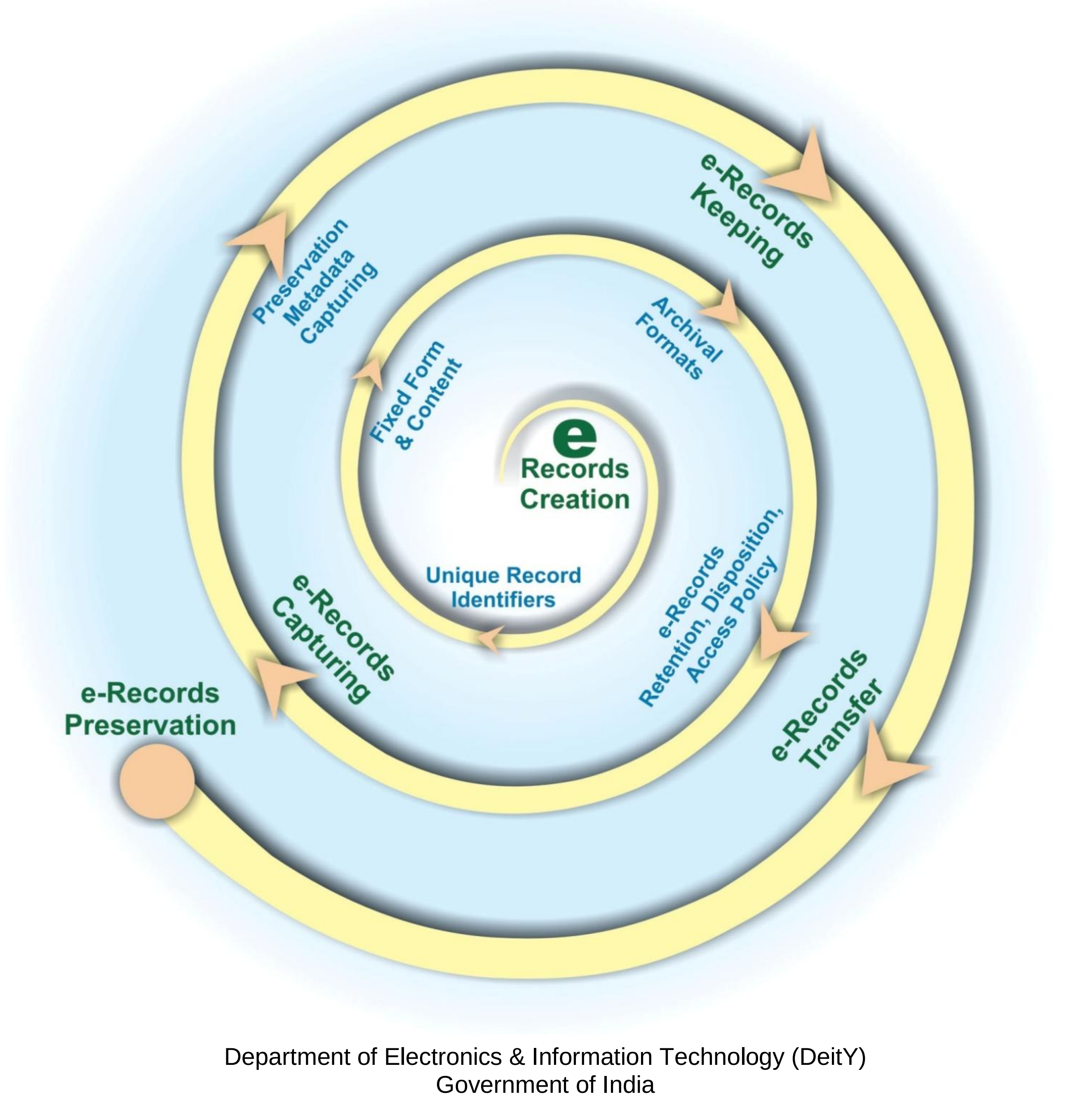
eGov.DP.01-01
Version: 1.0
December, 2013

Best Practices & Guidelines for Production of Preservable e-Records (PRoPeR)

PRoPeR Best Practices & Guidelines



Department of Electronics & Information Technology (DeitY)
Government of India

Metadata of the Document

S. No.	Data elements	Values
1.	**Title**	Best Practices & Guidelines for Production of Preservable e-Records (PRoPeR)
2.	**Title Alternative**	PRoPeR Best Practices & Guidelines
3.	**Document Identifier** *(To be allocated at the time of release of final document)*	eGOV.DP.01-01
4.	**Document Version, month, year of release** *(To be allocated at the time of release of final document)*	Version 1.0 December 2013
5.	**Present Status**	Notified
6.	**Publisher**	Department of Electronics and Information Technology (DeitY), Ministry of Communications & Information Technology (MCIT), Government of India (GoI)
7.	**Date of Publishing**	06/12/2013
8.	**Type of Standard Document** *(Policy / Technical Specification/ Best Practice /Guideline/ Process)*	Best Practices and Guidelines
9.	**Enforcement Category** *(Mandatory/ Recommended)*	Mandatory
10.	**Creator** ***(****An entity primarily responsible for making the resource)*	The Expert Committee for Digital Preservation Standards and Guidelines under the Chairmanship of Dr. Gautam Bose, Deputy Director General, National Informatics Centre (NIC)
11.	**Contributors** ***(****An entity responsible for making contributions to the resource)*	Centre of Excellence for Digital Preservation, Sponsored by DeitY, established at C-DAC Pune.
12.	**Brief Description**	This standard provides the best practices and guidelines for production of preservable electronic records and its management in the context of e-Governance. It is applicable for those e-records that need to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. The core concepts of 'preservability' are based on the requirements specified in IT ACT, ISO/TR 15489-1 and 2 Information Documentation - Records Management and ISO 14721 Open Archival Information Systems (OAIS) Reference Model. It introduces 5 distinct steps of e-record management

