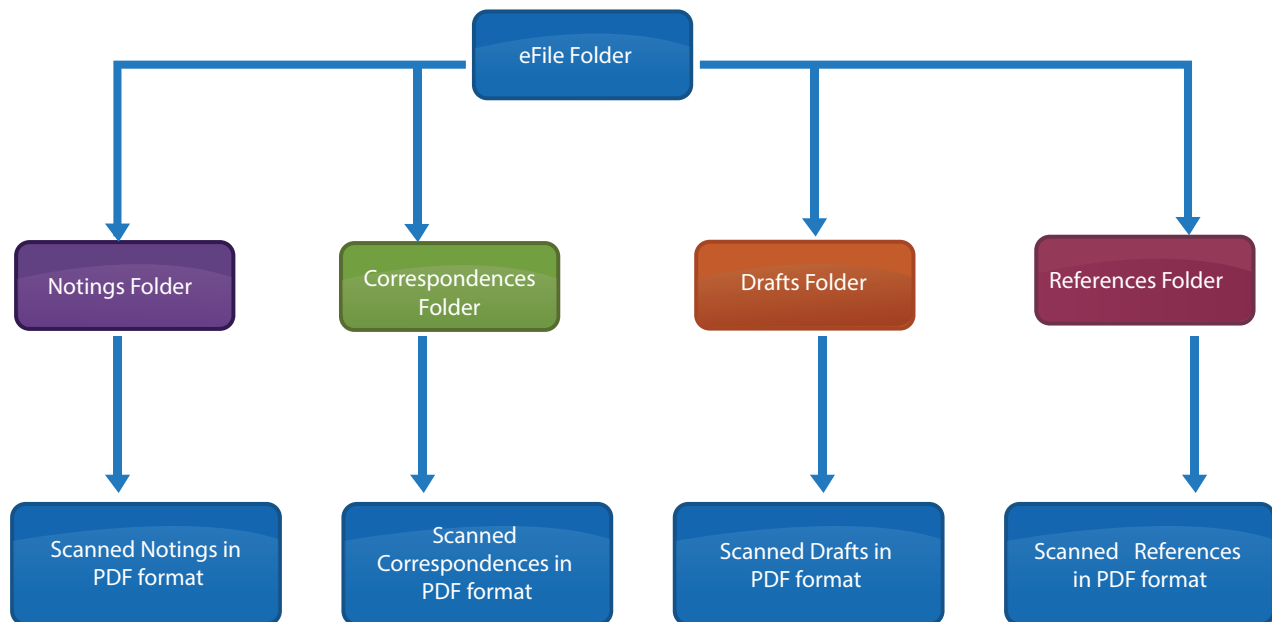
3.1.4 Entry of the Metadata
The required metadata is entered against each document. The indicative metadata is as follows:

Active Files	Closed Files	Recorded files
- File no. - Subject Description - Subject Category - Subject Sub Category - Language - Remarks - Previous Reference - Next Reference - Created On	- File no. - Subject Description - Subject Category - Subject Sub Category - Language - Remarks - Previous Reference - Next Reference - Created On - Closed by - Closed On - Closing Remarks	- File no. - Subject Description - Subject Category - Subject Sub Category - Language - Remarks - Previous Reference - Next Reference - Created On - Closed by - Closed On - Closing Remarks - Recorded By - Recorded On - Recording Remarks

3.1.5 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.

```
                          ┌──────────────┐
                          │ eFile Folder │
                          └──────────────┘
        ┌──────────────┬───────┴───────┬──────────────┐
        ▼              ▼               ▼              ▼
┌──────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────┐
│Notings Folder│ │Correspondences│ │Drafts Folder │ │References    │
│              │ │Folder        │ │              │ │Folder        │
└──────────────┘ └──────────────┘ └──────────────┘ └──────────────┘
        ▼              ▼               ▼              ▼
┌──────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────┐
│Scanned       │ │Scanned       │ │Scanned Drafts│ │Scanned       │
│Notings in    │ │Correspondences│ │in PDF format │ │References in │
│PDF format    │ │in PDF format │ │              │ │PDF format    │
└──────────────┘ └──────────────┘ └──────────────┘ └──────────────┘
```

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 labeling 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 labeling 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.

3.1.6 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.

When implementing systems, departments/ministries must therefore take special care to ensure they are secure, reliable and capable of producing files that are acceptable for legal, audit and other purposes.

4. 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.

4.1 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.

4.2 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 labeling 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.

