Version 1.1

DIGITAL PRESERVATION

Standard Operating Procedure (SOP)

DIGITIZATION * STORAGE * PRESERVATION * SEARCH & RETRIEVAL

E-Committee, Supreme Court of India

24th September 2021

Version 1.1

e-Committee, Working Group on Digital Preservation
(Digitization, Storage, Preservation, Retrieval)

Working Group Members

- **Dr. Dinesh Katre**
  Senior Director & HoD (Scientist G)
  Co-Mission Director – Digital Preservation, C-DAC

- **Kuldeep Singh Kushwah**
  Member (Systems), e-Committee (Convenor)

- **Rishi Prakash**
  Associate Director (Scientist F), C-DAC

- **Ashish Shirdhonkar**
  Senior Technical Director (Scientist F) & HOD, e-Courts, NIC

Table of Contents

List of figures

List of Acronyms

UELMA	Uniform Electronic Legal Material Act
JTC	Joint Technology Committee
NDPP	National Digital Preservation Program
IG	Information Governance
ERM	Electronic Records Management
ESI	Electronically Stored Information
OAIS	Open Archival Information System
JDPS	Judicial Digital Preservation System
JDR	Judicial Digital Repository
DVI	Digitization Verification Information
CNR	Case Number Record
CIS	Case Information System
WORM	Write Ones Read Many
NAS	Network Attached Storage
SOP	Standard Operating Procedure
OCR	Optical Character Recognition
AI	Artificial Intelligence
ML	Machine Learning
ISO	International Standards Organization
BS	British Standard
C-DAC	Centre for Development of Advanced Computing
NIC	National Informatics Centre
NCSC	National Center for State Courts

ULC Uniform Law Commission

NIST National Institute for Standards & Technology

1.Version Control

Version No.	Date dd/mm/yy	Section Changed / Suggestion reference	Change No.	Brief description
1.0	1 Dec. 2020	Executive Summary	1	Data size projection included / discussed in brief
	2 Dec. 2020	Data Protection Strategy	1	Data protection diagram and table is enhanced
	9 Dec. 2020	Budget	1	Budget revised to accommodate Bombay High Court's revised storage requirement
	2 Jan. 2021	Need	1	Data centre survey findings incorporated
	2 Jan. 2021	Executive Summary	2	Data centre survey findings incorporated
	12 March 21	Need	1	Data centre survey findings incorporated
1.1	19 July 2021	Suggestion 1	1	Refer 11.11
	19 July 2021	Suggestion 4	1	Addition of search and retrieval techniques. Refer 8.3.3
	19 July 2021	Suggestion 15	1	Setting up of digitization offices at district courts. Refer 11.9.2
	20 July 2021	Suggestion 16 and 41	1	Training by state judicial academy. Refer 16.2
	20 July 2021	Suggestion 20	1	Rank and designation of officer in charge. Refer. 11.9.2

	20 July 2021	Suggestion 43	1	Need of permanent cadre of technical manpower. Refer. 11.9.2
	21 July 2021	Suggestion 30	1	Refer. 11.9.2
	21 July 2021	Suggestion 31	1	Refer. 11.9.4
	21 July 2021	Suggestion 37	1	Refer. 11.9.3
	24 July 2021	Suggestion 58 and 60	1	Refer. 11.6
	24 July 2021	Suggestion 75	1	Refer. 11.7
	28 July 2021	Policy to separate data on cloud and offline media	1	Refer 4.2.1
	28 July 2021	Suggestion 26	1	Refer 4.3.1
	28 July 2021	Suggestion 72	1	Refer 8.3.3
	28 July 2021	Locations for DR	1	Refer 9.4
	28 July 2021	Table on estimated data storage	1	Figures are updated as per the additional inputs received
	28 July 2021	Table on scope of digitization	1	Figures are updated as per the additional inputs received
	28 July 2021	Table on budget	1	Revised as per the changes in scope of digitization and additional requirements

2.Executive Summary

The federal organization of the Indian judiciary must be kept in mind, while proposing an approach for effective management of the humongous volumes of digital records produced and managed by the courts. Every High Court has superintendence over all courts and tribunals under its territory. It also exercises original, appellate and revisional jurisdiction. The Supreme Court has original, appellate and advisory jurisdiction. Its exclusive original jurisdiction extends to any dispute between the Government of India and one or more States; or between the Government of India and any State or States on one side and one or more States on the other; or between two or more States, if and insofar as the dispute involves any question (whether of law or of fact) on which the existence or extent of a legal right depends. In addition, Article 32 of the Constitution gives an extensive original jurisdiction to the Supreme Court in regard to enforcement of fundamental rights. It should be noted that District Courts use regional languages of the State, whereas, language of the High Courts and the Supreme Court of India is English. There are around 25 High Courts and 672 District Courts in India.

The Digital Preservation Standard Operating Procedure (SOP) for the Indian Judiciary has been developed on the basis of international standards and best practices considering the enormous volume of digital records being produced by them. Two rounds of surveys (3 forms) were conducted to collect information from the High Courts and District Courts for understanding the scope, volume and current state of digital preservation.

Estimated Data Storage Requirements			
Sr.No	High Court Location	AIP (Disposed case Digitization) on Cloud (TB)	DR Site DataSize (TB)
1	NorthEast States	191.14	582.36
2	Orissa	511.19	1,563.06
3	Delhi	440.14	1,193.69
4	Punjab	1,287.74	3,945.30
5	Karnataka	1,599.95	4,901.19
6	Madhyapradesh	1,396.99	4,094.27
7	Uttarakhand	141.27	431.98
8	Uttarpradesh	2,227.64	6,738.22
9	Kerala	208.30	638.17
10	Rajasthan	1,975.47	6,047.16
11	Jharkhand	553.27	1,682.89
12	Chattisgarh	503.88	1,457.07
13	Madras	1,200.28	3,676.14
14	Andhra-Pradesh	1,404.59	4,303.28
15	Bihar	2,692.46	8,248.99
16	Calcutta	3,132.34	9,555.38
17	Telangana	984.12	3,010.95
18	Bombay	6,758.39	20,705.92
19	Himachalpradesh	275.57	842.63
20	Jammu & Kashmir	87.22	267.23
21	Gujarat	1,859.37	5,761.54
	Total	**29,431.32**	**89,647.42**

Based on the information received from 21 High Courts during Survey - I, only 73,44,57,063 pages, which constitutes 5.9% of the total number of legacy pages, have been digitized by the courts. As per the information received during Survey – II and III from 21 High Courts, around 12,42,93,90,000 pages require digital preservation, which mainly include legacy records (disposed cases). According to the estimated projections for data storage in the table given above, almost every High Court along with the district courts under its administrative control will require 1-7 petabytes of cloud storage with efficient search and retrieval mechanisms in the near future. The High Court with smaller storage requirements i.e., 25 to 300 TB may be provided with basic server + storage-based solution.

The North East states are grouped together on the basis of survey response received from courts of Assam, Arunachal Pradesh, Mizoram, Sikkim, Manipur Tripura, Nagaland and Meghalaya.

The SOP plans to establish dedicated Judicial Digital Repositories (JDRs) at the High Court level to manage and preserve the digital records of the High Court as well as the district courts under its administrative control. Similarly, the Supreme Court of India is also required to establish a Judicial Digital Repository (JDR) to meet its own digital preservation requirements.

The Judicial Digital Repositories (JDRs) have to be audited and certified as per the ISO 16363 for their overall trustworthiness and reliability in the long-term to ensure legal admissibility of digital records. ISO 16363 Audit and Certification of Trustworthy Digital Repositories also requires compliance with ISO 27001 and other ancillary standards. This approach ensures that there is appropriate level of decentralization at the High Court level and involvement of stakeholders in effective data management with a long-term vision.

Therefore, the SOP recommends defining and implementing a comprehensive Information Governance (IG) policy considering the volume of e-filing and production of digital records by the courts. It should include clear guidelines on Electronic Records Management (ERM) as per the ISO 15489 and digital preservation as per ISO 16363. The SOP provides comprehensive technical specifications, open standard based file formats and guidelines for digitization. It provides guidelines for preparation of Submission Information Package (SIP) and method for transfer of digitized records to JDR. The SOP provides a comprehensive Data Protection Strategy (DPS) along with step-by-step implementation guidelines. Presently the scope of digitization is kept limited to documents but eventually the same infrastructure can be gradually scaled to preserve other forms of digital data including electronic evidence. Standardized metadata parameters are also defined to ensure efficient searchability, classification and interoperability across JDRs. The SOP also provides a new proforma, which can be used by officer in-charge of digitization, for issuing a certificate to assure lawful control over the computers used for digitization and the integrity of digitized records as required under section 65B of the Indian Evidence Act.

Specially designed Judicial Digital Preservation System (JDPS) and Access Portal developed as per the ISO 14721 Open Archival Information System (OAIS) Reference Model should be used for preservation, search and retrieval, which is necessary for obtaining ISO 16363 certification. Judicial Digital Preservation System and the Access Portal should be implemented in a uniform way to ensure interoperability across all Judicial Digital Repositories (JDRs). The Access Portal will be customized for multilingual search & retrieval as per the requirement of different states. The Data Disaster Recovery (Data DRs) will be managed by the High Courts on reciprocal basis for each other with proper measures for accountability.

This flagship project needs to be implemented in collaboration with a specific High Court to establish model infrastructure, systems and best practices that can be developed and replicated across remaining High Courts. International collaborations should be established

with organizations like NCSC, NIJ and NIST based in the US to learn about technologies, standards and best practices of international judicial organizations in developed countries.

The SOP provides a comprehensive digital preservation budget covering the requirements of all the High Courts and District Courts. Based on the information received from High Courts, a budget of Rs. 2677.76 Crore, spreading across 5 years has been estimated. Overall cost turns out to be 0.86 paise per page. It includes the cost of digitization, metadata creation, curation, establishment and management of Cloud Infrastructure by the High Courts, JDR management, audit & certification, development, deployment and technical support for digital preservation tools and software solutions. Multiple options for availing cloud services or facility management services are provided to the High Courts.

The digital preservation SOP offers insights on how the Judicial Digital Repositories (JDRs) built as per the international standards and best practices, could be leveraged for AI / ML based applications to provide intelligent and accurate decision support and efficient methods to accelerate the justice delivery for common citizens in the near future.

3.Objective

The Digital Preservation Standard Operating Procedure (SOP) focuses on the following objectives.

- Digital preservation of judicial records to cover digitized as well as born digital data (computer generated electronic records) and address the looming challenges and threats of rapid technological obsolescence.
- Envisage an interoperable implementation model and trustworthy mechanism for digital preservation of records for the Supreme Court of India, 25 High Courts and 672 district courts for boosting efficiency, consistency and exchange of records / data between judicial entities.
- Provide coverage to all major aspects of digital preservation, as defined by the international standards. The document incorporates select ISO standards which are globally accepted, auditable, recognized as best practices and are readily available.
- Create Judicial Digital Repositories (JDRs), which will be audited and certified as per ISO 16363 for trustworthiness, interoperability and reliability.
- The certified JDRs could be leveraged for building AI /ML based intelligent applications to increase efficiency and accuracy of justice delivery system.

Note: -

As per the universally recognized nomenclature "Digital Preservation" is a single term which encompasses scanning / digitization, preservation, storage, search & retrieval etc. It not only covers digitized data but also the born-digital data (most commonly referred as computer-generated e-records), which is far more vulnerable on account of technological obsolescence, requiring even greater attention and sustained efforts for preservation.

4.Scope

4.1. Need

The need of digitization / digital preservation can be summarized based on the survey reports received from various High Courts and District Courts. Refer Annexure I for sample digital preservation survey forms, which are specially designed and circulated across all the High Courts for data collection.

High Court wise Existing Digitization Status and Linguistic Diversity				
Sr. No.	High Court Location	Total Number of Digitized Pages	Size of Digitized documents (TB)	Languages
1	Gauhati High Court Itanagaar, Auranchal Pradesh	0	0	English
2	Calcutta High Court West Bengal	1,22,00,000	20	English, Bengali
3	Chattisgarh High Court of Jurisdicture at Allahabad	19,68,00,000	42	English, Hindi, Urdu
4	Delhi High Court, New Delhi	17,90,00,000	75	English, Hindi, Urdu
5	Guwahati High Court, Assam	2,92,17,338	15.4	English
6	High Court of Himachal Pradesh	75,34,000	0.8	English, Hindi, Punjabi, Urdu
7	Gauhati High Court Kohima Bench	2,80,000	0.08	English
8	High Court of Madhya Pradesh, Jabalpur	15,40,00,000	90	English, Hindi
9	High Court Meghalaya Shillong	0	0	English, Khasi, Garo, Pnar
10	Gauhati High Court Aizwal Bench, Mizoram	29,867	0.016	English, Mizo
11	Orissa High Court ,Cuttak ,Odhisa	1,22,00,000	1.6	English, Odia
12	High Court of Sikkim, Gangtok	6,83,861	0.057	English, Nepali, Bhutia
13	Madras High Court , Chennai	50,98,000	0.581	English, Tamil
14	High Court of Telangana, Hyderabad	4,01,50,753	2	English, Telgu, Urdu
15	High Court of Uttarkhand, Nainital	1,32,00,000	0.4	English, Hindi
16	Bombay High Court	0	0	English, Marathi
17	Daman Diu High Court	0	0	English, Marathi
18	High Court Jharkhand, Ranchi	5,50,00,000	5.9	English, Hindi, Bangla, Urdu
19	High Court of Karnataka	1,13,22,389	0.308	English, Kanada
20	Manipur Hifgh Court	16,40,855	0.668	English,Manipuri, Hindi
21	Rajasthan High Court	1,61,00,000	2.6	English, Hindi
		73,44,57,063	257.21	

The table above shows that based on the inputs received from 21 High Courts in the Survey Form-I, approx. 73,44,57,063 pages are already digitized and 257.21 TB data is available. Most importantly the diversity of languages used in the legacy records across various courts requires to be noted.