S. No.	Data elements	Values
		i.e. e-record creation, e-record capturing, e-record keeping, e-record transfer to designated trusted digital repository and e-record preservation which need to be adopted in all e-Governance projects.
13.	**Target Audience** *(Who would be referring / using the document)*	• e-records producers and data managers • Departmental Record Officers (DROs) record keepers, archivists and preservation officers • All stakeholders in central and state government, as well as public and private organizations involved in execution, design, development and implementation of e-Governance applications. • Central, state, district level archiving organizations
14.	**Owner of approved standard**	DeitY, MCIT, New Delhi
15.	**Subject** *(Major Area of Standardization)*	Digital Preservation
16.	**Subject Category** *(Sub Area within major area)*	Preservability and Management of Electronic Records
17.	**Coverage. Spatial**	INDIA
18.	**Format**	PDF
19.	**Language** *(To be translated in other Indian languages later)*	English
20.	**Rights. Copyrights**	DeitY, MCIT, New Delhi
21.	**Source** **(***Reference to the resource from which present resource is derived)*	• InterPARES 2, International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records, 2008 • ISO/TR 15489-1 and 2 Information and Documentation - Records Management, 2001 • ISO 14721:2012 Open Archival Information System (OAIS) Reference Model • ISO 19005-1:2005 PDF/A-1 • ISO 19005-2:2011 PDF/A-2 (Use of ISO 32000-1: 2008) • ISO/DIS 16363 Audit & Certification of Trustworthy Digital Repositories Adaptation of above sources is based on the research carried out by the team of Centre of Excellence for Digital Preservation Project at C-DAC Pune.
22.	**Relation** *(Related resources)*	This document has to be used in conjunction with eGOV-PID: Metadata Dictionary & Schema.

Table of Contents

Statement of Purpose

This document provides a set of best practices and guidelines for production of preservable electronic records and its management in the context of e-governance. It is applicable for those e-records that need to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. The core concepts of 'preservability' are based on requirements specified in IT ACT, ISO/TR 15489-1 and 2 Information Documentation - Records Management and ISO 14721 Open Archival Information Systems (OAIS) Reference Model. Best practices and guidelines specified in this document are to be used in conjunction with e-Governance Standard for Preservation Information Documentation (eGOV-PID) for Electronic Records.

Acronyms and Definitions

Archival

The e-records are captured and removed from the routine workflow and placed in safe, separate, yet accessible and searchable storage.

Born digital

The term born-digital refers to materials that originate in a digital form.

Current e-records

The current or active e-records are regularly used for the current business of an agency, institution or organization and continue to be maintained in their place of origin or receipt. The current e-records can be subjected to further modification and processing.

Digital object

An object composed of a set of bit sequences. An e-record with fixed information content is also called as 'digital object'.

Electronic Record (e-Record)

The ISO 15489-1:2001 defines records as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business". As per the IT ACT 2000 "electronic record" means data, record or data generated, image or sound stored, received or sent in an electronic form or micro film or computer generated micro fiche. The electronic records or digital content are produced in the form of text, images, documents, e-files, audio, video, 3D models, web pages, maps, datasets, computer generated micro fiche and various other forms.

Electronic Records Management (ERM)

The electronic records management is the practice of maintaining the e-records of an organization from the time they are created up to their eventual disposal which includes classifying, storing, securing, archival, preservation and destruction.

Reformatted digital

Digital reformatting is the process of converting analogue materials into a digital format as a surrogate of the original. The digital surrogates perform a preservation function by reducing or eliminating the use of the original.

Rendered e-record

The e-records which are stored with proper rendering in terms of human sensory attributes are considered as rendered e-records.

Long Term Digital Preservation

Long Term Digital Preservation is a secure and trustworthy mechanism to ingest, process, store, manage, protect, find, access, and interpret digital information such that the same information can be used at some arbitrary point in the future in spite of obsolescence of everything: hardware, software, processes, format, people, etc. It should be preserved along with the details which will facilitate the identification of the origin, destination, date and time of such electronic record. The e-record has to be preserved in such a way that it will remain accessible, reliable, authentic and usable for a subsequent reference.

Open Archival Information System (OAIS)

An Open Archival Information System (OAIS) is an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a designated community. The OAIS Reference Model is defined by recommendation CCSDS 650.0-B-1 of the Consultative Committee for Space Data Systems, which is also accepted as ISO 14721:2012.

Non-current e-records

The non-current or inactive e-records are complete in all respects and no longer required for day-to-day conduct of an active business.

Non-rendered e-record

Typically the values or information stored in the database which require further processing or another software for meaningful interpretation are considered as non-rendered e-records.

Preservable

Capable of being preserved, ready to be preserved, complying with the requirements of preservation.

Trustworthy Digital Repository

A digital repository is an organization that has the responsibility for long term preservation of digital resources, as well as making them available to communities agreed upon by the depositors to the repository. The trustworthiness of a digital repository, as defined in ISO 16363: 2012 is established through periodic audit and certification which guarantees the capacity of a digital repository to deal with the threats and risks within its systems, to monitor, plan, and maintain the digital resources, as well as the ability to act and implement the strategy for digital preservation.

Unique record identifier for e-record

A unique record identifier is a numeric or alphanumeric string that is associated with a single entity i.e. an e-record within a given system. The unique record identifier is persistently linked or associated with the e-record which helps in its reference, location, identification, authentication, access and control. It is also used as the filename for storing the e-record.

NAS

Network Attached Storage

SAN

Storage Area Network

Best practices & guidelines

1. Aim

Best practices and guidelines provided in this document are aimed at defining the technical parameters for capturing of electronic records (e-records) so that they could be retained, managed and preserved as per the record retention, disposition and access rules.

2. Scope

This document covers a set of best practices and guidelines for defining the aspects of preservability of born digital records; e-record management practices; techniques for capturing the e-records in the form of text, images, documents and e-records stored in database that have to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. Best practices and guidelines provided in this document are also applicable for reformatted digital records.

3. Exclusions

This document does not include the digitization specifications of conversion from analogue to digital for paper records, audio and video materials.