4.3 Guidelines for Maintaining Quality Control
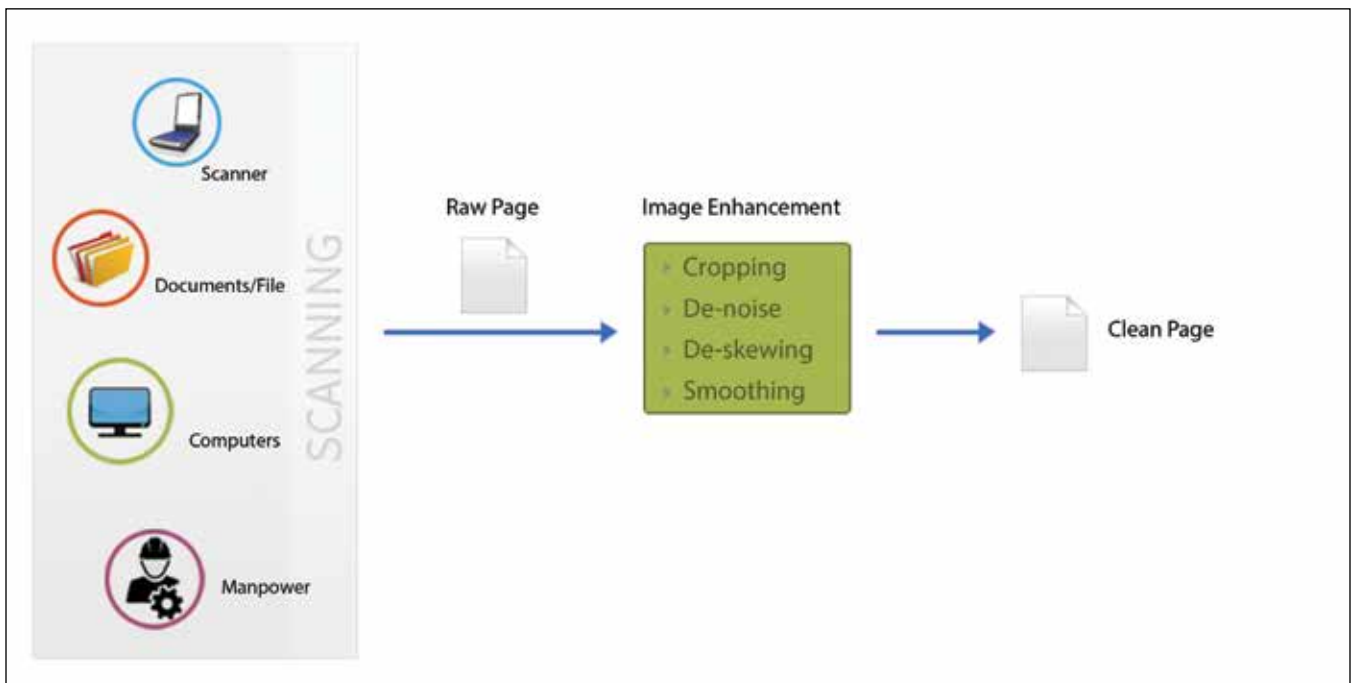
Quality Control Process

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.

4.4 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.

4.5 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.

6.   Integration into eOffice

For Integration with eOffice, digitisation strategy has to be followed with migration as follows:

Guidelines for Migration Process :

Step 1. Create main folder with the name of the "Folder" same as that of the "Physical File" in the local system.

Step 2. Create sub folders inside the main folder for Notes, Receipts, Issues, and References (Fig. Folder Structure for Migration).

Step 3: All files (Active/Closed/Legacy) are scanned along with their respective Notes, Receipts, Issues, References and stored in separate folder/sub folders (Fig. Folder Structure for Migration).

Folder structure for storing the scanned notes, receipts, issues, and references is depicted in diagram below:
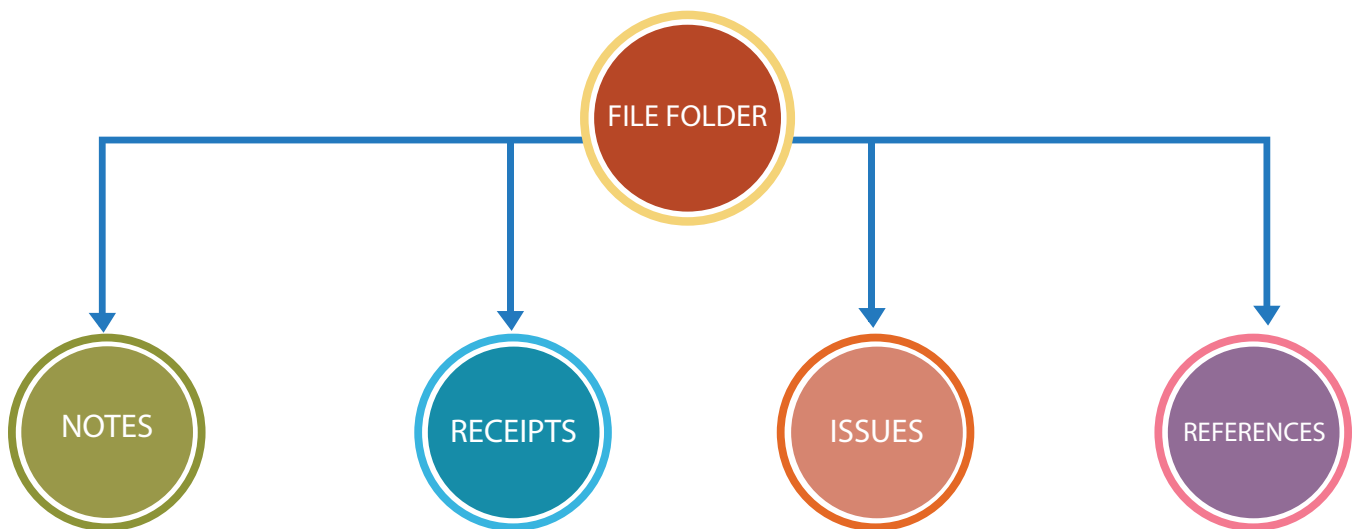
Folder Structure for Migration

Example: If the Physical file number is A-21001(3)/5/2008/DFG, then user has to create the main folder along with their respective sub folders as depicted in the **Table 1**:

Main Folder Name	Sub Folder Names
A-21001(3)_5_2008_DFG (Name of the folder should be same as that of the Physical file number, following certain conventions). (For Conventions, ReferTable 2)	For notes, Notes For receipts, Receipts For issues, Issues For references, References

<div align="center">

Table 1

</div>

For migration, user has to take care of certain set of Conventions, Assumptions, and Special Conditions as enlisted in the **Table 2**:

Conventions	Assumptions	Special Conditions
If any "/" (Backslash) is present in the physical file number, user will replace it by "_" (Underscore) in the folder name. • / • \ • * • : • " • | • < • > • ?	Physical File number contains the following characters: • a-z • A-Z • 0-9 • () • – (hyphen)	**File can be migrated only if all the DFAs in the file are issued.**

<div align="center">

Table 2

</div>

Step 4: Scanning Process:

All Notes, Correspondences (Receipts/Issues), References are scanned and saved in pdf format in FTP client.