As per our assessment, the High Courts and district courts have digitized only 5.9 % of the total number of legacy records. Therefore, the scope for digitization is very vast if we consider the total volume of legacy records, pending and new cases.

4.1.1. Scope of long-term digital preservation

The scope of long-term digital preservation is limited to disposed of cases. As per the records management standards only those records that have reached the end-of-life cycle (non-current records) are selected for preservation.

4.1.2. Scope of digitization

The overall scope of digitization in the judiciary is much wider as it includes pending and newly registered cases also. These records are current and still in process and therefore they are not considered in the scope of long-term digital preservation.

The table below indicates the number of pages from the legacy records and projection for the next 5 years (with 10% rise every year) including newly instituted cases and new pending cases. Refer Annexure I for the survey forms used for data collection. The data is collected using Survey Forms II and III for page estimation. Total page count is reduced by 40% after weeding out the unwanted records.

Scope of Digitization						
Sr.No	High Court Location	Disposed (Legacy Records)	Pendency	Institution Projection for Next 5 years	Total Pages	Total Pages after Weeding out
1	Arunachal pradesh	67,00,000	21,70,000	1,94,67,675	2,83,37,675	2,56,57,675
2	Gauhati	7,98,00,000	4,85,00,000	40,13,07,333	52,96,07,333	49,76,87,333
3	Manipur	89,00,000	14,00,000	1,51,79,474	2,54,79,474	2,19,19,474
4	Meghalaya	67,00,000	68,00,000	3,41,53,816	4,76,53,816	4,49,73,816
5	Mizoram	49,80,000	5,60,000	1,41,35,885	1,96,75,885	1,76,83,885
6	Nagaland	1,72,00,000	28,00,000	2,27,69,210	4,27,69,210	3,58,89,210
7	Sikkim	38,00,000	16,20,000	1,00,56,401	1,54,76,401	1,39,56,401
8	Tripura	1,08,00,000	6,00,000	66,41,020	1,80,41,020	1,37,21,020
9	Orissa	37,34,00,000	12,41,00,000	30,16,92,038	79,91,92,038	64,98,32,038
10	Delhi	26,75,00,000	7,74,00,000	85,28,96,673	1,19,77,96,673	1,09,07,96,673
11	Punjab	94,34,00,000	16,31,00,000	14,42,04,999	1,25,07,04,999	87,33,44,999
12	Karnataka	1,17,19,00,000	31,50,00,000	75,73,60,861	2,24,42,60,861	1,77,55,00,861
13	Madhyapradesh	95,75,00,000	7,61,00,000	60,81,27,661	1,64,17,27,661	1,25,87,17,661
14	Uttarakhand	10,32,00,000	3,36,00,000	21,82,04,933	35,50,04,933	31,37,24,933
15	Uttarpradesh	1,60,12,00,000	44,68,00,000	1,36,61,52,624	3,41,41,52,624	2,77,36,72,624
16	Kerala	15,26,00,000	11,22,00,000	70,86,91,674	97,34,91,674	91,24,51,674
17	Rajasthan	1,44,54,00,000	20,04,00,000	1,56,25,37,064	3,20,83,37,064	2,63,01,77,064
18	Jharkhand	40,10,00,000	14,20,00,000	79,12,30,061	1,33,42,30,061	1,17,38,30,061
19	Chattisgarh	33,83,70,000	2,56,00,000	20,11,28,025	56,50,98,025	42,97,50,025
20	Madras	87,89,00,000	30,58,00,000	73,71,53,187	1,92,18,53,187	1,57,02,93,187
21	Andhra-Pradesh	1,02,90,00,000	7,86,00,000	1,06,06,65,718	2,16,82,65,718	1,75,66,65,718
22	Bihar	1,97,25,00,000	35,15,00,000	73,43,07,035	3,05,83,07,035	2,26,93,07,035
23	Calcutta	2,28,01,00,000	20,71,00,000	83,20,24,897	3,31,92,24,897	2,40,71,84,897
24	Telangana	71,95,00,000	5,59,00,000	70,58,45,522	1,48,12,45,522	1,19,34,45,522
25	Bombay	4,95,12,00,000	42,30,00,000	1,29,49,98,842	6,66,91,98,842	4,68,87,18,842
26	Himachalpradesh	20,13,00,000	10,14,00,000	28,08,20,262	58,35,20,262	50,30,00,262
27	Jammu & Kashmir	6,39,00,000	7,62,00,000	32,54,09,965	46,55,09,965	43,99,49,965
28	Gujarat	72,49,00,000	12,39,00,000	1,14,69,98,974	1,99,57,98,974	1,70,58,38,974
	Total Pages	20,71,56,50,000	3,50,41,50,000	15,15,41,61,828	39,37,39,61,828	31,08,77,01,828
	Total Pages after Weeding	12,42,93,90,000	3,50,41,50,000	15,15,41,61,828	31,08,77,01,828	

The need of digital preservation can be summarized as under –

- Most of the legacy records are yet to be digitized.
- Many High Courts are still to start the digitization process.
- The existing digitized records are stored in available storage systems.
- Volume of e-filing and born digital records is also growing at an alarming rate.
- Standardized and auditable mechanism for long term digital preservation needs to be introduced.
- The linguistic diversity also poses a major challenge for OCR, search & retrieval and translation.

A separate survey was carried out to study the availability of Data Centre facility / cyber infrastructure of the High Courts. 17 High Courts have responded to the survey. As per the survey, High Courts in Calcutta, Chhattisgarh, Rajasthan, Karnataka and Uttarakhand do not have data center facility. High Courts in Bombay, Guwahati, Patna, Andhra Pradesh, Telangana, Panjab, Jharkhand, Karnataka and Jammu & Kashmir have server rooms comprising of 1 to 4 servers with multiple software applications already running on them. Many have reported that they do not have cooling system in the server rooms. Many of them have very old servers verging on obsolescence. Delhi, Madhya Pradesh and Kerala High Courts have proper data centre facility along with servers running multiple software applications. Delhi High Court has 22 servers, 335 TB SAN storage and 65 software applications. Madhya Pradesh High Court has 200 TB SAN storage. Jharkhand High Court has 128 TB SAN storage. Delhi, Patna, Guwahati and Telangana High Courts have 1 Gbps network connections. Andhra Pradesh High Court has 512 Mbps network connection. Kerala, Madhya Pradesh, Rajasthan, and J & K High Courts have 100 Mbps network connections. Some High Courts have 40 or 20 Mbps or even lower bandwidth or NIL network connections. Barring these minor exceptions, we can conclude that High Courts do not have proper cyberinfrastructure for digital preservation of massive volumes of digitized records. Also, it should be noted that none of the High Courts have setup a cloud but they are using separate physical servers for running software applications.

Sr.No	High Court Location	Data Centre	Servers	Existing Applications	Existing Storage (TB)	AIP on Cloud (TB)	Network	Remark
1	NorthEast States	Yes	2	2	0	557.04	1Gbps	1 of 2013
2	Delhi	No	22	65	335	76.46	1Gbps	Servers : 4 of 2012
3	Punjab	No	15	12	49	2,150.24	300Mbps	Servers : 8 before 2015
4	Karnataka	No	-NA-	5	0	2,302.79	100Mbps	-NA-
5	Madhyapradesh	Yes	26	11	200	1,634.59	500Mbps	Servers : 10 of 2012, 16 after 2015
6	Uttarakhand	No	-NA-	-NA-	0	428.04	-NA-	
7	Uttarpradesh	No	10	9+	21	3,339.20	-NA-	Servers : 4 of 2012, 9 after 2017
8	Kerala	Yes	11	12 +	0	1,846.10	100Mbps	
9	Rajasthan	No	3	3	14.6	1,899.80	68MBps	
10	Jharkhand	Yes	1	1	128	768.01	No	
11	Chattisgarh	No	-NA-	-NA-	16.2	698.05	-NA-	
12	Andhra-Pradesh	Yes	6	6	0	952.98	512Mbps	
13	Bihar	Yes	14	55	40	8.64	1Gbps	Servers : 6 of 2009
14	Calcutta	No	-NA-	-NA-	0	1,120.17	-NA-	
15	Telangana	Yes	12	8	7.5	346.93	1Gbps	Servers : 4 of 2014
16	Bombay	No	2	2	100	12,490.02	50Mbps	
17	Jammu & Kashmir	Yes	1	1	24	252.99	100Mbps	

By and large, it is observed that standardized specifications for digitization, metadata schema, archival systems, repository infrastructure, data protection strategy, disaster recovery, efficient retrievability, sustenance and auditable procedures for long term digital preservation need to be defined. The standardization of all the aspects of digital preservation across all High Courts and district courts will definitely help in achieving greater interoperability between data repositories and benefit the judiciary in overall acceleration of the justice delivery system. A uniform and long-term approach for preservation of digitized as well as born-digital records will certainly reduce the costs in contrast to non-standard and fragmented efforts of digitization.

The SOP guidelines presented in this document can be applicable for digital preservation at Supreme Court of India, High Courts and District and Subordinate Courts.

4.2. Digital Preservation

The following types of born-digital or digitized records require long term preservation:

- Disposed records (daily disposal);
- Disposed records (before digitization initiated); and
- Electronic documents created (digitally signed) by different stakeholders

Permanent records and the records requiring long term retention, after they have reached the end-of-life cycle, may be considered for digital preservation in the Judicial Digital Repository (JDR). All the High Courts have their own records retention schedules / policies.

Therefore, the scope of records to be brought under the purview of digital preservation may be decided by the respective High Courts.

4.2.1. Policy for digital records on cloud storage and on offline media

The High Courts need to formulate a policy to segregate the digitized records for optimizing the cloud storage based on the usage given below:

A. **Digitized records to be stored on active / online cloud storage based on frequency or importance or likely use**

B. **Digitized records that are less likely to be used to be stored on offline storage media**

   In case of digitized records stored on offline storage media, the cataloging metadata of the record and storage media should be maintained in the Judicial Digital Preservation System (JDPS) for retrieval as and when required.

4.3. Digitization

The various types of archival material/data in the courts comprise of data or judgments/orders and other records collectively called as "COURT Holdings", or simply "artifacts." The following types of holdings are being considered for digitization:

- Fresh filed cases.
- Documents received in pending cases.
- Pending records
- Disposed records (daily disposal).
- Disposed records (before digitization initiated).
- Electronic documents created (digitally signed) by different stakeholders.

4.3.1. Active and passive component of pending records

In case of pending records, there is a need to segregate active and passive ingredients of a file. Supreme Court has carried out an exercise of defining active and passive data of a case. The findings of that exercise can be the basis for adopting uniform standards of active and passive data. A separate system may be necessary for managing the pending records considering that it involves both active and passive records.

5.International & National Scenario

5.1. JTC Report on Electronic Records Preservation and Disposition Plan by NCSC

The Joint Technology Committee (JTC) of the National Center for State Courts (NCSC) in its report released and adopted in 2014 discusses the question of development of an Electronic Records Policy at length and provides recommendations for Digital Preservation Policy for the state courts in the United States of America. The JTC report also mentions that there is a need for audit and control of digitally preserved records through ISO Certified trustworthy digital repositories.

5.2. Uniform Electronic Legal Material Act (UELMA) in the United States

The Uniform Electronic Legal Material Act (UELMA) is a model Act drafted and approved by the Uniform Law Commission (ULC) in the United States to create standards for authenticating and preserving digital legal documents, such as official statutes, codes, regulations and decisions. The model Act was approved by the ULC in July 2011. So far twelve states in the US have since passed legislation based on the model act.

UELMA recognizes that use of digital information formats has become fundamental and indispensable to the operation of state government. This Act addresses the critical need to manage electronic legal information in a manner that guarantees the trustworthiness of and continuing access to important state legal material. Technology changes quickly enough, thus State Governments must address this issue as existing electronic legal information is already in danger of being lost. Such a uniform act has been drafted to allow state governments to develop similar systems of authentication and preservation, aiding the free flow of information across state lines and the sharing of experiences and expertise to keep costs as low as possible.

The white paper published by UELMA Preservation Group in 2018 prescribes the need to adopt the mechanism of Trustworthy Digital Repository and Open Archival Information System for preservation of electronic legal records, which has been recommended in this SOP too.

5.3. British Standard on Legal Admissibility of Electronic Information

BS 10008 is the British Standard that outlines best practice for the management and storage of electronic data. It is designed to help verify and authenticate all information to avoid the legal pitfalls of data storage. BS 10008 outlines best practice for transferring electronic data

between systems and migrating paper records to digital files. It also gives guidelines for managing the availability and accessibility of any records that could be required as legal evidence. The latest version of this standard has been released in May 2020.