4. Normative references

- Information Technology Act, 2000, Government of India
- Information Technology Act Amendment (ITAA) 2008, Standing Committee Recommendations, Government of India
- IT Act Notifications GSR 582, 6th September, 2004, Published by Ministry of Communications and Information Technology, Government of India
- Right To Information Act 2005, Government of India
- ISO/TR 15489-1 and 2: 2001 Information and Documentation - Records Management
- Extensible Markup Language (XML), World Wide Web Consortium (W3C)
- ISO 19005-1:2005 PDF/A-1 (Use of PDF 1.4)
- ISO 19005-2:2011 PDF/A-2 (Use of ISO 32000-1: 2008)
- ISO 14721:2012 Open Archival Information System (OAIS) Reference Model
- ISO/DIS 16363: 2012 Audit & Certification of Trustworthy Digital Repositories

- InterPARES 2, International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records, 2008

5. Need for production of preservable e-records

5.1. Threat of digital obsolescence

The national and state level initiatives of e-governance, e-service delivery, computerization and digitization across various domains are producing enlarging volumes of e-records which must be preserved as per the retention rules and to fulfill various legal obligations. The e-records can be quickly lost much before the assigned retention period due to obsolescence of file format, storage media, database, software and vendor lock-in as result of dependence on proprietary solutions. The obsolescence of e-records and the evidentiary proofs can create problems in administrative, judiciary and legislative functions in addition to loss of valuable information, intellectual property and heritage. Therefore, it is necessary to ensure that the e-records which require to be retained for long duration are produced by incorporating the guidelines to ensure its long term preservability.

5.2. Legal requirements

The best practices and guidelines for the production of preservable e-records are defined in order to comply with IT Act which specifies the requirements for retention of electronic records (section 7) as under –

1) Where any law provides that documents, records or information shall be retained for any specific period, then, that requirement shall be deemed to have been satisfied if such documents, records or information are retained in the electronic form, if:

(a) the information contained therein remains accessible so as to be usable for a subsequent reference;

(b) the electronic record is retained in the format in which it was originally generated, sent or received or in a format which can be demonstrated to represent accurately the information originally generated, sent or received;

(c) the details which will facilitate the identification of the origin, destination, date and time of dispatch or receipt of such electronic record are available in the electronic record:

As per the IT Act Notifications GSR 582, the e-records creation system or software should take into account the following features of e-records-

- life time
- preservability
- accessibility
- readability
- comprehensibility in respect of linked information
- evidentiary value in terms of authenticity and integrity
- controlled destructibility and
- augmentability

As per the IT Act Amendment 2008, Standing Committee Recommendations audit of electronic documents or e-records is essential as under -

- Where in any law for the time being in force, there is a provision for audit of documents, records or information, that provision shall also be applicable for audit of documents, records or information processed and maintained in electronic form.

As per the Right To Information Act 2005, Chapter II, Section 4(1) every public authority is obliged to maintain all its records duly catalogued and indexed in a manner and the form which facilitates the right to information under this Act and ensure that all records that are appropriate to be computerized are, within a reasonable time, computerized and connected through a network all over the country on different systems so that access to such records is facilitated.

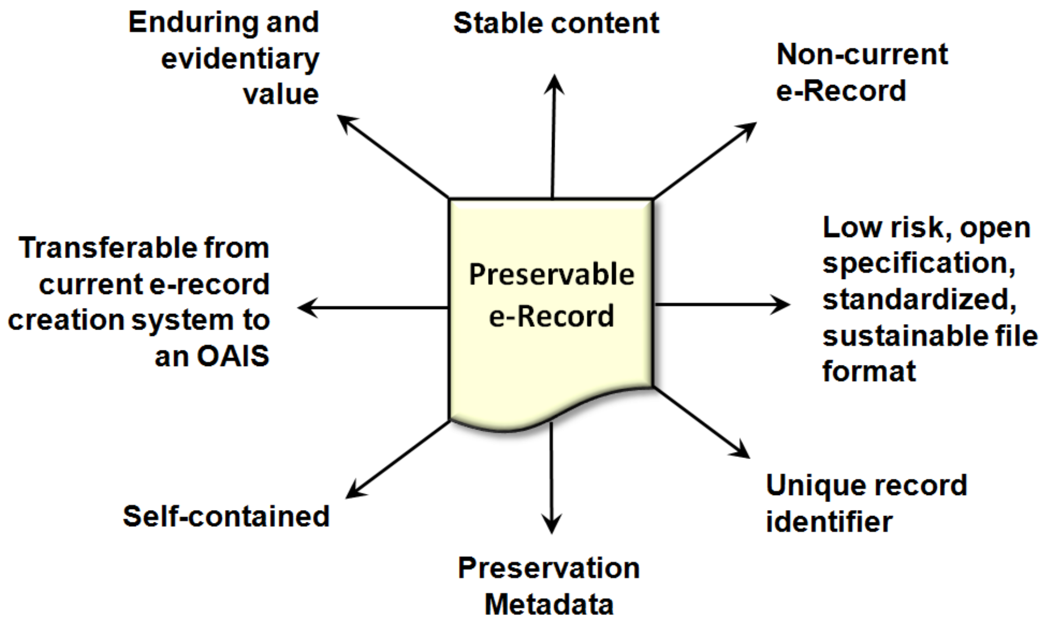

Enduring and evidentiary value
Stable content
Non-current e-Record
Transferable from current e-record creation system to an OAIS
Preservable e-Record
Low risk, open specification, standardized, sustainable file format
Self-contained
Preservation Metadata
Unique record identifier

Figure 1: The characteristics of preservable e-records

6. Characteristics of preservable e-records

An e-record is considered to be preservable when it complies with the following characteristics -

- It should be a non-current e-record.
- The e-record should be stored in the form of a digital object containing fixed information content (the e-record which is complete in all respects and does not require any further processing).
- It should be stored in low risk, open, standards based formats.
- It should be registered with unique record identifier or accession number.
- It should have an enduring and evidentiary value.
- The preservation metadata in terms of reference, context, provenance, fixity, representation information, authenticity and access rights is captured along with the e-record.
- It should possible to remove such e-record along with associated metadata from the current e-government system and store it on a separate storage for archival purpose.
- As far as possible, it should be self contained.

6.1. Difference between current and non-current e-records

Refer figure 2 to understand the difference between current and non-current e-records. The definitions of current and non-current e-records are given below.

6.1.1. Current e-records

The current or active e-records are regularly used for the current business of an agency, institution or organization and continue to be maintained in their place of origin or receipt. The current e-records can be subjected to further modification and processing. The current e-records are maintained within the e-records creation system or in the data centre for live transactions.

6.1.2. Non-current e-records

The e-records which are complete in all respects and no longer required for day-to-day conduct of an active business are referred as non-current or inactive records, which are required to be transferred to digital repository for preservation. The non-current e-record is the final output of the e-records creation system.