- Notes are scanned into a single pdf, if possible. Otherwise notes can be scanned one by one.
- Correspondences (Receipts/Issues) should be scanned, one at a time, as a single pdf.
- References should be scanned one at a time, as a single pdf.
- Issues should be scanned one at a time, as a single pdf.

Step 5: Uploading scanned files from FTP Client to the eOffice Server:

eOffice Migration Module facilitates to upload the files from FTP client to the eOffice server.

Important Note and Points to be taken care:

The verification of the documents in a file which have already been scanned need to be done to ensure the following:

- The scanned document is identical in content to the original and not tampered with.
- There is no missing scanned document.
- Care should be taken to ensure the completeness of the scanned documents when
- compared to the original physical document in terms of the content and number of
- pages.

The Roles and Responsibilities in the overall eOffice Governance Structure to ensure the above are as follows (See Annexure 1 for more details):

Nodal Coordinator
- Should be of the rank of Deputy Secretary or equivalent.
- Should do the final verification of the digitised documents.

eOffice Digitisation Coordinator
- Should be of the rank of Under Secretary or equivalent.
- Supervision of the scanning, digitisation, and integration activity.
- Verify that the documents in files/records are not tampered, misplaced, removed during the process of scanning.

Verify the completeness of the scanned file with the physical file.

Annexure 1: eOffice Governance Structure

To ensure an effective implementation of eOffice, it must be driven and monitored by a well-defined Governance Structure. The following four key institutional key mechanisms as outlined in Section 1 below, are recommended as a part of the eOffice Governance Structure. The recommended supporting structures are detailed in Section 2.

1. Key institutional key mechanisms

- Project Steering Committee
- Department Nodal Officer
- Nodal Coordinators
- Project Implementation Committee

Let's discuss about Key institutional key mechanisms one by one.

Project Steering Committee

It is the Project Steering Committee (PSC) that gives the overall direction and leadership to the implementation of eOffice. This committee, being advisory in nature, must be formed in a manner wherein all implementing divisions and key stakeholders are involved. The leadership is to be entrusted to the top authority.

The suggested constitution of the PSC is as follows:

- **Secretary of the Line Ministry/Department– Chairman**
- **Nodal Officer (Joint Secretary level) –Member Secretary**
- Heads of the Units(Bureau/Division/Wing/Section/Desk)
- Director General NIC
- Designated DARPG Official
- NIC HoD of the Ministry/Department

The key tasks and responsibilities of PSC should be as follows:

- Setting the scope , long term goals & targets.
- Provide Project oversight, direction and guidance as needed for the overall implementation of the project: Transition to eOffice in the defined time frame, ensuring the absorption and sustainability of the project.
- Budget allocation and approvals.

- Change Management & Process Re-engineering, wherever required.
- Allocation of the required resources for the smooth implementation: Manpower, ICT Infrastructure.
- Decision of the model and approach for transition to eOffice: Transition Strategy & Time table.
- Finalization of the Digitisation Strategy Framework.
- Digitisation Time Table and Mode of Digitisation (In-house, Outsourced or Mixed).
- Designate the Department Nodal Officer for the overall coordination, supervision and monitoring of the implementation and reporting the deliverables.
- Approval and Issue of the notifications required.
- Establish and Notify the required Policies , Guidelines ,Best practices and Standard Operating Procedures for eOffice.
- Communication of the project to all stakeholders (both internal as well as external to the organization).
- Ensure meeting the Legal requirements of eOffice: Compliance to Open Storage Formats, Prescribed eGovernance Standards , IT Act 2000 and its amendment.
- Promote a Participative strategy in the implementation so that the responsibility and ownership is shared across the organization (A System driven rather than a Person driven approach is to be enabled to ensure the sustainability).
- Regular Review of project deliverables & enforcement for smooth implementation.
- Every 15 days for the first three months.
- Every 30 days for three to six months.
- Every 45 days beyond the first six months.

Departmental Nodal Officer (DNO)

Department Nodal Officer (DNO) is the second recommended institutional mechanism for implementing the eOffice programme, which shall be the focal point for the overall steering and required coordination of implementation efforts in consultation with the Steering Committee. It is suggested that the DNO, who shall be the key person for the implementation and sustenance is not less than the rank of Joint Secretary so that the eOffice implementation gets the management attention it deserves.

The key tasks and responsibilities of DNO are mentioned below:
- Responsible for the transition and implementation of eOffice as per the Project Plan.
- Responsible for meeting the target deliverables and time lines laid out in the Transition & Digitisation Time Table, on behalf of the Project Steering Committee(PSC).

- Act as the reference point for all project implementation tasks.
- Set priorities, identify the Nodal Coordinators from each Unit (Bureau/Division/Wing/ Section/Desk).
- Will keep the PSC informed of the progress of the project.
- Notification of the appropriate support structures with the approval of the PSC.

Nodal Coordinators (NC)

For the day- to- day coordination and implementation of eOffice smoothly, representation from all the Units of the Ministry/Department are needed. Nodal Coordinators (NC) from each Unit are to be identified and designated. The key tasks and responsibilities of the NCs with respect to their individual unit are mentioned below:

- Should be of the rank of Deputy Secretary or equivalent.
- Should do the final verification of the digitised documents.
- Steering and coordination of implementation in their respective Unit.
- Provide the necessary assistance in infrastructure assessment, data collection and its vetting.
- Ensure DSC forms are collected and submitted on time as per the guidelines for completing the DSC forms.
- Ensure that NIC eMail accounts are created for all users in their Units.
- Identify the Content Manager, Digitisation Coordinator, Master trainer in their respective Unit and inform the PIC for necessary notifications with of the roles and responsibilities. (**Note:** These roles can be performed by one or many depending on the size of the Unit).
- Standardization Activities.
- Identify the File Heads in the Unit as per the Standard File Heads.
- Standardization of the Templates being used in the Unit (Letter, Forms, Sanction Orders).
- Transition Approach.
- Identify the categories and subjects that will be moved to eFile in a phased manner.
- Draw up the Transition Time Table for their respective Unit to place to the PIC for the finalization of the Department level Transition Time Table.
- Building the Knowledge Management System (KMS) (Support from Content Manager).
- Folder structure for the Unit in KMS.
- Identify the Documents in the Unit that need to be digitised for uploading in the appropriate folder of the KMS.
- Ensure that the Documents for the Central document and Knowledge repository are uploaded by the Content Managers.