5.4. National Digital Preservation Program (NDPP) by MeitY

Ministry of Electronics and Information Technology (MeitY), Government of India initiated the National Digital Preservation Program (NDPP) in 2010. The 'National Study Report on Digital Preservation Requirements of India' was prepared by involving the stakeholders from various domains with recommendations from national and international experts. Subsequently, as per the recommendations, in 2011, Centre of Excellence for Digital Preservation was sanctioned as the flagship project spearheaded by C-DAC Pune along with C-DAC Noida to work on pilot digital repositories in diverse domains such as archives, cultural heritage, judiciary and e-governance.

As a part of this project, C-DAC developed the Open Archival Information System (OAIS) as per the ISO 14721 Reference Model for managing the massive audiovisual digital archive established by IGNCA, which received world's 1st ISO 16363 Trustworthy Digital Repository status after due auditing and certification by Primary Trustworthy digital Repositories Authorization Body (PTAB), UK in 2017. C-DAC has designed the technical architecture for the repository, conducted digital preservation training for the IGNCA staff and supported the audit process. The IGNCA's NCAA Project Team took tremendous efforts in coordination with 21 partner institutes, evolving the data and metadata specifications, selection and digitization of audiovisual materials and quality control through-out the digitization of almost 2 petabytes of audiovisual data.

This information has been provided to establish the availability of the digital preservation expertise, knowhow of the standards and technological capabilities within India.

The 2nd certification was achieved by the digital repository of United States Government Publishing Office (USGPO) in 2019.

6.Legal Framework

National Policy and Action Plan for Implementation of information and communication technology (ICT) in the Indian Judiciary–2005 was submitted by e-Committee, Supreme Court of India, with a vision to transform the Indian Judiciary by ICT enablement of Courts. As per this policy, e-Courts are being established to make justice delivery system affordable and cost-effective. The computerization of judicial procedures is producing massive volumes of e-records. The digitization of legacy records with the courts is also resulting in humongous volumes of data.

The risks associated with digital data due to rapid technological obsolescence are recognized globally. The obsolescence of digital records and the evidentiary proofs can create problems in administrative, judicial and legislative functions in addition to loss of valuable information, intellectual property and heritage. Therefore, it is necessary to ensure that the digital records, which require to be retained for long duration are preserved as per the international best practices and standards.

In addition, the relevant extracts from the Indian Evidence Act 1872, IT Act 2000, IT Act Amendments 2008, RTI Act 2005 and Public Records Act 1993 have been reproduced here, which make it mandatory to preserve the digital records. The specific sections and clauses from these acts are referred in this section.

Fig. 1. Legal requirements providing for establishment of a suitable mechanism for digital preservation

6.1. Indian Evidence Act, 1872

The section 65B of Indian Evidence Act is reproduced here to highlight the conditions defined for admissibility of electronic records.

Section 65B. Admissibility of electronic records. —

(1) Notwithstanding anything contained in this Act, any information contained in an electronic record which is printed on a paper, stored, recorded or copied in optical or magnetic media produced by a computer (hereinafter referred to as the computer output) shall be deemed to be also a document, if the conditions mentioned in this section are satisfied in relation to the information and computer in question and shall be admissible in any proceedings, without further proof or production of the original, as evidence or any contents of the original or of any fact stated therein of which direct evidence would be admissible.

(2) The conditions referred to in sub-section (1) in respect of a computer output shall be the following, namely: —

(a) the computer output containing the information was produced by the computer during the period over which the computer was used regularly to store or process information for the purposes of any activities regularly carried on over that period by the person having lawful control over the use of the computer;

(b) during the said period, information of the kind contained in the electronic record or of the kind from which the information so contained is derived was regularly fed into the computer in the ordinary course of the said activities;

(c) throughout the material part of the said period, the computer was operating properly or, if not, then in respect of any period in which it was not operating properly or was out of operation during that part of the period, was not such as to affect the electronic record or the accuracy of its contents; and

(d) the information contained in the electronic record reproduces or is derived from such information fed into the computer in the ordinary course of the said activities.

(3) Where over any period, the function of storing or processing information for the purposes of any activities regularly carried on over that period as mentioned in clause (a) of sub-section (2) was regularly performed by computers, whether—

(a) by a combination of computers operating over that period; or

(b) by different computers operating in succession over that period; or

(c) by different combinations of computers operating in succession over that period; or

(d) in any other manner involving the successive operation over that period, in whatever order, of one or more computers and one or more combinations of computers, all the computers used for that purpose during that period shall be treated for the purposes of this section as constituting a single computer; and references in this section to a computer shall be construed accordingly.

(4) In any proceedings where it is desired to give a statement in evidence by virtue of this section, a certificate doing any of the following things, that is to say, —

(a) identifying the electronic record containing the statement and describing the manner in which it was produced;

(b) giving such particulars of any device involved in the production of that electronic record as may be appropriate for the purpose of showing that the electronic record was produced by a computer;

(c) dealing with any of the matters to which the conditions mentioned in sub-section(2) relate and purporting to be signed by a person occupying a responsible official position in relation to the operation of the relevant device or the management of the relevant activities (whichever is appropriate) shall be evidence of any matter stated in the certificate; and for the purposes of this sub-section, it shall be sufficient for a matter to be stated to the best of the knowledge and belief of the person stating it.

Note:-

As required by section 65B, the proforma for issuing a certificate ascertaining the lawful control over the computers and monitoring of the digitization activities is provided in chapters 16 and 17 of this SOP, which may be helpful in establishing the evidentiary value of digitized records.

6.2. Information Technology Act 2000

IT Act 2000 specifies the requirements for retention of electronic records (section 7) as under.

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;

6.3. Information Technology Act Amendment 2008

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.
- 67C Preservation and Retention of information by intermediaries

  (1) Intermediary shall preserve and retain such information as may be specified for such duration and in such manner and format as the Central Government may prescribe.

  (2) Any intermediary who intentionally or knowingly contravenes the provisions of sub section (1) shall be punished with an imprisonment for a term which may extend to three years and shall also be liable to fine.

6.4. IT ACT Notification, GSR 582

As per the IT Act Notifications GSR 582, the e-record making 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

Note:-

The notification of General Statutory Rules (GSR) 582 notified under the IT ACT are extremely important as most of the software systems being used by the Judiciary or the associated agencies are producing born digital documents in proprietary PDF formats, which are not suitable for preservation. Judiciary needs to instruct strict compliance with PDF/A-1a or PDF/A-1b profile for digitally signed documents or any other digitized documents provided by associated agencies. The software systems should immediately comply with the requirements of "preservability" as stated in the GSR notification.

6.5. Right To Information ACT 2005

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.

6.6. Public Records Act 1993

The Public Records Act 1993 makes it mandatory that every record creating agency of the central government, any ministry, department or office of the Government must provide proper arrangement, maintenance and preservation of public records. Section 2 (e) of this Act clearly mentions that material produced by a computer or by any other device produced by the record creating agency must be preserved.

In conclusion, the existing legal framework makes it mandatory to undertake the following activities –

- Consider digital information as records requiring preservation
- e-Records must be produced in preservable formats
- Apply the records retention policies to electronic records
- Preserve the digital information
- Protect the evidentiary value of e-records
- Ensure safe and secure custody of the digital records
- Audit requirements are applicable to e-records and digital documents
- Facilitate efficient access to digital records from anywhere in the country

7. Certified Judicial Digital Repositories

7.1. Information Governance (IG) Strategy

Information Governance is a set of multi-disciplinary structures, policies, procedures, processes and controls to manage information at an enterprise level that supports an organization's current and future regulatory, legal, risk, environmental and operational requirements. Digital preservation of records is an integral part of the Information Governance (IG) Strategy of any organization. The judiciary needs to create model policies for Information Governance which can be uniformly applied across courts.

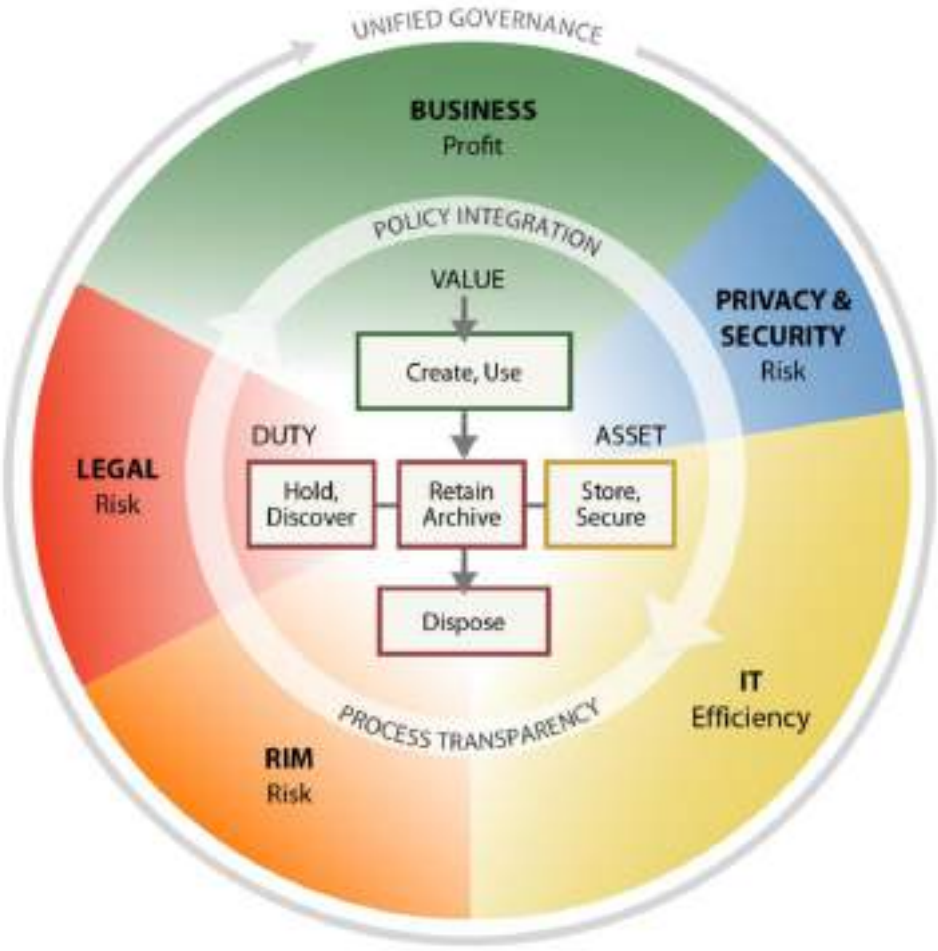


Fig. 2. Information Governance Reference Model (IGRM) Model

(The generic IGRM Model is applicable for all organizations. Courts to interpret "business" as "judicial activities" and "profit" as "benefits" gained towards its objective. Image Courtesy EDRM.net)

7.2. Need for Electronic Records Management (ERM)

As per the records management standard ISO 15489, there is a clear distinction between current and non-current 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 (Case Information System) or in the data center for live transactions.

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.

It is important to note that e-records creation system and digital preservation system are two different systems with a distinct focus and role which should not be mixed.

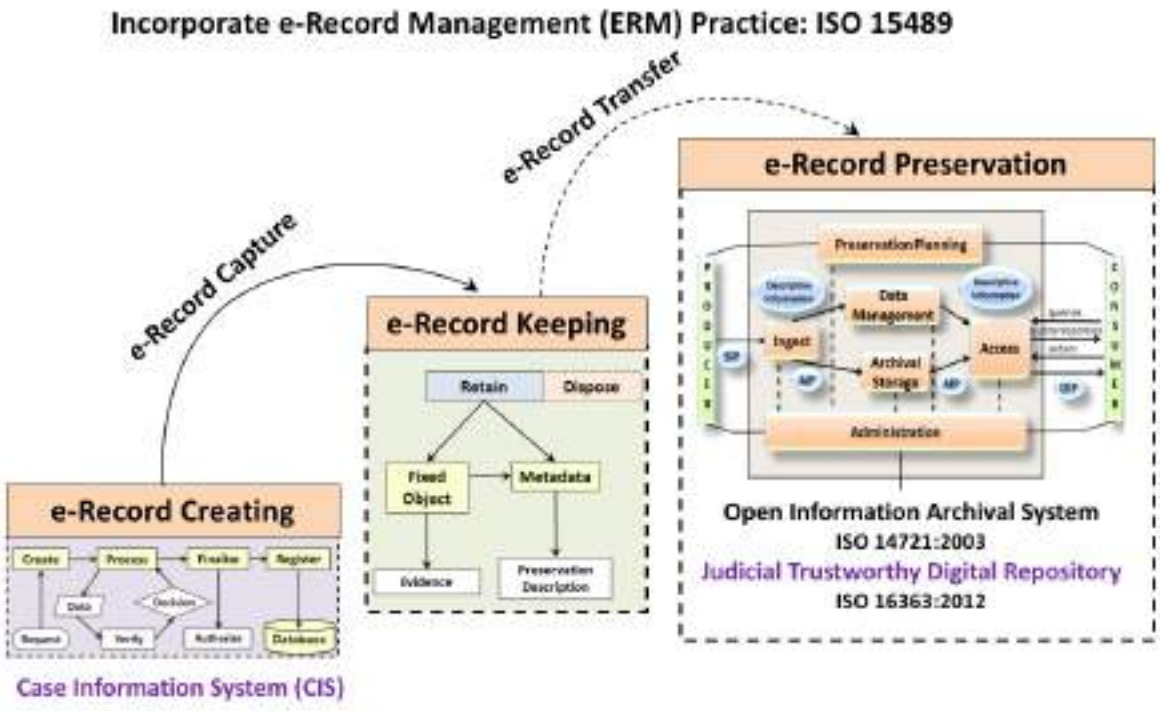


Fig. 3. The Lifecycle of Records Management

(Image Courtesy PRoPeR Best Practices & Guidelines, eGov.DP.01-01 Version: 1.0 December, 2013, MeitY)

7.3. Call for Attention

Advanced Information & Communication Technologies are significantly helpful in the modernization of judicial procedures. ICT is regarded as the key Information Governance (IG) stakeholder for the courts. While it is so critical to leverage the benefits of ICT based solutions for acceleration of justice to citizens, the potential threats to judicial electronic records must also be recognized and necessary measures for long term digital preservation of judicial electronic records needs to be implemented and institutionalized on priority. The judicial Records and Information Management (RIM) is critical irrespective of whether it comprises e-files /born-digital records produced by software systems or the digitized copies of legacy paper-based records or digitized microfilms. The potential threats for judicial digital records include –

- Digital obsolescence due to rapidly changing technologies
- Proprietary file formats
- Data corruption
- Storage media failure
- Computer virus

Therefore, Judicial Digital Repositories (JDRs) duly audited and certified as per ISO 16363 become the key mechanism for governing the information / digital records of long-term value. The figure 4. shows various layers of activities involved in the establishment of trustworthy digital repository.