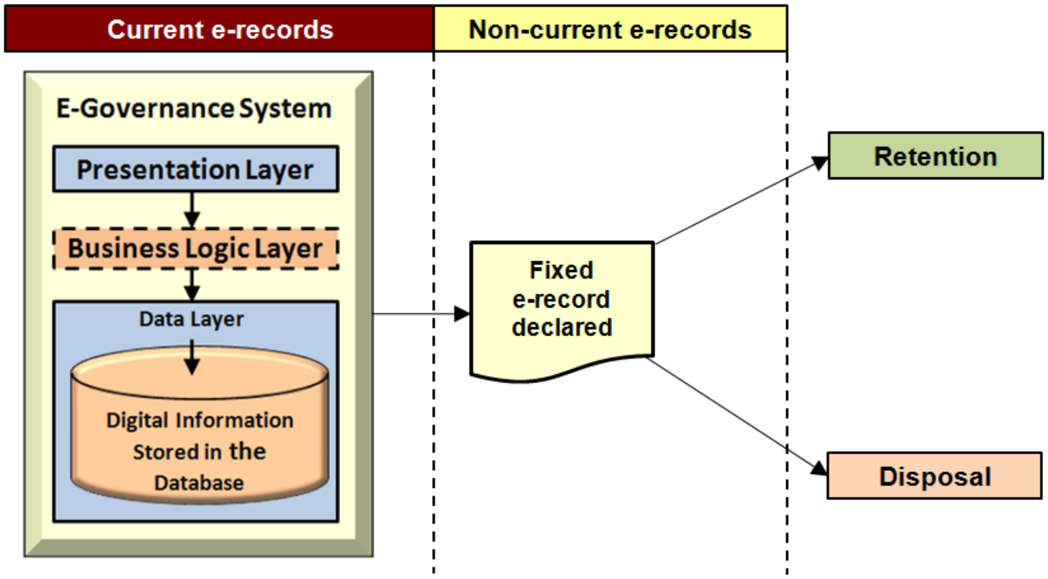


Figure 2: Current and non-current e-records

6.2. Fixed content and form for an e-record

The e-records are a result of the business logic which involves workflow, formulas applied for calculated values, dependencies between values and various functions in force. Therefore any change in the business logic, representation logic and rendering logic can change the e-record in an undesirable manner. Therefore, after the e-record is finalized in all respects, it is necessary to capture it in the form of a digital object for the purpose of preservation.

6.3. Retention, disposition and access policy for electronic records

The concerned administrative departments must define and notify the retention, disposition and access policy / rules applicable for electronic records, which should cover the following aspects-

- The types of electronic records which should be captured for preservation
- The duration for which the e-record should be retained
- The guidelines for e-records retention should include –
    - the legal requirements that need to be fulfilled.
    - the procedure for disposition of e-records after retention duration is over.
- List of e-records which will remain in the current e-gov system and the duration
- List of e-records which will be transferred to designated trustworthy digital repository
- The duration for which the metadata of transferred e-records to be maintained by the organization
- The access policy applicable to various e-records should include guidelines on matters like public access, privacy / confidentiality, provision of e-records to another organization for usage, bulk sharing, provision of extracts, access over intranet, access over Internet, access duration, downloading of e-records, re-publishing, fees, etc.

For example, refer the retention schedule for electronic records (Section J) in the Record Retention Schedule in Respect of Records Common to All Ministries / Departments - 2012 released by Department of Administrative Reforms and Public Grievances (DARPG), Government of India. National Data Sharing and Accessibility Policy (NDSAP - 2012) by Department of Science & Technology, Government of India can be referred as an example, however the access policies for e-records produced by different organizations and businesses will require to address domain specific concerns of all stakeholders and legal aspects of e-records being preserved. Annexure C may be referred for more examples.

In the context of e-governance, with the help of e-records retention, disposition and access policy it will be possible to -

- anticipate the costs involved in storage and archival of e-records.
- optimize the storage requirements.
- manage and control access to e-records
- become aware of the risks involved in case the e-records are destroyed unknowingly.
- evaluate the legal, social, political, and financial gains from preserving the e-records.

6.4. Selection criteria for capturing the e-records as fixed objects

The e-records should be captured in the form of a fixed digital object on the basis of following criteria-

- The retention and disposition rules pertaining to e-records
- The legal obligations and implications of failing to reproduce such e-record in its original and authentic form in future.
- The value of information contained in the e-record
- Whether the e-record serves as the basis for other transactions
- The historical significance of the e-record

The e-records that need to be retained for ten years or more are generally considered to be more vulnerable to change in technologies and obsolescence.

6.5. Store the e-records in low risk, open and standards based formats

While storing the e-records in specific file formats it is necessary to know whether they are non-rendered or rendered e-records as shown in figure 3.

6.5.1. Non-rendered e-records

Typically the values or information stored in the database which require further processing or another software for meaningful interpretation are considered as non-rendered e-records. Such non-rendered e-records can be captured in XML document format for preservation.

6.5.2. Rendered e-records

The e-records which are stored with proper rendering in terms of human sensory attributes are considered as rendered e-records. Such rendered e-records are captured in PDFA or Image formats.

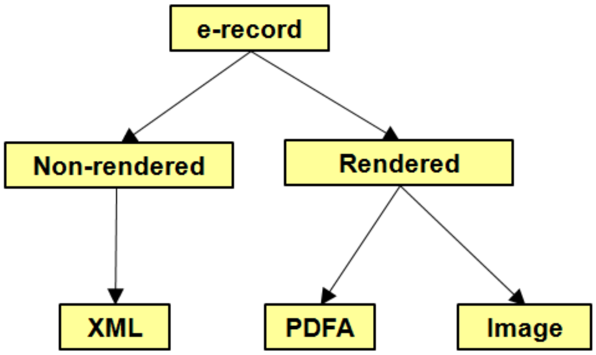


Figure 3: Non-rendered and rendered e-records

6.5.3. File formats for e-record capture

The file format guidelines for e-record capturing include:

a) Capture of e-records stored in database as XML document format

- Finalized e-records stored in a database should be exported and published as XML document for machine readability, portability, re-use, interoperability and preservability.
- The domain specific metadata captured as XML document should have appropriate / standardized metadata schema to enable proper representation and interpretation of its contents. The Metadata and Data Standards – Demographic, Version 1.1, November 2011 published by Ministry of Communications and Information Technology, Department of Electronics and Information Technology, Government of India should be used as applicable.
- The XML schemas used for e-record capturing should be maintained along with its URI and version information for its use and validation in future.
- Images and digital signature forming an integral part of the e-record can be encoded in generic base64 format for storing within XML document.
- The successive relative path of the linked objects (enclosures) should be maintained within the XML.
- The XML document may be digitally signed with detached digital signature.
- The e-record (XML and linked objects together forming a single entity) can be packaged as Open Packaging Conventions (OPC) standardized in the ISO/IEC 29500:2008 or in ZIP file format along with deflate compression being standardized as per the ISO/IEC NP 21320-1.
- The data should not be encrypted or password protected.