- Digitisation (support of Digitisation Coordinator).
- Identification documents (See Annexure 2 & 3 for details) for Digitisation.
- Categorize the documents into those which should be done in-house and those to be outsourced: These inputs will go towards the estimation of the volume of documents for Digitisation.

Projejct Implementation Committee (PIC)

The Project Implementation Committee (PIC) is the fourth important institutional mechanism to give the steering support for the implementation of eOffice. This committee under the chairmanship of the Department Nodal Officer must be formed in a manner wherein there is representation from all Units involved.

The suggested constitution of the PIC is as follows:
- Designated Nodal Officer (Joint Secretary level) -– Chairman
- Nodal Coordinators of each Unit
- Digitisation Coordinator
- NIC HoD of the Ministry/Department
- eOffice Project Manager coordinating the implementation

The key tasks and responsibilities of PIC should be as follows:
- Coordination and Monitoring of implementation.
- Finalization the designated roles in the Support Mechanism that are identified by the Nodal Coordinators of each Unit and placing to the PIC for approval.
- Incentive & Rewards System.
- Draw up the Training and Capability Schedule.
- Inter-Unit coordination.
- Facilitating and ensuring that gaps in infrastructure are met by the department as per the Gap Analysis Report.
- Ensure the availability of the Data from the individual units for eOffice environment.
- Ensure the Time bound transition by all departments to eOffice as per the Transition Time Table.
- Preparation of the Digitisation Time Table (Inputs from Nodal Coordinators of the Unit).
- Analysing the ministries/department's digitisation requirements based on the assessment.

- Who will undertake (in-house for current/outsource to agency).
- Estimate the budget and resources for a sustainable organization-wide digitisation strategy.
- Mechanism for Storage of the records.
- Identify the agency to undertake the digitisation activity (Outsourced).
- Regular review of project deliverables and enforcement for smooth implementation.
  - Every 7 days for the first three months.
  - Every 15 days for three to six months.
  - Every 30 days after the first six months.

2. Support Mechanisms

In addition to the key institutional mechanisms outlined in the above section, the following support structures are considered important for a focused implementation. The roles defined in the support mechanisms along with the responsibilities need to be designated and notified by the department.

- eOffice Administrator
- eOffice Master Trainers in each Unit for sustainability
- eOffice Content Managers
- eOffice Digitisation Coordinators
- Digitisation Implementing Agency
- eOffice Service Book Nodal Coordinator

Lets discuss different roles and their responsibilities one by one

eOffice Administrator

- Overall responsibility for subsequent administration & management of eOffice Data under coordination of NIC Division of concerned Ministry/Department.
- Assigning roles & privileges.
- Managing top level directory structure (department wise) in KMS.
- Managing Inter-departmental transfers, postings.
- Ensure the updation of employee data.

eOffice Master Trainers

- Identification to be done in the initial stage of the project.
- Will undergo initial training and capability in eOffice from the eOffice Implementation team.
- Should have the aptitude for conducting trainings.

- Will assist the eOffice implementation team from start to provide trainings.
- Will be instrumental in providing continuous trainings to the department.

eOffice Content Managers

- To upload the content into the relevant folder of KMS.
- Ensure that the latest version is updated in KMS.
- Provide the privileges at the Folder and Document level.

eOffice Digitisation Coordinators

- Should be of the rank of Under Secretary or equivalent.
- Identify the documents in the Unit.
- Categorise the documents into two main categories.
  - Those that can be done in-house (Current and Active).
  - Those that need to be outsourced (Normally the legacy records. This will be as per the Digitisation Decision of the PSC).
- Assess and Estimate the volume of documents.
- Assess the type of scanner required in their respective unit and the number of scanners.
- Required training and awareness to users in the Unit.
- Ensure that the Standard Scanning Guidelines are followed with Quality Control.
- Time bound digitisation activity (both of in-house and the outsourced activity).
- Will report to the DNO on the progress in their respective Unit as per the Digitisation Timetable.
- Supervision of the scanning, digitisation, and integration activity.
- Verify that the documents in files/records are not tampered, misplaced, removed during the process of scanning.
- Verify the completeness of the scanned file with the physical file.

Implementing Agency (External Agency)

- Managing all digital files/records identified by the organization.
- Creating metadata about digital files/records.
- Storing the digitised files and records in standard folder structure till the migration to eOffice takes place.
- Taking effective security measures to ensure that the digital files/records are safe from damage and tampering.

- Ensure that the digitised records are stored in open standard format.
- Managing digital files/records in accordance with their specialized document management requirements.
- Adherence to the Policies and Guidelines of the "Digitisation Framework Document".

eOffice eService Book Coordinator
- Estimate the records as a part of eService Book Digitisation.
- Decide the cut-off date for manual entry in eService Book.
- Estimate the efforts for digitizing the Service records and entering into eOffice.
- **Plan for resources :** Manpower and ICT infrastructure.

The above Governance Structure (Key and Support Mechanism) are to ensure that the implementation of the eOffice is carried out in a structure manner across the organization in a manner that makes each user in the department accountable and responsible for implementation.

Annexure 2: Transition Strategy

a. What is the Transition Strategy?