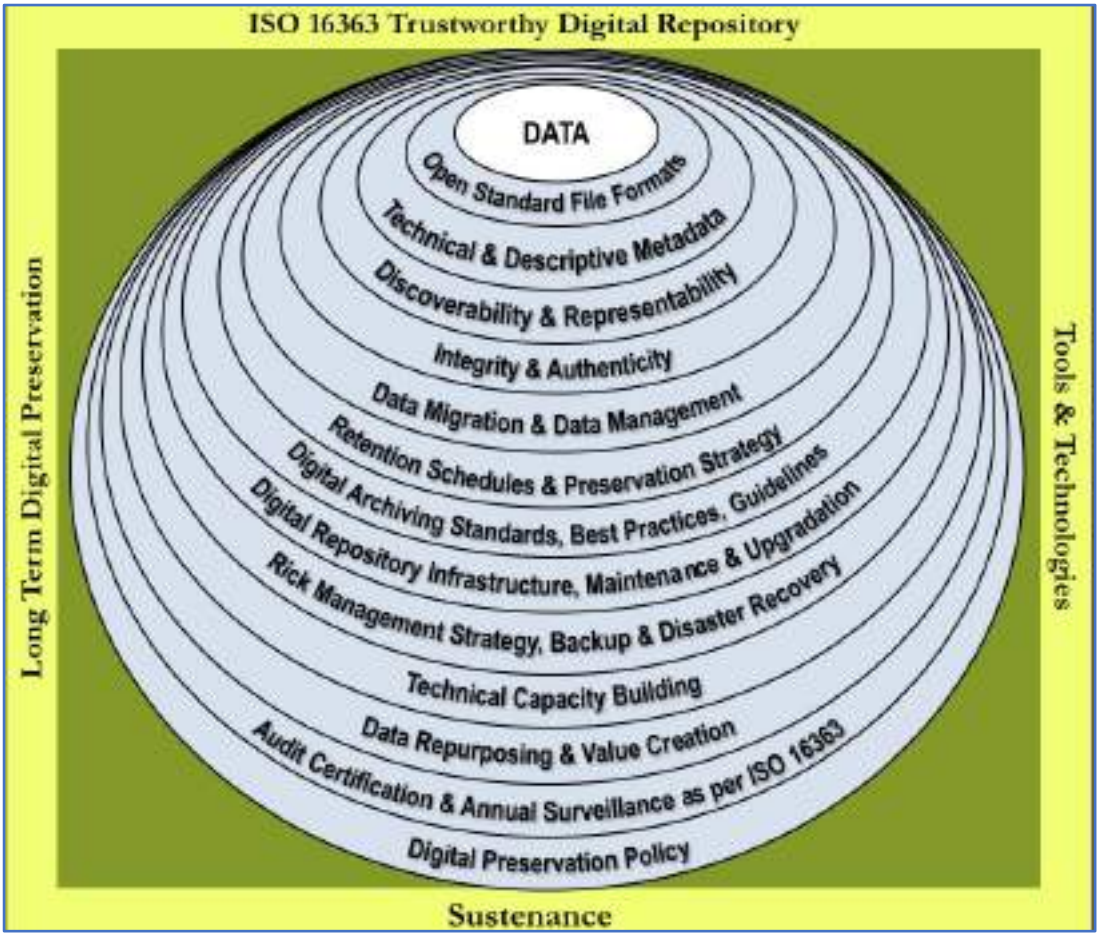


Fig. 4. Layers of ISO 16363 Certified Trustworthy Digital Repository

(The layered representation of Trustworthy Digital Repository, Image Courtesy Proceedings of C-DAC/APA International Conference on Developing Trustworthy Digital Repositories for Digital Preservation, 2014)

7.4. ISO 16363 Certified Trustworthy Digital Repository

A digital repository has the responsibility for long term preservation of digital resources, as well as making them available to communities agreed upon by the depositors of 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. It provides a framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access.

The Trustworthy Repository Audit & Certification (TRAC) metrics are split into three subject groups:

- Organizational Infrastructure - the repository's administrative, staffing, financial, and legal functions
- Digital Object Management - the handling of digital objects from ingests to access
- Technology, Technical Infrastructure, and Security - the technology used to handle ingested objects

Third party audit is conducted by the accredited audit organization, which provides certification after successful completion of the audit. After the certification is received, the digital repository is audited on yearly basis to retain its validity. This approach is extremely helpful in protecting and preserving the digital repositories over a long period.

7.5. Digital Preservation Policy for Courts

Therefore, in order to avoid fragmented and short-lived efforts, Supreme Court of India and High Courts have to define and adopt a comprehensive digital preservation policy, with a high level of commitment to establish and sustain Judicial Digital Repositories duly certified as per ISO 16363 along with the supporting ecosystem for long term digital preservation of judicial records.

7.6. Ecosystem for Trustworthy Judicial Digital Repositories

The ecosystem for Trustworthy Judicial Digital Repositories include -

- State-of-the-art digital repository infrastructure in terms of data centre environment, cloud infrastructure, storage, disaster recovery site, high speed network connectivity
- State-of-the-art software tools and systems necessary for digital preservation, data processing, data migration, integrity and authenticity, search and retrieval, e-discovery and annotation.
- Digital preservation and information security best practices and guidelines
- Access control as per the designated users of judicial digital repositories
- Open and standard based data format specifications
- Cataloging, descriptive and technical metadata standards to enable proper representation and comprehension of digital records
- Well-defined data flow mechanisms to link related business processes and to ensure proper data deposits, preservation and retrieval
- Qualified & trained human resource for managing judicial digital repositories

7.7. Trained human resource for JDR

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.

- **System administrators**

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 system / storage / network administration and cloud management.

7.8. Digital Preservation Standards

The Judicial Digital Repository infrastructure, High Courts and the JDR management staff should collectively gear up the capacity for comply with international standards and best practices related to digital preservation as under -

- ISO 16363 Audit & Certification of Trustworthy Digital Repositories
- ISO 14721 Open Archival Information System (OAIS) Reference Model
- ISO 13008 Digital Records Conversion & Migration Process
- ISO 13028 Implementation Guidelines for Digitization of Records
- ISO 15489 Records Management
- BS 10008 Evidential Weight & Legal Admissibility of Electronically Stored Information (ESI)
- ISO 27001 Information Security Management

The list of international standards needs to be updated regularly for keeping pace with new technologies and the evolving nature of best practices. Also, the relevant national level digital preservation standard and guidelines notified by MeitY such as eGOVPID Metadata Dictionary & Schema and Production of Preservable e-Records (PROPeR) must be suitably incorporated as these are designed to meet the requirements of above listed ISO standards.

7.9. Digital Preservation Planning

The judicial digital repositories to regularly prepare detailed digital preservation plans with consideration for data volume estimation, threat assessments, data migration requirements, refreshing of storage media, infrastructure upgradation, manpower and capacity building, risk management / mitigation, disaster recovery and budget provisions for sustenance.

7.10. Audit & Certification

In order to ensure trustworthiness, the Judicial Digital Repositories to obtain ISO 16363 certification from the accredited third-party organization through regular surveillance audits.

It must be noted that ISO 16363 builds comprises of many ancillary standards along with ISO 14721 OAIS Reference Model at its core.

7.11. Advantages

Establishment of Certified Judicial Digital Repositories will provide following advantages and benefits to Indian Judiciary.

- Duly audited and certified (as per ISO 16363) Judicial Digital Repositories will enable the Supreme Court of India and High Courts of India in effective management and consolidation of the digital information / digital records (born-digital & digitized both) with proper accountability.

- It will also reduce the scattered efforts to digitize and manage the records locally with inconsistent and non-standard practices.

- Traceability and availability of digital information / digital records of long-term value will significantly improve.

- It will reduce the existence dark archives, which get created out of non-catalogued records packed in boxes and digital data stored on offline / obsolete storage media.

- It will be possible for the judiciary staff to take proper care of the legally sensitive digital records. Very often all information whether sensitive or not is lying in the software systems. The practices of digital preservation regime can help in identifying such valuable information /digital records to ensure protection and preservation.

- It will be possible to develop AI/ML based intelligent decision support and analytics for accelerating justice delivery on top of the Judicial Digital Repositories (JDRs).

- The judicial digital repositories to maximize the benefits of digital records preserved in its custody for strengthening the judiciary, while adhering to the rules and policies governing the data / digital records.
    - Efficient access and availability of digital records
    - Semantic linking and referencing between digital records
    - Improved reliability of digital records
    - Increased confidence in legal admissibility of digital records

7.12. Flagship Project to Establish Judicial Digital Repository

The flagship / lead project to implement Judicial Digital Repository may be undertaken with one of the High Courts on priority. It will help in developing the model infrastructure along with the software tools, and tried & tested procedures, which can be speedily replicated and deployed in other High Courts. It will also help in improving the overall approach for implementation by avoiding the mistakes and lessons learnt. More detailed SOPs and guidelines can be evolved.

7.13. International Collaborations

As a part of the flagship project, international collaborations with organizations like National Centre for State Courts (NCSC), National Institute of Justice (NIJ) and National Institute of Standards and Technology (NIST) in the US may be explored for conducting joint workshops, conferences, study visits and training programs. Such initiative will help the teams to learn from the implementation models in developed countries and give wider exposure to technology trends in judiciary. Organizations like NCSC, NIJ and NIST have been evolving policies and standards for dealing with key topics like technology standards for advancing justice, legal admissibility of e-records, e-evidence preservation and access since long ago.

8.Digital Preservation

8.1. Definition of Digital Preservation

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.

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, corresponding to the original record, along with all associated information necessary for proper comprehension. The e-record has to be preserved in such a way that it will remain accessible, reliable, discoverable, authentic and usable for subsequent reference.

Therefore, archival software and digital preservation tools are required to be developed and deployed. The ISO 14721 OAIS Reference Model has to be adopted by the Indian Judiciary to meet its digital preservation and access requirements.

8.2. ISO 14721 Open Archival Information System (OAIS) Reference Model

ISO 14721:2012 defines the reference model for an open archival information system (OAIS). An OAIS is an archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a designated community. It meets a set of such responsibilities as defined in this International Standard, and this allows an OAIS archive to be distinguished from other uses of the term "archive". The term "open" in OAIS is used to imply that ISO 14721:2012, as well as future related International Standards, are developed in open forums, and it does not imply that access to the archive is unrestricted.

The OAIS Reference Model provides

- a framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access,
- the concepts needed by non-archival organizations to be effective participants in the preservation process,
- a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future archives,

- a framework for describing and comparing different long-term preservation strategies and techniques,
- a basis for comparing the data models of digital information preserved by archives and for discussing how data models and the underlying information may change over time,
- a framework that may be expanded by other efforts to cover long term preservation of information that is not in digital form (e.g. physical media and physical samples),
- expands consensus on the elements and processes for long term digital information preservation and access, and promotes a larger market which vendors can support, and
- guides the identification and production of OAIS-related standards that require to be tailored to meet the domain specific requirements.