b) XML based domain specific formats

XML based domain specific and standardized formats such as XBRL, ODF (ISO/IEC 26300: 2006), Office Open XML (ISO/IEC 29500: 2008), SVG, etc can also be used depending upon the type of application.

c) Capture of e-records in PDF for Archival (PDFA) format

ISO 19005-1:2005 Use of PDF 1.4 (PDF/A-1) with Level A conformance is recommended for archival of "born digital documents" due to following reasons –

- PDF/A-1a is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (implemented in Adobe Acrobat 5 and latest versions) and is defined by ISO 19005-1:2005.
- The specification is open and fully documented.
- PDFA preserves the visual appearance of document.
- It includes visible contents like text, raster images, vector graphics, fonts, color information.
- Supports Unicode character map
- Documents logical structure
- It is self contained as it embeds the fonts, images, metadata, color profile, rendering specifications in the document itself.
- Ensures long term reproducibility - internationally accepted as a Standard for long-term electronic archiving

ISO 19005-1:2005 Use of PDF 1.4 (PDF/A-1b) with Level B conformance is recommended for archival of reformatted digital documents due to following reasons -

- PDF/A-1b preserves the visual appearance of the document
- Digitized documents in image format can be composited as PDF/A-1b

ISO 19005-2:2011 Use of ISO 32000-1 (PDF/A-2) is also recommended for preservation of documents requiring the advanced features supported in it.

PDF/A-2a is based on ISO 32000-1 – PDF 1.7 and is defined by ISO 19005-2:2011. Its features are as under -

- Support for JPEG2000 image compression
- Support for transparency effects and layers
- Embedding of OpenType fonts
- Provisions for digital signatures in accordance with the PDF Advanced Electronic Signatures – PAdES standard
- Possibility to embed PDF/A files in PDF/A-2 for archiving of sets of documents as individual documents in a single file

PDF/A-2 does not replace the PDF/A-1 standard but it co-exists alongside with an extended set of features.

PDF/A-1a and PDF/A-1b compliance are minimum essential for e-government records as recommended in the IFEG technical standard of DeitY.

d) Image file formats

JPEG2000 (ISO/IEC 15444-1:2004) and PNG (ISO/IEC 15948:2004) which support lossless compression are recommended as raster image file formats for e-governance applications as specified in Technical Standards for Interoperability Framework for e-Governance (IFEG) in India, published in 2012 by e-Gov Standards Division, Department of Electronics and Information Technology (DeitY).

Captured e-records should be stored inside separate folders named as per the respective unique record identifiers on the file system.

6.6. Assign unique record identifiers to e-records

6.6.1. Unique record identifiers

A unique record identifier is a numeric or alphanumeric string that is associated with a single entity i.e. an e-record within a given system. The unique record identifier serves as a “filename” for an e-record. Thus it is persistently linked or associated with the e-record so as to help in its reference, location, identification, authentication, access and control.

The e-record producing organization should define a set of rules or conventions for generating the unique record identifiers as per the guidelines given below -

- Select at least 3 to 4 relevant elements as per the examples given in 6.6.2 for defining the rules or conventions for generating the unique record identifiers.
- Standardize the abbreviations or short forms to be used for referring the classification of e-record.
- Standardize the separators between the filename elements and numbering.
- Avoid the use of controlled characters and empty spaces.
- The length / character sets of the unique identifier should be compatible across operating systems / file systems as it has to be used as filename of the e-records.

6.6.2. Examples of common elements in the unique record identifier

A unique record identifier can be generated by selecting any 3 to 4 relevant elements from the examples given below -

- Name of organization
- Type of document
- Service code
- Accession number or registration number or reference number
- Place
- Date of creation
- Name of creator / organization
- Description or title of content
- Release date
- Publication date
- Project number
- Department number
- Records series
- Version number
- Other domain specific elements preferred by the organization

6.6.3. Benefits of assigning unique record identifiers to e-records

Following are the benefits of assigning the unique record identifiers to e-records -

- It serves as a kind of metadata about the e-record.
- It helps in locating the e-records.
- It helps in establishing the referential integrity of e-records and distinguishing between versions.
- The digital rights and access controls are directly linked with the unique identifier of e-record.
- It helps in ensuring the trustworthiness and legal admissibility of an e-record.
- The documents and folders associated with an e-record also use the derivatives of its unique identifier.
- It helps in parsing the e-records for meaningful classification or various batch operations.
- It is not possible to aggregate the e-records in a trustworthy digital repository unless they have unique identifiers.

6.7. Capture the preservation metadata

The e-government system should automatically capture the associated preservation metadata of the e-record during its production process, which is otherwise likely to be lost forever as it remains scattered in different parts of the system. The preservation metadata includes information pertaining to cataloging, enclosures, provenance, fixity, representation, authenticity and access rights. Refer e-Governance Standard for Preservation Information Documentation (eGOV-PID) of Electronic Records, which provides the common metadata dictionary and schema for this purpose.