This encompasses the decisions taken by the department under the PI&MC with regard to the transition to the electronic mode of working, particularly with respect to the electronic files.

In this step the department has to decide on the approach to be adopted for the transition. It has two broad options in this regard:

- Complete transformational (big-bang) approach by going for a completely electronic mode of working from the start of eOffice implementation.
- Custom approach by phasing the transition. Department will have to decide the number of divisions that will be covered and the number of phases for doing so. The department could either go for all divisions or selected sections in each division and the sequencing of implementation.

The Strategy decisions are to be based on several factors: Readiness of the department in terms of infrastructure, availability of resources, employee readiness, etc. The department will also have to identify files for conversion from physical mode to electronic mode in each division. It will also have to prepare a timeline for the phased conversion and coverage of files in each phase.

b. Why is it required?

A sound strategy plan makes the induction and transition to eOffice smoother and more seamless. The experience gathered from the implementations indicates that some amount of phasing is inevitable even in case of a complete transformational model.

c. How to draw up a Transition Strategy and Time Table?

The first step is the decision on the Transition Strategy that the department will adopt:

Transformational approach- All divisions and all subjects to be covered under eFile implementation

- All inward receipts will be scanned. There will be no movement after the first point of receipt entry and diarisation.
- All files in all the sections to move as e-Files end-to- end.
- No manual file will be created nor move internally.
- File movement should be to up to the level of Minister.

Custom Approach-Selected divisions /subjects to be moved as eFiles.

To arrive at the strategy and time table for the transition, some questions that need to be considered are as follows:

- Which files need to be converted from physical mode to electronic?
    - o (Select subjects/categories in each section; prepare a time plan for the phased conversion).
- What will be the timeline for coverage of files in each phase? Time table for conversion?
- Does the transition have to happen at one go or can it be done in phased manner?
- When is a file termed as completely electronic (Complete migration)?
- Is there a separate process of conversion for active/closed/legacy files?
- Should all the papers in the active/closed and legacy files to be scanned?
- How is the phased approach to be planned?
    - What are the parameters to be considered while making the decision on the strategy and approach?
- What sort of hardware and software will be required?
- Who will coordinate the transition implementation?
- Who will do the conversion of the physical files to Electronic mode (Department / Outsource to agency)?
- What happens to the papers/files after the conversion?

Complete Transformational Approach
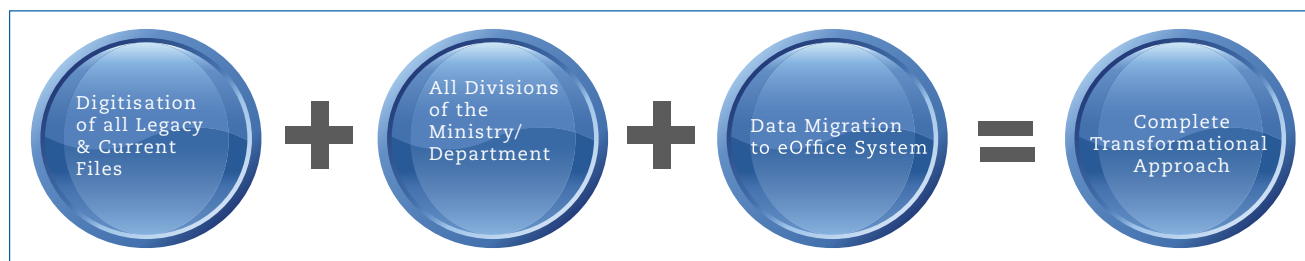
This model represents the transition approach in which the department decides to make a complete transition to an electronic mode of working of files and documents from the start. This approach involves the digitisation of all categories of the files (active, closed and legacy) of all divisions of the department and migration of the digitised files into eOffice. There is no physical file movement in case of adoption of this model by the department.

Complete Transformational approach	Divisions	File Categories
1	All Divisions	All Categories

Phased Transformational Approach
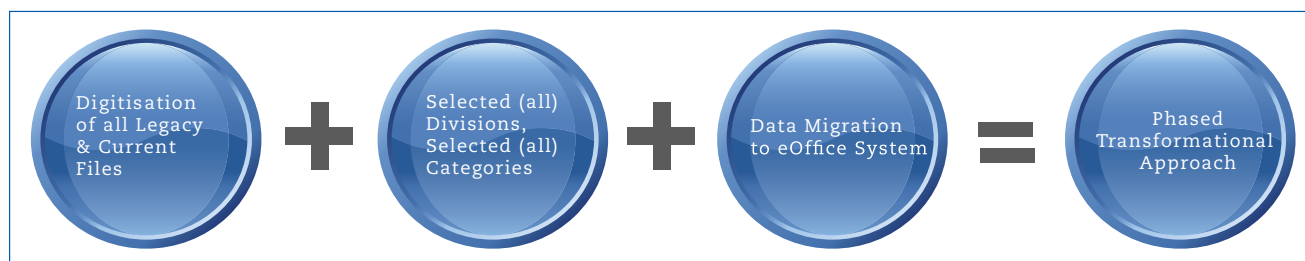
This model is one in which the transition of files is done in a phased manner. The department may choose selected divisions only to begin with, or may further choose to have only selected categories of files within a division to be taken up for transition to electronic mode.

Source : eOffice Implementation Handbook Version 1.0

Transition Models and Approach for Digitisation

The final outcome is the drawing up of a Transition Time Table through a consultative approach. Final transition need to be monitored through the transition time table of each department.

A Comparative Analysis of the different Transition Approaches

Implementation Model	Resource requirement	Turnaround time for Transition	Management Commitment
Transformational Approach	High	Low	High
Hybrid Approach	Moderate	Moderate	High
Custom Approach 1 (Selected division and transformational approach)	Moderate	Moderate	High
Custom Approach 2 (Selected division and hybrid approach)	Moderate	High	High

Annexure 3: Types of Documents

Receipt

eFile in most cases is created with a receipt after it has been diarized. A receipt is an incoming communication and may be any one of the following:

- Letter
- E-mail
- Fax
- Video
- Images
- Others

The receipts are first received by the Section Officer (SO), who scrutinizes and assigns the receipts to the Dealing Hand (DH) depending on the subject of the receipt. The DH eventually creates the eFile.