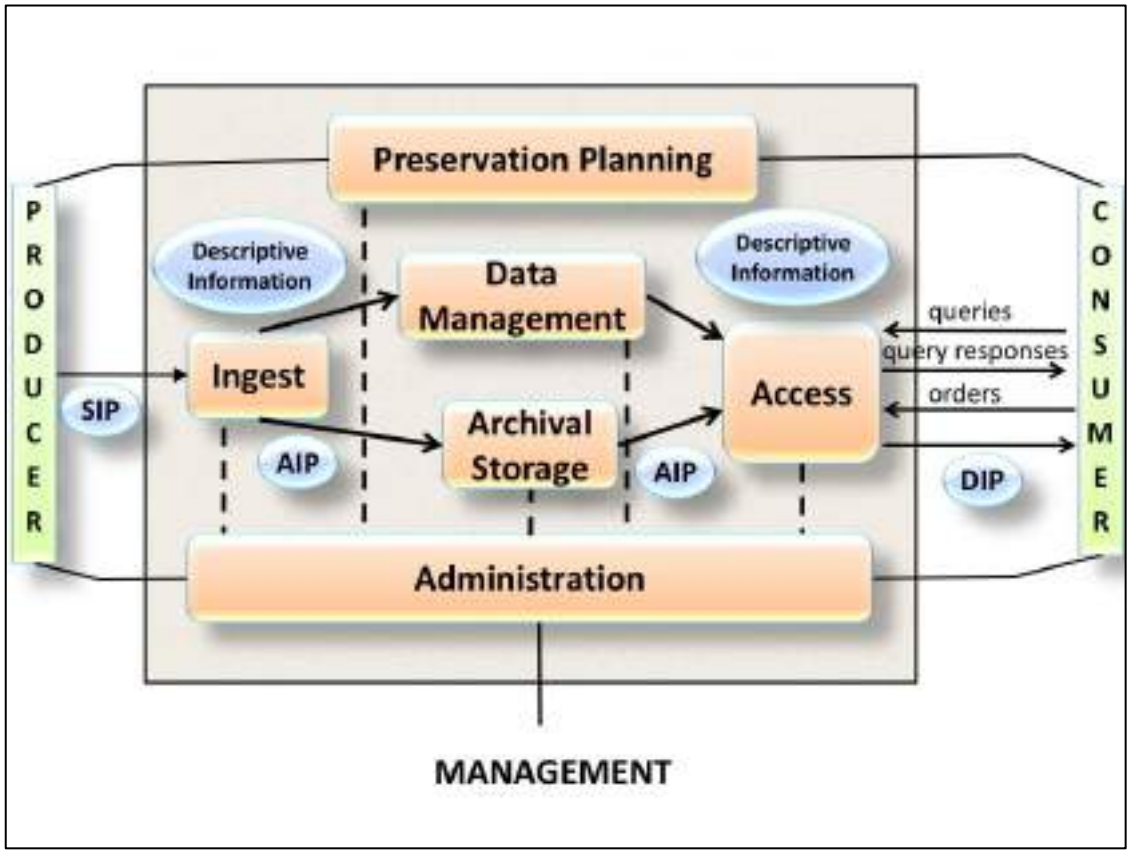


Fig. 5. High-level representation of OAIS Model

(Image Courtesy, ISO 14721:2012 Open Archival Information System (OAIS) Reference Model. OAIS is resourceful with functional guidelines and procedures, wherein each high-level block is exploded with minute operational details.

8.3. Specially Designed Judicial Digital Preservation System (JDPS)

A comprehensive software system based on OAIS Reference Model needs to be especially designed and developed for the Indian Judiciary based on exhaustive study of the requirements of the legal domain. The software can be developed using open technologies with well-documented source code for regular upgradation and consistent version control. The Judicial Digital Preservation System development requires two distinct modules namely Archive Management and Access Portal. A separate module may be considered for managing the pending case records.

8.3.1. Archive Management

The Judicial Digital Preservation System (JDPS) to provide necessary functionalities for Submission Information Package (SIP) validation, metadata creation, automatic metadata extraction, ingest, storage and archive management, Archival Information Package (AIP) and Dissemination Information Package (DIP) configuration, access control, annotation, integrity and authenticity information.

8.3.2. Access Portal

The access portal to allow acceptance of Submission Information Packages (SIPs) and online metadata creation by subordinate courts. It should use reliable mechanisms for user authentication, user management and access control. The portal to provide access to digital records along with descriptive metadata or entire Archival Information Package (AIP) depending on the user privileges. Configurable Dissemination Information Package (DIP) to be provided for the designated users.

8.3.3. Search & Retrieval

The access portal to provide the following types of search mechanisms to retrieve relevant information / digital records from the Judicial Digital Repository-

- Query in English, Hindi and Regional Languages
- Fuzzy search
- Full text search
- Cataloging metadata search
- Parameter-based / facetted search with filtering mechanism
- Boolean search mechanism with options
- Wild card search
- Logical search, proximity search
- Search within search, nested keywords

- Elastic search

Note:-

There are many types of search methods available as mentioned above. It is necessary to select appropriate type of search methods for the given digital repository as the search functionality depends on the quality of data and metadata e.g., full text search depends on availability of extracted text (OCR in case of digitized records) and the accuracy of OCR for the documents being searched. In case of noisy / inaccurate OCR, it only consumes the compute resources without providing greater searchability. Text summarization techniques are used to optimize and reduce the load on full text search.

8.3.4. E-Discovery

With the rise in the volume of Electronically Stored Information (ESI) in Judicial Digital Repositories, e-discovery tools will be necessary to analyze, review and extract meaningful and relevant information from preserved data in terms of digital documents, audio video material and information in other electronic formats. The data collection, documentation and preservation methods have to be properly organized for effective e-discovery.

8.3.5. Scalability

The scalable framework of Judicial Digital Preservation System (JDPS) should allow addition of new ingest procedures depending on type of data.

8.3.6. Access Control

The JDPS should allow configuration of different policies for public and private records with controlled access as per the designated users.

8.3.7. Customizability

The Judicial Digital Preservation System to allow customizability in terms of regional language support, record retention schedules and policies which may differ from court to court within the common framework.

8.4. Benefits of Standardized OAIS Implementation

The digital repositories should be established using the standardized Judicial Digital Preservation System (JDPS) across High Courts and Supreme Court of India. This approach provides following tangible benefits to Indian Judiciary:

- **Interoperability between Judicial Digital Repositories**

  Homogenous and standardized implementation of the judicial digital preservation system across High Courts,district courts and the Supreme Court of India will ensure interoperability and easy exchange of data using standardized protocols and APIs.

- **Scalable Framework**

  OAIS provides a generic reference model with exhaustive documentation of functional procedures. It is designed for scalability and adaptation, as per the domain specific requirements.

- **Reduction of Cost and Duplicate Efforts**

  Standardized implementation of judicial digital preservation system will avoid similar duplicate efforts causing greater expenditure and repetition of inconsistent efforts resulting in incongruent and non-interoperable systems.

- **Greater Stability and Reduced Technical Support**

  It is possible to provide more stable system as there is a distinct pattern in support requirements and errors in its performance. It makes it lot easier to provide technical support and reduce the maintenance cost for the organizations.

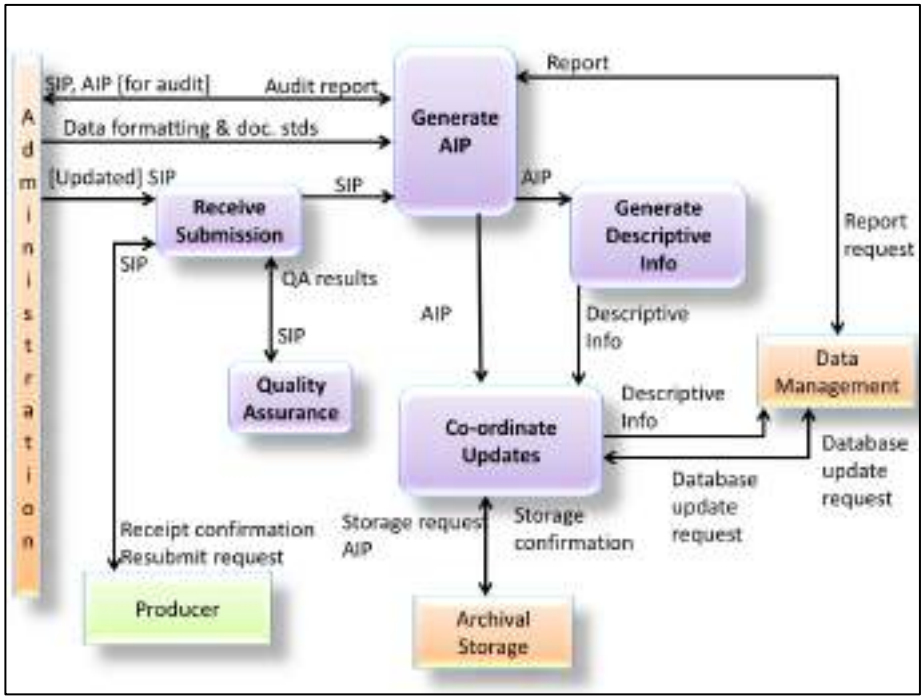


Fig. 6. Functions of Ingest as per OAIS standard

(Image Courtesy, ISO 14721:2012 OAIS Reference Model)

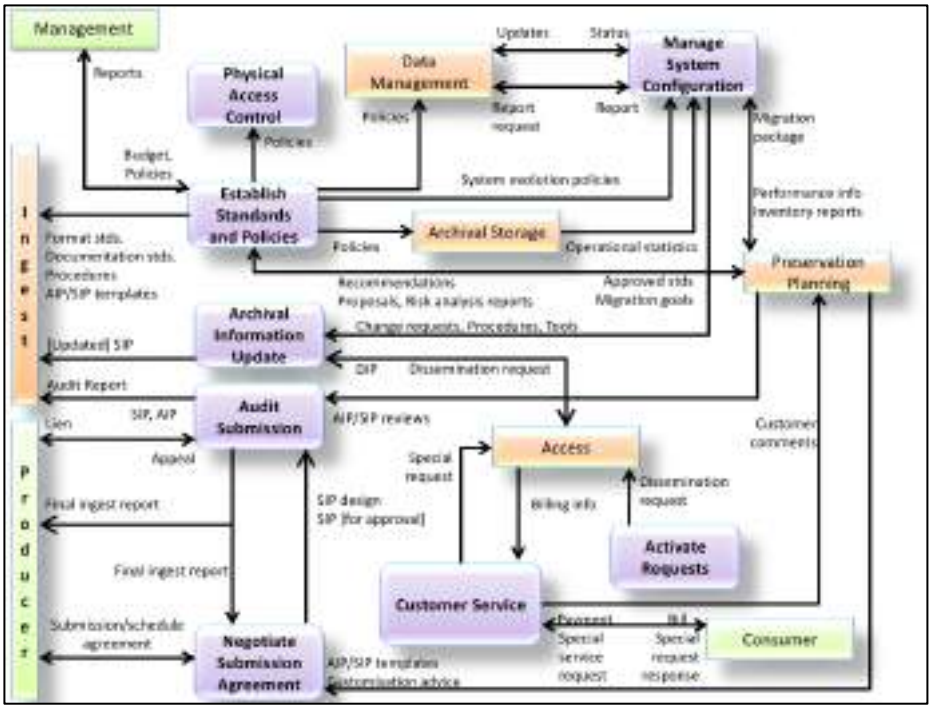


Fig. 7. Administration SOP as per OAIS

(Image Courtesy, ISO 14721:2012 OAIS Reference Model)

9.Implementation Model

9.1. Disadvantages and Problems in Non-Standard DMS implementations

A digitization survey was conducted for knowing the current state and the volume of digitization requirements across different High Courts. Refer to Annexure I for the survey form which was used for deriving the insights. The survey revealed that many courts have massive unmet digitization requirements. It can be estimated that the storage requirement of digitized records may run into several petabytes. Some courts have conducted partial digitization but they don't have any Document Management System (DMS). Some courts are yet to digitize their records. Some are using either open source or proprietary DMS solutions. Let's understand the issues involved in the inconsistent and fragmented approach of implementation.

9.1.1. Open Source DMS Implementations

It is observed that open source DMS software are being used by some courts. Very often open-source solutions get used at the local level because of their free availability. But with such an approach, the metadata requirements, scalability, interoperability, security,

reliability and audit requirements get compromised, which are essential for managing the judicial records. Also, the open source DMS software may not fulfill the requirements of ISO 14721:2012 OAIS Reference Model.

9.1.2. Proprietary DMS Implementations

The commercial DMS solutions may end-up with vendor lock-in, as a result the courts become vendor dependent due to proprietary nature of the system. It is observed that venders store the digital records in closed source format or intentionally create complex technical barriers to ensure dependence of the customer. In such a case, the record owners cannot take out the records or metadata from the DMS for migrating it into another system. In order to achieve this the customer has to pay heavy fees for data migration. Usually, the proprietary DMS solutions do not fulfill the requirements of ISO 14721:2012 OAIS Reference Model. The courts will have to frequently purchase new versions and upgrades from the vendor for maintaining the accessibility of its own data. The non-standard and fragmented DMS implementations can create islands or silos of data repositories. The interoperability between such heterogenous systems poses a major challenge. The non-standard approach may also end-up in wastage of resources, duplication and duplication of efforts.

Based on abovementioned observations, we can conclude that it is not practical and viable to establish, operate and maintain digital repositories at 672 district courts. This requires centralization at the High Court level for effective management. The implementation model proposed in this SOP is illustrated in Fig.8.

9.2. Administrative Organization of Indian Judicial System

Every High Court has superintendence over all courts and tribunals throughout the territories over which it exercises jurisdiction. The Supreme Court has original, appellate and advisory jurisdiction. There are 25 High Courts and approximately 672 District Courts.