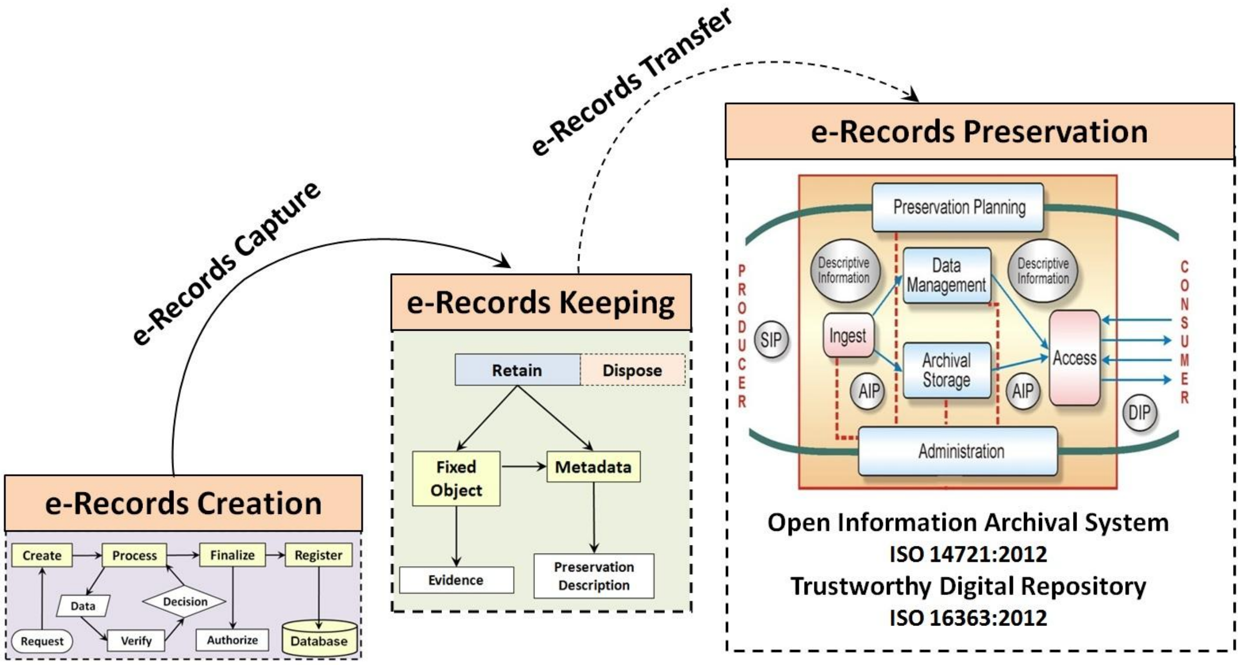


Figure 4. Electronic Records Management Practice

7. Incorporate the e-records management practice

The electronic records management is the practice of maintaining the e-records of an organization from the time they are created up to their eventual disposal which includes classifying, storing, securing, archival, preservation and destruction. The distinct stages of e-record management are defined as under-

7.1. e-Records creation

e-records creation is a process that is responsible for production of e-records. The typical e-government systems are designed to create and maintain the e-records for current transactions.

7.2. e-Records capturing

In order to ensure the “preservability” of e-records, the e-record producing organizations should properly analyze the record requirements of the business in terms of the records that need to be captured as legal evidence for future use. The captured e-records should be stored into open, fully-specified, standards-based file formats. The e-records creation systems should be designed to capture the e-records that need to be retained and preserved.

7.3. e-Records keeping

The electronic records that need to be retained for ten years or more should be captured and then removed from the current e-records creation system and kept in a safe, separate and yet accessible storage.

7.3.1. Guidelines for e-records storage

The archiving of e-records i.e. disposed/closed e-files and correspondences have to be taken at regular intervals on SAN or NAS storage in the data centre / e-record room and one more copy is to be maintained at the location of disaster recovery site as required in the Central Secretariat e-Manual of Office Procedure (CSeMOP), DARPG, Government of India, 2012. The e-records need to be maintained properly till they are transferred to the designated Trustworthy Digital Repository.

If the e-records are maintained in removable storage media e.g. CD, DVD, Blu-Ray Disk, Flash Disks, LTO Tapes, etc then the e-record keepers must take care of the following-

- The digital storage media should be selected on the basis of the following criteria-
  - its proven experience of longevity
  - capacity (appropriate for the quantity of e-records)
  - durability (low susceptibility to physical damage)
  - viability (availability of support for its long-term readability, data recovery in case of media failure)
  - mature and established technology
- The digital storage media containing the e-records should be properly numbered, classified, labeled and indexed in a register.
- At least two copies of the e-records should be maintained into two different types of media.
- The digital storage media should be preferably maintained in two different locations.

- The digital storage media should be kept in a safe, secure and temperature controlled environment as specified by the manufacturer.
- The e-records stored in the media need to be retained for longer duration than the lifetime of the media then it must be migrated into another media before its expiry.

Refer ISO/IEC 27002: 2005 - Code of practices for information security management for ensuring the security of the e-records archived on digital storage.

7.3.2. Periodic auditing of digital storage media

The digital storage media should be audited periodically which involves testing, refreshing and migration until the e-records are transferred to the designated Trustworthy Digital Repository for preservation.

7.4. e-Records transfer

The captured e-records should be transferred with a fixed periodicity to designated Trustworthy Digital Repository for long term preservation.

7.5. e-Records preservation

The Trustworthy Digital Repository as specified in ISO 16363: 2012 is a special facility with dedicated infrastructure and specialized human resource where the e-records are preserved for long term purposes. Refer Annexure A. for comparison between data centre and trustworthy digital repository for further clarity.

7.6. Roles and responsibilities

The ISO 16363: 2012 Audit and Certification of Trusted Digital Repositories requires to define the roles and responsibilities of the staff engaged for digital preservation. The personnel in the roles of e-records producer, e-records keeper and e-records manager are required at the records creating agency, whereas the personnel in the roles of digital archivist, digital curator, digital repository manager, digital repository administrator are required for the trusted digital repository. Generic definitions of the roles along with board responsibilities are defined in this section to allow suitable adaptation in different organizational contexts.

7.6.1. Records creating agency

The roles described in this section require technical skills and knowledge of Information Technology to be able to effectively manage the e-records.

- **e-Records producer**

The role played by those persons that analyze the business requirements and produce the records to be preserved. The records creating agency is also referred as records producer.

- **e-Records keeper**

The digital record keeper is responsible for identifying, capturing and maintaining the e-records as per the record retention and disposition rule.

- **e-Records manager**

The e-records manager for an organization is the person responsible for the management of records in the organization. The records management activities include policy, procedures, creation, use, retention, disposition and transfer of records.

The roles of e-records keeper or e-records manager are similar to the role of 'record officer' as defined in Public Records Act.

7.6.2. Trusted digital repository

The roles described in this section require relevant domain expertise, knowledge of digital preservation best practices, ability to conceptualize and develop technological solutions and manage the digital preservation infrastructure.