A receipt, once received, is marked/ indexed in the following ways:

- If the receipt initiates the creation of the eFile it is marked as Paper Under Consideration (PUC 1) automatically by the system.

- If the receipt does not initiate the creation of an eFile, but is linked to an existing PUC, the receipt is marked as Fresh Receipt (FR) for the corresponding PUC.

- If the receipt does not initiate the creation of an eFile and the receipt is not linked to an existing PUC, then it is indexed as PUC.

Serial numbers are generated by the system for all correspondence whether it is a receipt or an issue. A receipt is marked as 'R' and an issue is marked as 'I'. Page numbers are also generated serially. A receipt that initiates an eFile is assigned a new serial number, which will be displayed on the top right hand corner of the first page of the receipt, in bold and in red colour. All subsequent receipts will be numbered serially by the system. All receipts relevant to PUC 1 will be numbered FR 1, FR 2 and so on. All PUCs and FRs marked in the correspondence section are referenced to notes in the notes section. When the PUC 1 is closed all FRs linked to that PUC will also be closed. All the receipts and issues will reside in the Correspondence section of the eFile and the pages will be serially numbered in common series. A letter issued from an eFile will have the eFile number and the letter number along with a serial number suffixed with it.

During indexing, all receipts constituting a file may be indexed or the user can select a batch of receipts or even a specific receipt for indexing.

Files

Current files – Current files are those which are currently in circulation/ in use. They deal with cases that are open and decisions on the issues being considered in them are yet to be finalized. An active file may contain many documents or a few documents depending on its use.

Closed files – Files are marked as closed/recorded once all actions are complete on the file and the Section Officer /Designated Authority gives approval for closing the file. Closed files have no outstanding fair communications to be issued and no further action due.

Notes

Notes and orders are recorded in the notes section of the eFile. The system generates serially numbered paragraphs and page numbers in the notes section. All receipts (R) on which the note is being prepared are docketed above the note and all issues are docketed below the note in Red, that culminated in that issue (I).

When the DH puts up the case to a higher officer, he references all relevant correspondences and drafts and mentions the approving authority. The system affixes name, designation, date & time next to the note drafted by the user.

All referencing in the notes section is done using the PUC and DFA numbers in the correspondence and DFA sections respectively.

The user can create an appendix to notes with referencing to the notes section.

Draft for Approval (DFA)

A DFA is created in response to a PUC or FR. A DFA is identified by a system generated draft number such as DFA 1, DFA 2, etc. The DH fills the name, designation, telephone no., fax, email id of the issuing officer in the DFA and uploads any other enclosures (letters, communications, and annexures) relevant to the draft. The system links the enclosures to the draft by mentioning the enclosure no. on the bottom left corner of the first page of the DFA. The DFA is referenced to the notes section.

For each edition made in the DFA ,a different version of the DFA is stored with the digital signature of the officer. Before the DFA is signed by the issuing authority the DH fair types the DFA. The fair DFA is indexed with a system generated letter number which is the same as the eFile number with serial number affixed. The fair communication is then moved to the correspondence section and

is marked as an issue ('I') with the next serial number and page number. Once signed, the Issue is docketed in the notes section and the system records the serial number (I), date of issue, details of the addressee and the page number in red. For example, Serial Number 12 (Issue) dt 30.06.2008 to all State Govts./MHA/Min of Finance (page No. 38-40).

Records

Records are those files that have been archived after completion of their retention period. These files have usually been closed for years and rarely used for certain reference purposes. Once all required actions are complete and no further actions are expected, the DH closes the eFile after approval from the Section Officer. The DH enters the retention period and the system records the name of the user who closes the file along with the date & time of closing. The eFile is marked as closed and moved to the eRecord room.

Annexure 4: Comparative Analysis of Storage Formats

Parameter	TIFF	PDF (ISO 32000-1)
Open and ISO Standard	• Not an ISO standard. • Privately held/controlled specification • Platform independent	• PDF: ISO 32000-1 standard • PDF-A: ISO 19005-1: 2005 standard • Platform independent
Reliable File Format	• Numerous extensions to TIFF by both Adobe and 3rd parties make the file format incompatible and unreliable for consistent rendering of the content.	• PDF Specification updated regularly by ISO. • PDF-A exists for long term archiving of electronic data and consistency of content rendering.
File Size	• Document size is limited by 4GB. • Older compression schemes like CCITT G3, G4, JPEG supported.	• No inherent limitation on document size. • Modern compression schemes like JBIG2, JP2K supported. • MRC compression possible
Longevity	• TIFF viewers will be available for some time to come	• For long term standards archiving, PDF-A is the best choice.
Searchability and Indexing	• Only images can be stored. • Not inherently searchable	• Both text and image can be stored. • Inherently searchable. • Textual content can be accessed.
Security and Integrity of Documents	• No security model.	• Sophisticated security system • Different kinds of security ensure granular levels of control over the document. • Digital Signatures provide integrity of documents.
Document Structure and Navigation	• Only Page sequencing • No bookmark, hyperlink etc	• Complete document structure is present • Bookmark, hyperlink, annotation gives good user experience • Mimics the real world document.
Support for Accessibility	• No support for people with disabilities.	• Document structure enables effective use of assistive technologies such as screen readers.
Quality of Presentation	• Image quality compromised at higher resolution.	• By virtue of storing fonts, quality of view and print is maintained at higher resolution.