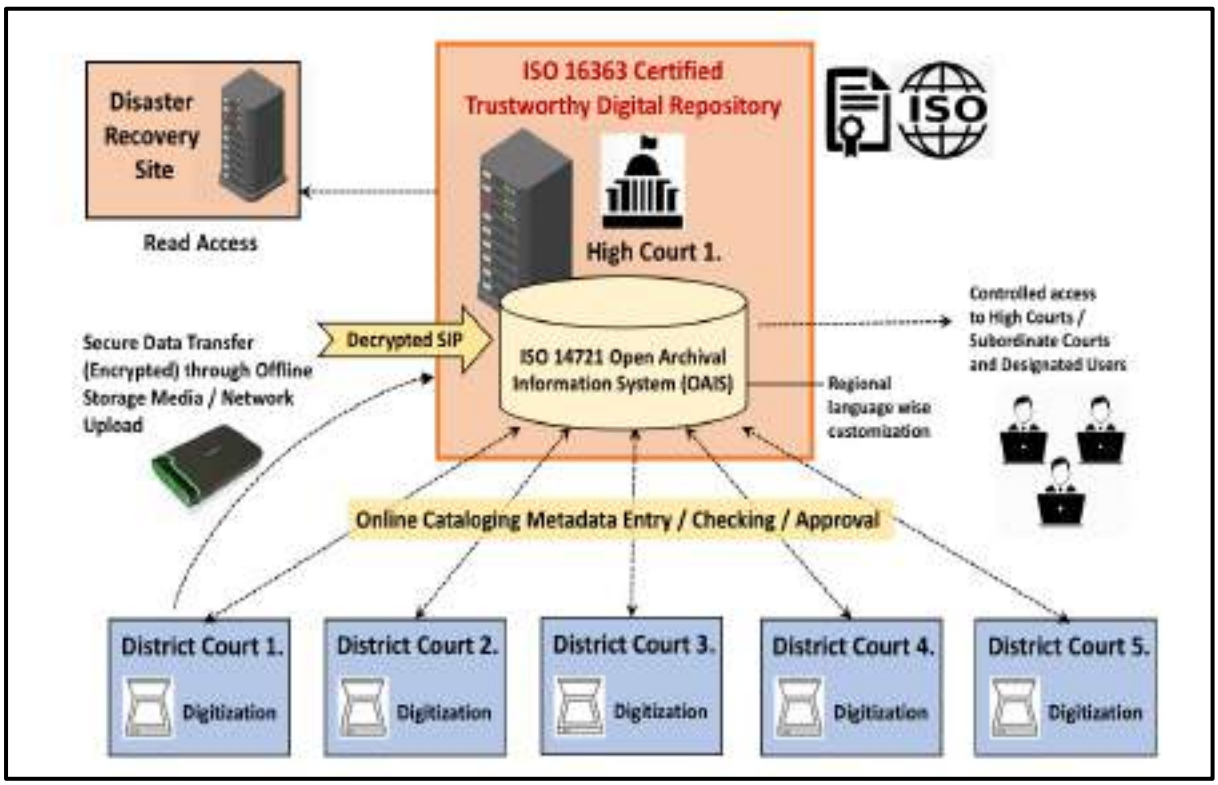


Fig.8 Implementation Model for Courts

9.3. Approach for Implementation

The implementation model is elaborated step-by-step as below.

- The digitization activity to be undertaken at Supreme Court of India, High Courts and District Courts as per the standardized digitization specifications and metadata parameters given in the digital preservation SOP.
- Supreme Court of India and High Courts should establish separate Judicial Digital Repositories (JDR), which should aim to achieve ISO 16363 certification for ensuring the trustworthiness of cyber infrastructure, data management practices, the procedures and digital continuity plans.
- Judicial Digital Repositories (JDR) established at the High Courts should preserve the records produced by High Courts as well as district courts using the standardized Judicial Digital Preservation System (JDPS) which is designed to comply as per the ISO 14721 Open Archival Information System (OAIS) Reference Model.

9.4. Disaster Recovery Sites

The disaster recovery sites should take care of the data requiring long term digital preservation. The DR sites should be established with proper manpower, infrastructure, environmental controls (temperature, humidity) and security for ensuring the safety of data

and storage media. The systems and processes for regular maintenance, refreshing and cataloging of storage media should be put in place.

The High Courts based in low-risk seismic zone II such as Bengaluru, Hyderabad, Jaipur, Ranchi, Allahabad and Bhopal may be considered for establishing the Data DR facilities to take care of the DR requirements of all High Courts.

10. Interoperability between JDRs

The interoperability between Judicial Digital Repositories (JDRs) is very critical for integration and sharing of records between the courts. Therefore, the standardization of metadata, data formats and repository management systems across the judiciary becomes very essential.

10.1. Metadata Interoperability

It is extremely important to standardize the common metadata elements and schema across the Indian judiciary for preservation purpose. It promotes interoperability and provides following advantages.

- The standardization of metadata can boost efficiency of search & retrieval across the digital repositories.
- It also enables the facetted search, category-based classification and sorting using common parameters.
- It also helps in performing unified search over metadata catalogues from various repositories.
- The standardization of metadata scheme can also boost the automation efforts and improve the quality of metadata, which is necessary to understand the record.

On the contrary, it is not possible to integrate, retrieve and share the records with non-standard metadata.

10.2. Data Interoperability

Standardization of open data formats as given in this SOP can provide the following advantages towards data interoperability.

- Reduces or eliminates the threat of file format obsolescence, which is a major risk in non-standard proprietary file formats.
- Data can be easily used and exchanged between the repository systems.
- Reduces expenditure in purchase of proprietary tools and investments in frequent upgradation as per the change of versions.
- Saves your data from vendor lock-in.

10.3. System Interoperability

- A specially designed and standardized digital repository management system for the judiciary can enable seamless integration and exchange of records between the courts.

- It also makes it easy to track down common problems which can be addressed by a shared solution. Also, it is possible to design and use common protocols and APIs across repositories.
- It will be easy for State Judicial Academy (SJA) to train officials due to standardization of processes and systems.
- It will improve the predictability, communication, effectiveness of overall digital preservation activity across all High Courts.

11. Digitization of Legacy Records

11.1. Planning for digitization

Before going for digitization task in a Judiciary, groundwork is required for selection of documents which needs digitization considering importance and future requirements of the documents, type of documents (files, rare books, historical documents, loose sheets, maps etc.) and various sizes of the documents (A0 - A10). An approximate number of total pages being digitized have to be worked out for the estimation of total efforts, work involved and duration of the project.

One of the biggest challenges / tasks of the Digitization project is that on one hand the artefacts are priceless but in poor physical condition, on the other hand the service provider is expected to meet the minimum target of digitizing 30-50 thousand pages in excellent quality at multiple distant locations of District Courts daily. Keeping this target in mind, the service provider must acquire a thorough understanding of the District Court holdings and the manner in which the current systems and procedures work. The various types of archival material/data in the courts comprise of data or judgments/orders and other records collectively these are called “COURT Holdings”, or simply “artifacts.” The courts primarily have following kinds of holdings:

1. Fresh filed cases.
2. Documents received in pending cases.
3. Pending records
4. Disposed records (daily disposal).
5. Disposed records (before digitization initiated).
6. Electronic documents created (digitally signed) by different stakeholder.
7. Mechanism of storage of audio and video record in the respective case records.

Brief descriptions of above holdings:

a) Judgments are typed or handwritten or printed or a combination of these (for instance a typed note sheet with handwritten marking in the margins).

b) A vast majority of the Holdings comprise of Files of Legal and A4 sizes.

c) A single record may consist of one or several pages.

d) The size of Judgment and order will be Legal / FS / A4 or equivalent.

e) Most of the documents are on papers.

f) Most of the documents/ records are in black & white. There might be possibility that some colour document might be there.

g) All electronic document like audio, video or images submitted during the trial period.

The following activities or groundwork is required initially to initiate a digitization project and for awarding the contract to the vendor (scanning agency).

- Weeding out the documents, files/records etc.
- Selection of documents for digitization
- Identification of type of documents (files/books/maps etc.)
- Identification of the size of documents (A4/A3/legal etc.)
- Total number of pages approx.
- Funds available

11.2. Selection of Records for Digitization as per Record Retention Policy

Supreme Court and High Courts have already defined their own record retention policies, which are helpful in differentiating the records, which require preservation. The records retention policies of the High Courts are slightly different from each other. Therefore, the following generic guidelines are provided, which can be adopted by the High Courts for defining the scope of digitization.

- Disposed cases / judicial records requiring permanent preservation
- Administrative records requiring permanent preservation
- Judicial records of historical importance.
- Judicial records requiring preservation after its disposal (completion / end of life cycle)
- Judicial records required for frequent reference by different parties simultaneously / frequently.
- Pending records

It must be noted that the legacy records must be complete in all respects and should have reached the end of life cycle for them to be considered for digital preservation.

The electronic records / born-digital records may be directly considered for digital preservation as per the record retention policy.

11.3. Pre-Scanning Activities

Before starting the scanning work, files/documents should be prepared or ready for scanning. The pre-scanning activities which may include (but not limited to) removal of dust, removal of tags, pins, threads, rubber bands, application of curative techniques to biologically infected or otherwise damaged documents etc., sorting and numbering of pages in the document file in correct order. Special care and attention are required in preparing the documents which are too old and that may not be in good physical condition or are very delicate and cannot be directly scanned. It is recommended to review physical documents to verify completeness, legibility and "scan- ability" to ensure that it will scan smoothly. We may summarize these activities as follow.

- Collection of documents from user record room/courts to be scanned
- Maintain record of received documents for scanning.
- Cleaning/Dusting, if required.
- Unbinding a document, if required.
- Numbering/sorting of each page in correct order.
- Flattening of wrinkles of folded pages.
- Staples/pins /rubber band/file tags/paper clips removing.
- Taping/pasting of torn pages.
- Weeding out of undesired pages.
- One should make photocopies of original documents/pages, if necessary, which have been identified by the Government Department as special documents but extreme fragile/ delicate and may get damaged upon movement through the scanner mechanisms.
- Application of curative techniques for biological infected or damaged pages if required.

Detailed description of the pre-digitization process is as follows.

Document Preparation

The District Court staff deputed in Record Room/Courts shall deliver the physical files on day to day basis to vendor after taking due acknowledgement from the Vendor. The vendor shall do document preparation work by unbinding/un-tagging/ un-dusting the physical files with due care. Document preparation work shall include the work of unbinding, repairing, cleaning, counting the number of pages of the physical file and also rebinding if requires. Proper tapes are to be affixed on torn pages.

Document Segregation

After the document preparation, the work of document segregation by flagging the physical documents of files with indexing parameter will take place as mentioned in the Annexures. There are different types of documents in a case file. In this process it is required to identify and tag document types in a file. List of document types will be provided by the District Court. Further, the record will be scanned as per the technical specification mentioned in Annexures.

Stamping/Segregation

The vendor will differentiate the original copy and photocopy in the case file which will flag in the software so that scanned pages can be identified whether it is photocopy or original document.

The work of pre-scanning activities may be in-house or can be assigned to the scanning agency. The documents would be handed over in lots as agreed mutually between the vendor and the user Government Department. The vendor will provide acknowledgement of number of documents and number of pages in each document received from user department. The scanning team will maintain a record of the collected documents received for scanning in a log register. This log register may contain the following details.

- Description/title of document collected
- File number, if any
- Date of collection
- Total number of pages
- Collected from (court officials)
- Collected by (Vendor representative)
- Date of return
- Return to (court officials)
- Return by (Vendor representative)

11.4. Digitization Strategy and Specifications

The digitization specifications are provided with 3 different digitization strategies as per multiple types of records based on whether they require Black & White or Grayscale or True

Colour scanning. All three digitization strategies can be applied depending on the assessment of records.

Note:-

In this section, the digitization specifications are provided with alternatives in terms of "highly desirable" or "minimum required" quality expectation. The courts have to adopt a comprehensive digitization strategy by selecting and combining the specifications given below. The digitization should at least comply the "minimum desired" specification.

11.5. PPI and DPI

The terminology of Pixels Per Inch (PPI) is used when it comes to onscreen viewing / rendering of scanned documents. The terminology of Dots Per Inch (DPI) is used when it comes to printing the document. We have come across different brands of scanners with their software using PPI and sometimes DPI terminologies while deciding the resolution. Both terms are used to convey the same meaning and to decide the resolution.

JP2K file format for the final output of digitization is strongly recommended considering that it provides lossless compression with much smaller file size if compared with the traditional TIF format.