- **Digital archivist**

A digital archivist is an expert competent to appraise, acquire, authenticate, preserve, and provide access to records in digital form.

- **Digital curator**

A digital curator has the domain knowledge to improve the quality of information and the data being stored in the digital repositories for present and future use.

- **Digital repository manager**

A digital repository manager has the technical expertise to manage and support the workflows, hardware and software infrastructure necessary for digital preservation.

- **Digital repository administrator / archive administrator**

The digital repository administrator or archive administrator looks after the administration of staff, budgets, facilities, logistics, and other support functions of the digital repository.

Digital preservation is a highly technology driven activity, and therefore, a trusted digital repository requires to be strongly supported and sustained by human resource with technical skills in software development and system / storage / network administration.

8. Summary of best practices and guidelines

1.	Design the e-government system or e-records creation system to enable capturing of e-records that need to be preserved for long durations.
2.	Incorporate the digital preservation best practices and standards provided in PRoPeR Best Practices & Guidelines and e-Gov-PID Preservation Metadata and Schema.
3.	Design the composition of e-record in terms of the main e-record, linked objects, enclosures, filenames, preservation information, folders and overall structure of the contents before starting to capture the e-records.
4.	Define the e-record retention, disposition and access policy according to the laws and regulations and the value of information contained in the e-record for posterity.
5.	Select and identify the type of e-records that need to be retained and preserved.
6.	Categorize the current and non-current e-records.
7.	Categorize the non-rendered and rendered e-records so as to decide the file formats for capturing them.
8.	The e-record that needs to be preserved should be captured with fixed content and form.
9.	Define a set of rules or conventions for assigning unique record identifiers to e-records.
10.	Use the unique record identifier of e-record as its filename at the time of its capture.
11.	The documents, images and folders associated with main e-record should be named using its unique identifier with appropriate suffix separated by underscore character.
12.	Store the fixed e-records in low risk, open and standards based formats.
13.	Capture the non-rendered e-records in XML format by using appropriate / standardized schema.
14.	Capture the rendered e-records in PDF/A format.
15.	Capture the preservation metadata while producing the e-record, as per the e-Governance Standard for Preservation Information Documentation (eGOV-PID) of Electronic Records.
16.	Remove the captured e-records from the current e-records creation system and keep them in a safe, separate and yet accessible storage.

17.	Storage media containing the captured e-records should be properly numbered, classified, catalogued and labeled.
18.	The digital storage media should be audited periodically which involves testing, refreshing and migration.
19.	Transfer the captured e-records to designated Trustworthy Digital Repository for long term preservation.
20.	Define the roles and responsibilities of personnel with requisite skills for e-record management and digital preservation.

9. References

1. Information Technology Act, 2000, Government of India
2. Information Technology Act Amendment (ITAA) 2008, Standing Committee Recommendations, Government of India
3. IT Act Notifications GSR 582, 6th September, 2004, Published by Ministry of Communications and Information Technology, Government of India
4. Right To Information Act, 2005, Government of India
5. ISO/TR 15489-1 and 2 Information and Documentation - Records Management: 2001
6. E-Government Electronic Records Management Initiative, National Archives Records Management (NARA), 2003
7. ISO 14721:2012 Open Archival Information System (OAIS) Reference Model
8. Extensible Markup Language (XML), World Wide Web Consortium (W3C)
9. ISO 19005-1:2005 Use of PDF 1.4 (PDF/A-1)
10. ISO 19005-2:2011 Use of ISO 32000-1 (PDF/A-2)
11. ISO/DIS 16363 Audit & Certification of Trustworthy Digital Repositories
12. ISO/IEC 27002: 2005 - Code of practices for information security management
13. Metadata and Data Standards – Demographic, Version 1.1, published by Ministry of Communications and Information Technology, Department of Electronics and Information Technology, Government of India, November 2011
14. Technical Standards for Interoperability Framework for e-Governance (IFEG) in India, Version 1.0, Published by e-Gov Standards Division, Department of Electronics and Information Technology, May 2012
15. Archivi, International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2, Edited by Luciana Duranti and Randy Preston, Published by Padova, Italy, 2008
16. InterPARES 2, International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records, 2008

17. Electronic Records Management: An Audit Guide, EUROSAI IT Working Group, Version 0.8
18. Electronic Records (Section J), Record Retention Schedule in Respect of Records Common to All Ministries / Departments, Published by Department of Administrative Reforms and Public Grievances (DARPG), 2012.
19. National Data Sharing and Accessibility Policy (NDSAP - 2012), Department of Science & Technology, Ministry of Science & Technology, Government of India

Annexure A. Comparison between Data Centre and Trustworthy Digital Repository

Table 1. Comparison between Data Centre and Trustworthy Digital Repository

	Data Centre		**Trustworthy Digital Repository**	
Parameters	**Business Applications**	**Backup**	**Archive**	**Digital Preservation**
1. Objectives	Business Continuity	Recovery	Discovery	Usability in future
2. Definition	Process live transactions Maintain the active or current data	Copy files to a second medium	Capture and move the e-records that are no longer actively used to a separate storage device	A secure and trustworthy mechanism to ingest, process, store, manage, protect, find, access, and interpret digital information such that the same information can be used at some arbitrary point in the future in spite of obsolescence of everything: hardware, software, processes, format, people, etc.
3. Why	-	Restoration of data in the event of deletion, corruption or loss	To keep the e-records unaltered and available electronically	Digital preservation is necessary for - access to digital information through technological obsolescence - ensuring administrative continuity - protection of digital intellectual assets - reuse of digital information - long term view - fulfillment of legal obligations - protection from litigation
4. What	**-**	- a snapshot of active data in bulk at that point in time - multiple versions of data - can include entire production system - application specific data	The e-records that require to be retained for long durations are captured and removed from the routine workflow and placed in safe, separate, yet accessible and searchable storage.	The final e-records - that require to be retained for long durations as per the retention and disposition rules. - for legal purposes. - with enduring value of information. - which have historical importance.
5. How long	**-**	Daily / weekly/ monthly (high frequency short	digital data can be kept in the archive for 3 to 5 years	(usually for minimum 10 years, 25 years, 50 years or permanent)