Suitability for Web Delivery	• No support of dynamic delivery over web	• Linearized PDF is optimized for faster Web View.
Rich Meta Data	• TIFF tags used to store simple metadata.	• Conforms to the XMP specification.
Multiple Colour Support	• Only one colour space per page	• PDF page can have both B/W and colour image data
Associating similar documents	• Multiple files cannot be assembled into an integrated unit.	• A PDF Portfolio contains multiple files assembled into an integrated PDF. • Documents which cater to similar things can be kept together.

Source: Whitepaper: Adobe Systems - TIFF versus PDF for Document Storage and Archival

TIFF

The TIFF format was created by Aldus Corporation in the mid-1980s in order to create a standard file format for storage of scanned images. The TIFF specification [1] is now controlled by Adobe although no major update to the specification has taken place since 1992.

The TIFF format itself is a file format, not an image format. A TIFF file can be thought of as a container for one or more images each of which may be of a different type.

The most common types of image included in TIFF files are as follows:

Group4: Most commonly used for black and white ("bitonal") scanned images.

Group3: Used for Faxes.

JPEG: Used for Grayscale and Colour scanned documents. There are two definitions of JPEG in TIFF, type 6 and type 7 and there have been interpretation issues with type 6 in particular. Consequently these images are not always well supported.

LZW: Used for Grayscale and Colour scanned documents. Due to historical patent issues this is not always well supported, although as the patents expired in 2004 more recent software should provide good support.

Uncompressed: Often output from graphics applications.

Others: Other less commonly used schemes include ZIP, Pack bits, and RLE.

PDF

The PDF format was created by Adobe in 1993 for the purpose of portable document exchange between systems and applications. The PDF specification has been updated and published by Adobe over the years but PDF is now an open standard that was officially published on July 1, 2008 by the ISO as ISO 32000-1:2008.

A PDF file encapsulates a complete description of a fixed-layout document that includes the text, fonts, raster images, and vector graphics which the document comprises. It includes support for JPEG, JPEG 2000, JBIG2, Group 3 and Group 4 images. A PDF file can be classified as:

Normal PDF: This is the most common type of PDF and is most typically created from a document such as Microsoft Word. It contains the full text of the page with appropriate coding to define fonts, sizes, etc. and will provide a faithful print of the original.

Image Only: This is a PDF that has been created from one or more images – most commonly as a result of scanning a document either directly to PDF or by converting a scanned TIFF image to PDF. These files do not contain any searchable text and most often comprise a set of Group4 or JBIG2 images in a PDF "wrapper".

Searchable PDF: A "Searchable" PDF is an "Image-Only" PDF that additionally contains a hidden layer of text generated by an OCR engine. This enables the file to be searched in the same fashion as a "Normal" PDF. Text can be copied and pasted.

PDF-A: It is a file format for the long-term archiving of electronic documents. It is based on the PDF Reference Version 1.4 and is defined by ISO 19005-1:2005, an ISO Standard that was published on October 1, 2005.

It is a subset of PDF, obtained by leaving out PDF features not suited to long-term archiving. It eliminates certain risks threatening the one-to-one future reproducibility of the content. PDF-A forbids dynamic content to ensure that the user sees the exact same content both today and for years to come. Everything that is required to render the document the exact same way, every time, is contained in the PDF-A file: fonts, colour profiles, images etc. PDF-A is also an ISO standard, guaranteeing that future software generations will know how to open and render PDF-A files.

TIFF has proven to be a robust file format for document storage over the last 20 years and still has a role to play in document storage. Increasingly, however, business needs dictate that the greater functional capabilities of the PDF format are required. The ISO standardization of the PDF format has put to rest any qualms about the long term openness of the format.

Some specific advantages of a PDF/A archive over a TIFF or a paper-based archive are:

- **Full-Text Search -** PDF/A stores objects (e.g. text, graphics), allowing for an efficient full-text search in an entire archive. TIFF is a raster format and must first be scanned with an OCR (optical character recognition) engine.

- **File Size -** PDF/A files require only a fraction of the memory space of original or TIFF files, without loss of quality. The smaller file size is especially advantageous by electronic file transfers (FTP, e-mail attachment etc.).

- **Optimization -** PDF/A format can be optimized. The optimization can be focused on images (e.g. scanned checks) or extracting structured data (e.g. voucher information). TIFF treats all file information the same.

- **Metadata -** Metadata like title, author, creation date, modification date, subject, keywords, etc. can be stored in a PDF/A file. PDF/A files can be automatically classified based on the metadata, without requiring human intervention.

Strength & Weakness – File Formats

Name and Current Version	TIFF 6.0	GIF 89a	JPEG JFIF	PNG 1.2	PDF 1.4
Extension	.tiff	.gif	.jpg	.png	.pdf
Bit-depth(s)	1-bit bi-tonal; 4- or 8-bit grayscale or palette colour Up to 64-bit colour	1-8 bit bi-tonal, grayscale, or colour	8-bit grayscale 24-bit colour	1/2/4/8-bit palette colour or grayscale 16-bit grayscale, 24/48-bit true colour	4- or 8-bit grayscale or palette colour Up to 64-bit colour support
Compression	Uncompressed Lossless: CCIT G3/G4, LZW, Packbits, JPEG	Lossless: LZW	Lossy: JPEG	Lossless: Deflate, an LZ77 derivative	Uncompressed Lossless: CCIT, LZW. JBIG Lossy: JPEG
Standard/ Proprietary	De facto standard, LZW compression may require license	De facto standard, LZW compression may require license	JPEG: ISO 10918-1/2 JFIF: de facto standard	ISO 15948	ISO 15930-1:2001. De facto standard
Web Support	Plug-in or external application	Native since Microsoft® Internet Explorer 3	Native since Microsoft® Internet Explorer 2	Native since Microsoft® Internet Explorer 4	Plug-in or external application
Metadata Support	Basic set of labeled tags	Free-text comment field	Free-text comment field	Basic set of labeled tags plus user-defined tags.	Basic set of labeled tags
Strengths	Long history and widespread support as the premier format for high quality digital images and faxes Lossless compression Multiple pages	Widespread support through web browsers and imaging applications Lossless compression	Widespread use on websites Native support in nearly all imaging applications and web browsers Variable compression rates	Native support in web browsers Royalty free Supports all commonly used bit depths Lossless compression	Widely used Viewer freely available It can be used to distribute documents with images and text that will print consistently. Multiple pages