11.5.1. Black & White Digitization

Record Type	**Purely textual, laser printed or typeset documents having clear visibility of text with high contrast between white paper background and information**
Digitized Master Copy for Preservation	
Minimum desired quality	1-bit bitonal mode - 300 ppi for documents with smallest significant character of 2.0 mm or larger
High quality	1-bit bitonal mode - 600 ppi for documents with smallest significant character of 1.0 mm or larger
Output format	JP2K (lossless compression) most preferred OR Uncompressed TIFF
Access Quality Output for Online Usage	
Compression	While producing the access quality PDF/A document, the digitized master

	copies of individual pages in the record are resampled at a reduced size JPG Compression at minimum 72ppi to 96 ppi resolution (The size of master image is reduced while ensuring the ease of readability of information in the record. The ppi resolution can be increased to ensure clarity of text.)
Searchable PDF/A	A composite searchable PDF for Archival as per ISO 19005 Minimum PDF/A-1a is acceptable as the basic profile for access quality searchable document. PDF/A-2a is highly desirable as this format supports the following specification: • JPEG 2000 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 • the option of embedding PDF/A files to facilitate archiving of sets of documents with a single file

11.5.2. Grayscale Digitization

Record Type	**Documents with poor legibility or diffuse characters (e.g. carbon copies, faxed copies, etc.), handwritten annotations or other markings, low inherent contrast, staining, fading, halftone illustrations, or photographs**
Digitized Master Copy for Preservation	
Minimum desired quality	8-bit grayscale mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger

High quality	8-bit grayscale mode – 400 ppi for documents with smallest significant character of 1.0 mm or larger
Output format	JP2K (lossless compression) most preferred OR Uncompressed TIFF
Access Quality Output for Online Usage	
Compression	While producing the access quality PDF/A document, the digitized master copies of individual pages in the record are resampled at a reduced size JPG Compression at minimum 72 ppi resolution (The size of master image to be reduced while ensuring the ease of readability of information in the record. The ppi resolution can be increased beyond the minimum recommended resolution to ensure clarity of text.)
Searchable PDF/A	A composite searchable PDF for Archival as per ISO 19005 Minimum PDF/A-1a is acceptable as the basic profile for access quality searchable document. PDF/A-2a is highly desirable as supports the following specification: • JPEG 2000 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 • the option of embedding PDF/A files to facilitate archiving of sets of documents with a single file

11.5.3. True Colour Digitization

Record Type	**Documents as described for grayscale scanning and/or where color is important to the interpretation of the information or content, or desire to produce the most accurate representation**
Digitized Master Copy for Preservation	

Minimum desired quality	24-bit RGB mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger
High quality	24-bit RGB mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger
Output format	JP2K (lossless compression) most preferred OR Uncompressed TIFF
Access quality output for online sharing	
Compression	While producing the access quality PDF/A document, the digitized master copies of individual pages in the record are resampled at a reduced size JPG Compression at minimum 72ppi resolution (The size of master image is reduced while ensuring the ease of readability of information in the record. The ppi resolution can be increased to ensure clarity of text.)
Searchable PDF/A	A composite searchable PDF for Archival as per ISO 19005 Minimum PDF/A-1a is acceptable as the basic profile for access quality searchable document. PDF/A-2a is highly desirable as this format supports the following specification: • JPEG 2000 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 • the option of embedding PDF/A files to facilitate archiving of sets of documents with a single file

Note:-

It is mandatory to store digital records in open standards-based formats in the Trustworthy Digital Repository and for clearing the audit and certification as per ISO 16363. Therefore, courts and other agencies like police stations should ensure that digitized records or digitally signed records are produced in compliance with the open standards-based formats as defined in this section.

The closed source and proprietary file formats are not sustainable over long term and can become obsolete. Converting the digitally signed documents into open formats in the later stage may result in loss of evidentiary value of the record. Therefore, existing software systems should be upgraded to produce the output in recommended file formats only.

11.6. Optical Character Recognition (OCR)

Optical Character Recognition refers to recognition on machine-printed text that uses various fonts, such as Arial, New Times Roman or regional language fonts. This text is created with a word processor, typewriter, or printer.

For best recognition results, use a dpi (dots per inch) between 200 and 300 dpi. 200 dpi is a minimum dpi for text that is 10 point or larger. If the text is 9 point or smaller, the dpi would need to be higher. Languages that have small intricate characters use minimum 300 dpi for 10 point text.

Production of searchable PDF/A document requires OCR to be performed on digitized images containing printed text. The documents in regional languages will require regional language OCR supported with properly trained data for good results.

It is quite likely that some records may contain text in English as well as regional languages. In such case, appropriate bilingual OCR should be used. The OCR may be selected on the basis of accuracy of the text output. New start-ups may be involved in OCR customization and optimization.

It is difficult to perform handwriting recognition on hand written documents.

11.7. File Naming as per CNR Guidelines

The file naming of digitized files must comply with Case Number Record (CNR) Numbering System. Incremental serial numbers may be added in the CNR string of characters to define page numbers. An example is illustrated below.

Each folder of the record to be exactly named as per the CNR of the record and it should contain the following folders.

Folder structure for each digitized record		
Record Name as per CNR System	**PBJL01-015294-2016**	
Main Folder Name	PBJL01-015294-2016	
Sub-folder Name 1	Master-PBJL01-015294-2016	Contents of folder PBJL01-015294-2016_001.jp2 PBJL01-015294-2016_002.jp2 PBJL01-015294-2016_003.jp2
Sub-folder Name 2	Access-PBJL01-015294-2016	Contents of folder Access quality PDF/A file and Digitization Verification Information (DVI) in XML format PBJL01-015294-2016.pdf PBJL01-015294-2016-DVI.XML

The district court should copy the digitized records arranged in folders as indicated above in LTO tapes or Blue Ray Disks or DVDs and transfer the digitized records to the High Court for preservation. They may also maintain a copy of this data. The JDPS should suffix an autogenerated number and timestamp after the ingest of SIP.

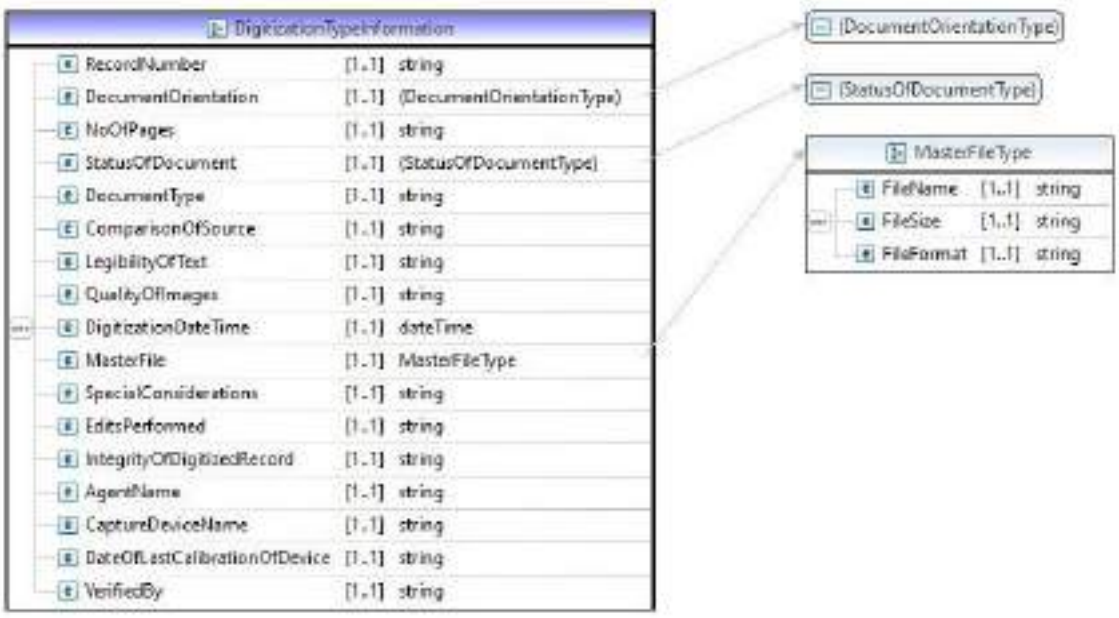
11.8. Verification of Digitized Records

The designated officer(s) of the court should verify each digitized record by comparing it against the original document as per following parameters (as applicable) and store the following information in a database.

The following information forms the basis for issuing the certificate by the designated officer for certifying the admissibility of digitized record in the court as per the Section 65B of Indian Evidence Act. Refer the section on legal framework for specific clauses applicable in this context. The rank and designation of the officers in charge of the particular digitization processes should be identified for uniform implementation.

Verification of Digitized Record	
1.	Record number
2.	Document orientation (Portrait or landscape)
3.	Number of pages
4.	Original or photocopy
5.	Comparison of source and digitized record
6.	Legibility of text in scanned document
7.	Quality of images
8.	Date and time of digitization
9.	Master file name, size, format
10.	Special considerations
11.	Edits performed
12.	Integrity of the final digitized record
13.	The name of the agent associated with the digitization process (e.g. name of the outsourced bureau or name of the in-house operator)
14.	Capture device name (HW / SW)
15.	Date of last calibration of device
16.	Verified by

The XML schema for DVI is given below:



The Digitization Verification Information (DVI) gets linked with the Digitization Information (Block No. 9) in the Preservation Metadata as defined in the chapter on metadata requirements.

11.9. Quality Control

Quality control procedures should be defined, documented and implemented. Quality control is necessary to ensure the digital copy of the non-digital source record is a true and accurate copy. This is critical such that the records possess integrity and are authentic.

Quality control procedures should not only be applied at the point where the digital output is produced, but also be documented and built into the ongoing operation of the digitizing process. Quality control procedures should, at minimum, address the following issues:

– any acceptable variations from normal procedures;

– scanner operation quality control;

– verification to ensure that the digital output matches the quantity of non-digital source record input;

– extent and frequency of sampling of digitized images;

– criteria for checking image quality;

– frequency and criteria for checks on metadata;

– processes for re-digitizing;

– operator training.

Quality checking should be completed before the digitized images are accepted into a business process, or as a master copy in the case of digitization projects. Quality checking should be complete before the destruction of the non-digital source records is considered.

11.9.1. Reviewing Quality Control Checking

The results of quality control processes and quality checks should be documented. A review of quality procedures for digitizing should be undertaken regularly to ensure that the procedures continue to meet the business purpose. Appropriate training should be provided to all staff who create, manage or work with digitized records. Documentation on the level

and the frequency of training provided to those staff involved with digitization should be created and maintained.

11.9.2. Digitization Cell

A dedicated digitization cell should be established at each High Court comprising of Project Manager, digitization and metadata experts and technical persons as a permanent cadre in order to monitor the day-to-day progress of the digitization project as per project management guidelines. The digital cell should be headed by an officer with technical background such as Registrar (IT), Registrar (Computer) or an officer (Technical) nominated by Hon'ble Chief Justice of the High Court.

Additionally, there is a need to set up Digitization Offices at the District Courts for the entire duration of this exercise. These offices will manage contracts with accepted vendors for specialized services while being accountable to the judiciary for accomplishing phase wise digitization milestones.

11.9.3. Synchronizing the legacy digitization

Many High Courts have digitized court records following their own format and parameters of metadata. Now in view of the SOP for digitization, the legacy digitized data should also be synchronized with standardized parameters of digitization provided in the SOP.

11.9.4. Digitization Progress Monitoring

There should be an online common mechanism to monitor the progress of digitization in all the courts. Proper logs with date, timestamp and accession register of digitized records should be maintained to report and monitor the progress of digitization happening in courts across India. The reporting of digitized records and transferred records should be compared for verification. Release of payments to vendors involved in digitization should be linked with online reporting on the progress of digitization.

12. Metadata Requirements

12.1. Adoption of Paris Cataloguing Principles for Judicial Records

The digital preservation of judicial records requires adoption of the Paris Principles for Cataloging which are helpful in selecting the common cataloging parameters (International Conference of Cataloging Principles 1961) for access purpose. The Paris Principles primarily focus on how to find a single resource (e-record) and how to find sets of resources (large volume of e-records) associated with a given person, family, or organization or all resources on a given subject. It also covers the finding of resources defined by other criteria such as, language, date, type, place etc. The cataloguing parameters for e-records provide adequate access points for classification and retrieving the bibliographic data. The cataloguing parameters are mandatory to be filled for the purpose of archival and access.

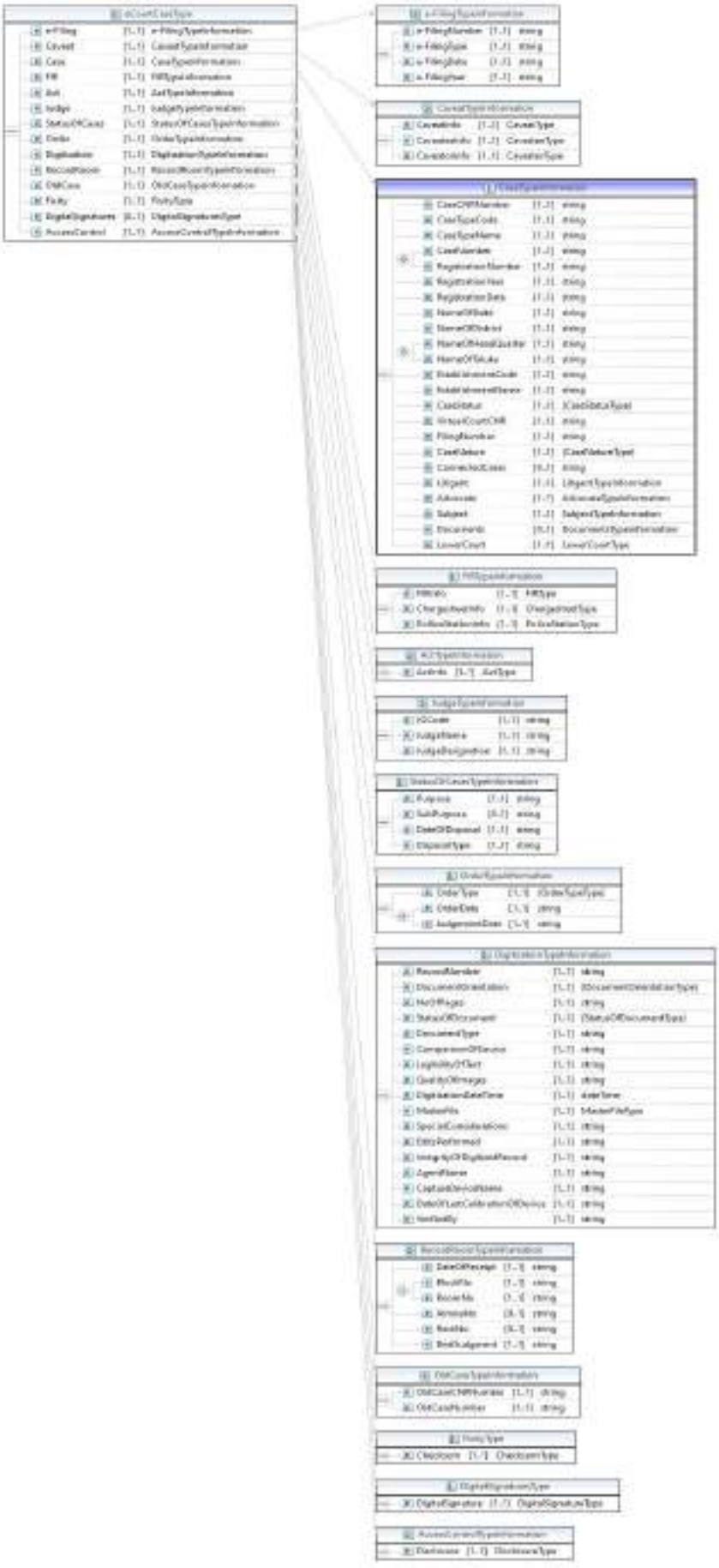
Therefore, while considering the Paris Principles for Cataloguing, the common metadata elements should be identified based on the inputs received from various High Courts. The metadata elements are clustered into 12 information blocks as listed below.