		life span)	Beyond 3 to 5 years it is difficult to guarantee the readability of storage media or formats	The e-records are preserved as per legal requirements or retention rules or depending on the heritage / historical value of the content.
6. How	**-**	For security and efficiency reasons, backup data may be encrypted and compressed.	The archival files are stored in their original format along with cataloging information.	The e-record has to be preserved in such way that it should be possible to find, read, represent, render and interpret the information accurately as original along with all associated information necessary for its comprehension. It should be preserved along with the details which will facilitate the identification of the origin, destination, date and time of such electronic record. The information contained therein to remain accessible, reliable and authentic so as to be usable for a subsequent reference. Periodic refreshment, migration and conversion as required.
7. Who	Intra-organization, 3rd party service provider	Intra-organization	Intra-organization or Designated archiving organization	Designated Trustworthy Digital Repositories or 3rd party digital preservation service provider who is regularly audited for trustworthiness
8. Standards	ISO /IEC 27001 Information security management	-	International Standard Archival Description (ISAD)	ISO/DIS 16363 – 2012 Audit & Certification of Trustworthy Digital Repositories ISO 14721:2012 Open Archival Information Systems (OAIS) Reference Model ISO/TR 15489-1 and 2 Information and Documentation - Records Management: 2001
9. Policy	-	Backup procedures are defined by respective teams, departments,	An organization derives its archival policy based on record retention rules and legal requirements.	As per the ISO 16363 – 2012 Standard for Trusted Digital Repositories, there has to be an authorized digital preservation policy (a framework of rules and

		organization.		regulations) which should guide and support the digital preservation activities.
10. Audit	Data centre audit covers infrastructure and information security aspects	-	-	The trustworthiness of digital repository is established through self audit and 3rd party audit as per the as per ISO 16363 – 2012 for Audit and Certification of Trustworthy Digital Repository.

Annexure B. Guidelines for PDFA implementation

PDF/A for e-governance applications

- Apache FOP 1.1 library can be used in the application logic for dynamically publishing the e-records in PDF/A format.

PDF/A for document creation

- Libre Office 4.0 supports the exporting of a document in PDF/A format.
- MS Office 2007 onwards the support for “save as” PDF/A is available.
- Adobe Acrobat Professional can be used for converting the PDF documents to PDF/A format.

Annexure C. Examples of record retention, disposition and access policies

Following record retention and disposition rules are referred during the background study.

1.	Record Retention Schedule in Respect of Records Common to All Ministries / Departments, 2012	Department of Administrative Reforms and Public Grievances (DARPG), Government of India
2.	Destruction of office records connected with accounts, General Financial Rule, 2005	Ministry of Finance, Government of India
3.	Period of Preservation of Records, Banking Regulation Act, 1949	Government of India
4.	Bank's Policy on Record Maintenance for the year 2012-13	All India Indian Bank Officers' Association
5.	Record Retention Schedule for Records Relating to Substantive Functions, 2006	Ministry of Labor & Employment, Directorate General of Employment & Training, Government of India
6.	Retention period/destruction schedule of recorded files, 2006	Central Vigilance Commission, Government of India
7.	Record retention requirements as per the Right To Information Act, 2005	Ministry of Law and Justice, Government of India
8.	Record retention rules of UPSC institutions	Union Public Service Commission (UPSC), Government of India
9.	Disposal of Records (in the Offices of the Registrars of Companies) Rules, 2003	Ministry of Finance and Company Affairs, Government of India
10.	Destruction of Records Act, 1917	Ministry of Law and Justice, Government of India

Variety of such records retention and disposition requirements are specified in various acts and rules adopted by different organizations. Several examples of access policies defined by other countries are available online for reference such as Registry of Birth, Death, Marriage (BDM)

Access Policy of Department of Justice, Government of Australia; Public Access policy of The Unified Judicial System of Pennsylvania, USA, Health Insurance Portability and Accountability Act (HIPPA) provides access policies for electronic health records, USA, etc.

National Data Sharing and Accessibility Policy (NDSAP - 2012) by Department of Science & Technology, Government of India can be referred as an example, however the access policies for e-records produced by different organizations and businesses will require to address domain specific concerns of all stakeholders and legal aspects of e-records being preserved.

Acknowledgements

Expert Committee for Digital Preservation Standards		
Dr. Gautam Bose	Deputy Director General, NIC	Chairman
Dr. Usha Munshi	Head – Librarian, Indian Institute of Public Administration	Member
Mr. U. K. Nandwani	Director, Standardization, Testing and Quality Certification (STQC)	Member
Mrs. Kavita Bhatia	Additional Director, Department of Electronics and Information Technology	DeitY Representative
Mrs. Kavita Garg	Deputy Secretary, Department of Administrative Reforms & Public Grievances	Member
Dr. Ramesh Gaur	Head – Librarian, Jawaharlal Nehru University	Member
Dr. Meena Gautam	Deputy Director, National Archives of India	Member
Mr. N. S. Mani	Microphotographist, National Archives of India	Member
Dr. Dinesh Katre	Associate Director & HOD, Centre for Development of Advanced Computing	Convener

Centre of Excellence for Digital Preservation Team at C-DAC Pune		
Dr. Dinesh Katre	Associate Director & HOD, Human-Centred Design & Computing Group, C-DAC Pune	Chief Investigator of Centre of Excellence for Digital Preservation Project, C-DAC
Mr. Shashank Puntamkar	Joint Director, HCDC Group	C-DAC
Ms. Jayshree Pawar	Project Engineer, HCDC Group	C-DAC
Mr. Saurabh Koriya	Project Engineer, HCDC Group	C-DAC
Mr. Suman Behara	Project Engineer, HCDC Group	C-DAC

Review and Guidance		
Mrs. Renu Budhiraja	Senior Director	DeitY
Mr. V. L. Kantha Rao	President & CEO, NeGD	DeitY
Mr. Gaurav Dwivedi	Director	DeitY
Dr. Ajai Kumar Garg	Additional Director	DeitY
Mr. Bhushan Mohan	Principal Consultant	NeGD, DeitY
Dr. Rajesh Narang	Principal Consultant	NeGD, DeitY
Mr. Rajesh Loona	Senior Consultant	NeGD, DeitY
Mr. T. Hussain	Assistant Director	National Archives of India
Mr. J. K. Luthra	Microphotographist	National Archives of India

The support and guidance received from the members of NeGD, R & D in IT Division, DeitY and PRSG members is duly acknowledged.