Limitations	Limited options for compression of colour and grayscale images Not natively supported by browsers LZW licensing restrictions JPEG version Not widely used	Not suitable for continuous tone images or photographs due to 8-bit maximum colour depth. LZW licensing restrictions	Used for photographs or images that has continuous tone, rather than text documents or graphics. Only lossy compression Not suitable for bi-tonal	Limited uptake Not supported within all imaging and scanning applications	Not an image format Requires post-processing for encapsulation Not natively supported by browsers Limited support for file creation in non-proprietary applications
Specification	http://partners.adobe.com/public/developer/tiff/index.html	http://www.w3.org/Graphics/GIF/spec-gif89a.txt	http://www.w3.org/Graphics/JPEG/	http://www.w3.org/TR/PNG/	http://partners.adobe.com/public/developer/pdf/index_reference.html

Source: Image formats Comparison: Adapted from http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html

Annexure 5: Imaging Software for Capture and Manipulation

Scanners are bundled with a piece of software (known as a driver) that is required for the controlling computer and the scanner to communicate. Additional software is typically included with the scanner which allows scanning, calibration and some post scanning image processing operations. This software will usually be tested thoroughly by the scanner manufacturer to work optimally with their hardware, with features appropriate to the type of scanner purchased. For example, software bundled with high speed sheet fed scanners would likely include features that would allow a choice to be made between single sided and duplex scanning and the recognition of barcodes, while the software that comes with slide scanners would typically include features for magnifying the originals and reversing the colours of negatives. If the bundled software does not fulfill eOffice scanning needs it may need to be supplemented or replaced by software that is purchased separately. Additional image processing software may also be used for tasks such as the conversion of file formats, derivation of related files, and modification or enhancement of images. Some degree of interoperability usually exists between most image processing and scanning software so that images can be acquired directly into the image processing software. To enable full text searching of scanned paper documents, OCR software would also be required.

Glossary

- **Active and Current files -** All active and current files will not be converted to electronic file from day one. The active/current files that will be taken up for transition to electronic file will depend on the Transition Approach adopted and finalized by the organization and the finalized Transition Time Table.
  - o The Digitisation activity of Active/Current files is normally best under taken by the Section concerned.

- **Department Nodal Officer (DNO) -** It is the second recommended institutional mechanism for implementing the eOffice programme, which shall be the focal point for the overall steering and required coordination of implementation efforts in consultation with the Steering Committee.

- **Digitisation Framework -** Digitisation Framework has been designed to be in sync with the overall eOffice Implementation. It depicts the areas that include the Governance Structure, Approach and Methodology, Implementation of the Digitisation to the final integration into eOffice.

- **eOffice Migration Module -** This module, a part of eFile is used to migrate the scanned physical files to electronic form.

- **File -** A collection of papers on a specific subject matter, assigned a file number and consisting of one or more of the following parts:
  - o Notes
  - o Correspondence
  - o Appendix to Notes
  - o Appendix to correspondence

- **Fresh Receipt (FR) -** A subsequent receipt on a case which brings in additional information to aid the disposal of a paper under consideration.

- **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.

- **IT ACT -** Information Technology Act 2000(http://deity.gov.in/content/information-technology-act-2000) came into effect from October 17, 2000.

IT (Amendment) Act 2008 is an Act further to amend the Information Technology Act,2000 (http://deity.gov.in/sites/upload_files/dit/files/downloads/itact2000/it_amendment_act2008.pdf)

- **LZW -** It is used for Grayscale and Colour scanned documents. Due to historical patent issues this is not always well supported, although as the patents expired in 2004 more recent software should provide good support.

- **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.

- **Open Standard -** A standard which meets all mandatory characteristics laid down in the Policy on Open Standards for e-Governance.

  See: Policy on Open Standards for e-Governance (http://egovstandards.gov.in)

- **Paper Under Consideration (PUC) -** A receipt pertaining to a case, the consideration of which is the subject matter of the case.

- **Project Implementation Committee (PIC) -** It is the fourth important institutional mechanism to give the steering support for the implementation of eOffice. This committee under the chairmanship of the Department Nodal Officer must be formed in a manner wherein there is representation from all Units involved.

- **Project Steering Committee (PSC) -** Project Steering Committee gives the overall direction and leadership to the implementation of eOffice. This committee, being advisory in nature, must be formed in a manner wherein all implementing divisions and key stakeholders are involved.

- **Receipt -** Refers to Dak after it has been received by the concerned section officer.

- **TIFF -** The TIFF format was created by Aldus Corporation in the mid-1980s in order to create a standard file format for storage of scanned images. The TIFF specification [1] is now controlled by Adobe although no major update to the specification has taken place since 1992. The TIFF format itself is a file format, not an image format. A TIFF file can be thought of as a container for one or more images each of which may be of a different type.

- **Transition Strategy -** This encompasses the decisions taken by the department under the PI&MC with regard to the transition to the electronic mode of working, particularly with respect to the electronic files.

Prepared by:

Suchitra Pyarelal   Scientist F, Head of eOffice Project Division, NIC

Kavita Garg   Director (eGovernance), DARPG

Saroj Kumar Patro   Scientist C, eOffice Project Division, NIC

Rohit Jandial   eOffice Project Division

Document Design by   Deepak Joshi, eOffice Project Division