12.2. Preservation Metadata Information Blocks

Sr. No.	Preservation Metadata Information Blocks	Remarks
1.	e-filing Information	
2.	Caveat Information	
3.	Case Information 3.1 Litigant Information 3.2 Advocate Information 3.3 Subject Information 3.4 Document Information	3.1 and 3.2 are repeatable blocks
4.	FIR Information	
5.	Act Information	It is a repeatable block.
6.	Judge Information	It is a repeatable block.
7.	Case Status Information	
8.	Order Information	It is a repeatable block.
9.	Digitization Information	
10.	Record Room Information	
11.	Old Case Information	

12.	Digital Signature Information	It is a repeatable block. It is applicable to digitally signed documents
13.	Integrity Information	It is a repeatable block. The hash value of the document is generated and maintained for verification of integrity. It is applicable to all documents being preserved.
14.	Access Control Information	This information block defines the access control for each digitized record in terms of Public or Private (Confidential records).

The XSD combining all the information blocks is provided on next page.



12.3. Preservation Metadata Elements

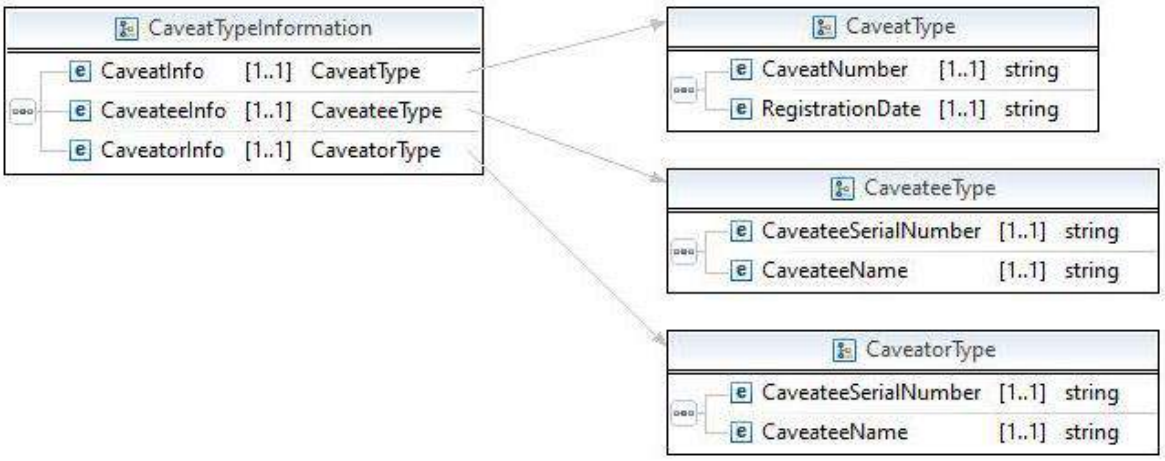
Each preservation information block is expanded in terms of common metadata elements and the design of XSD is shown for clarity.

1. e-Filing Information

e-Filing Information	
e-Filing No.	All petitions, applications, appeals and all pleadings/documents in fresh, pending and disposed of cases of all types are filed / uploaded electronically in PDF/A format with digital signature. The metadata elements defined here are meant to capture the information necessary for identification of e-filing.
e-Filing Type	
e-Filing Date	
e-Filing Year	

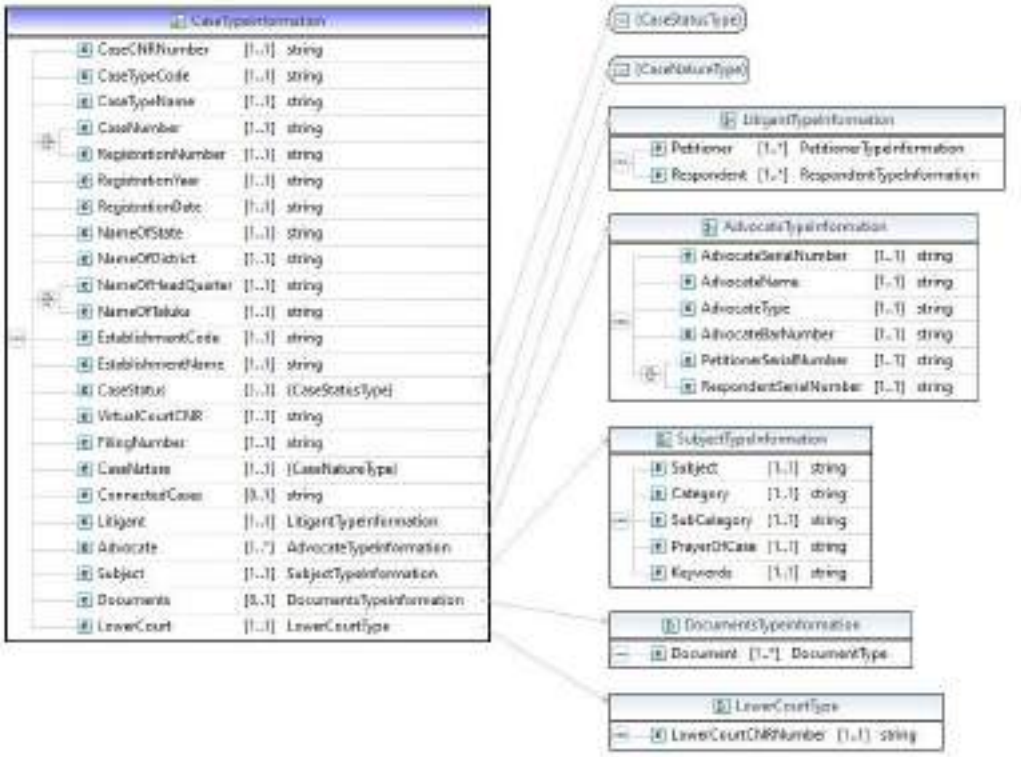
2. Caveat Information

Caveat Information	
Caveat No.	This information block captures the details necessary for identification of a caveat filed. This information is useful to know whether the caveat is filed before or after filing the case.
Caveatee Name	
Caveator Name	
Registration Date	
Caveator Serial No.	
Caveatee Serial No.	



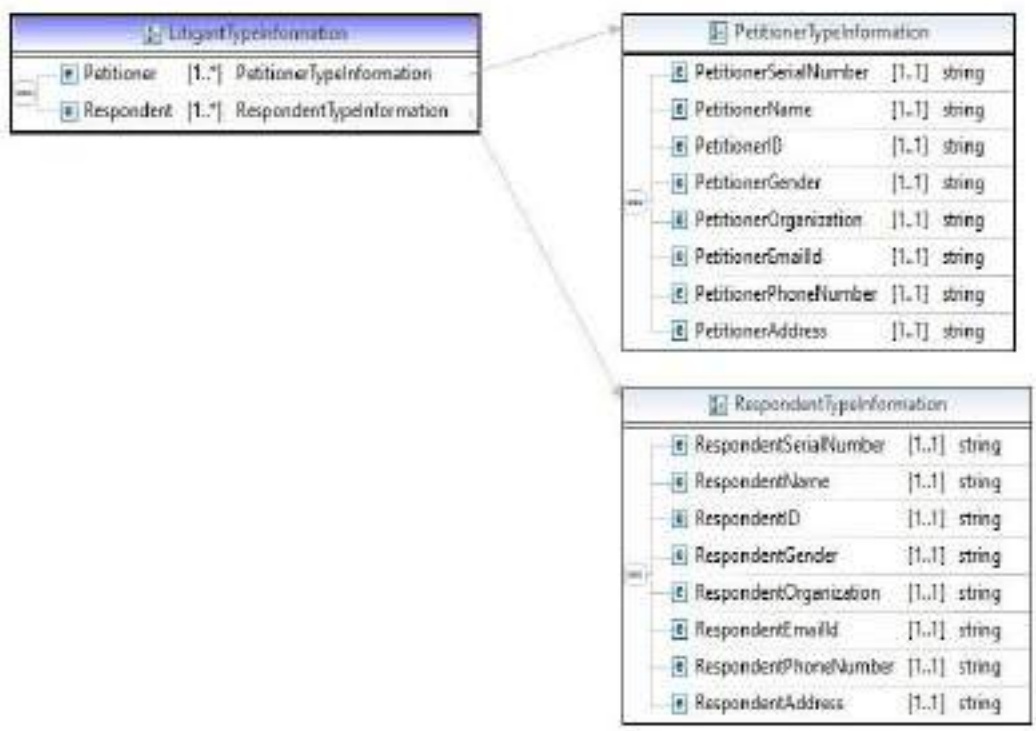
3. Case Information

Case Information	
CNR No.	Case information block captures all the identifiers and particulars of the case. The case is identified by the CNR number, which is a unique number in entire judicial system in the country. Information, such as master of case type, case nature, pending and disposal remark, state, district, taluka and established code, is to be preserved for identification of the case related Information.
Case Type Code	
Case Type Name	
Case/Registration No.	
Registration Year	
Registration Date	
Name of the State	
Name of the District	
Name of the Head Quarter / Taluka	
Establishment Code	
Establishment Name	
Case Status (Pending/Disposed)	
Virtual Court CNR	
Filing No.	
Case Nature (Criminal or Civil)	
Connected Cases	
Lower Court CNR	



3.1 Litigant Information

Litigant Information (R)	
Petitioner	This is a repeatable information block for all the petitioners and respondents associated with the case. It is a subset of Case Information. The serial number, identity and gender of the Petitioner/ Respondent is equally important when the case moves from a subordinate court to the higher court.
Petitioner ID	
Petitioner Gender	
Petitioner Organization	
Petitioner Serial No.	
Petitioner Email ID	
Petitioner Contact	
Petitioner Address	
Respondent	
Respondent ID	
Respondent Gender	
Respondent Organization	
Respondent Serial No.	
Respondent Email ID	
Respondent Contact	
Respondent Address	



3.2 Advocate Information

Advocate Information (R)	
Advocate Type (Caveator)	This information block captures the details of advocates with serial numbers to link them with the serial numbers of petitioners and respondents. It is a subset of Case Information.
Advocate Name	
Advocate Serial No.	
Advocate Bar No.	
Petitioner /Respondent Serial No.	

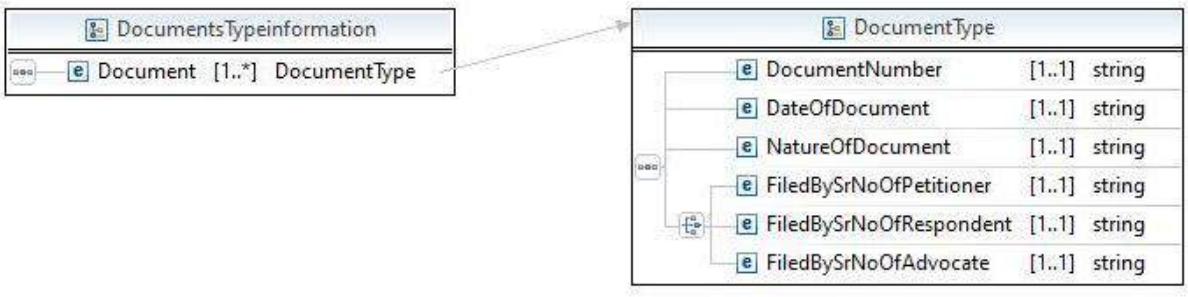
3.3 Subject Information

Subject Information	
Subject	This is most important data for preservation which may be utilized for automatic categorization of cases with the help of text analytics.
Category	
Sub Category	
Prayer of the Case	
Keywords	

3.4 Document Information

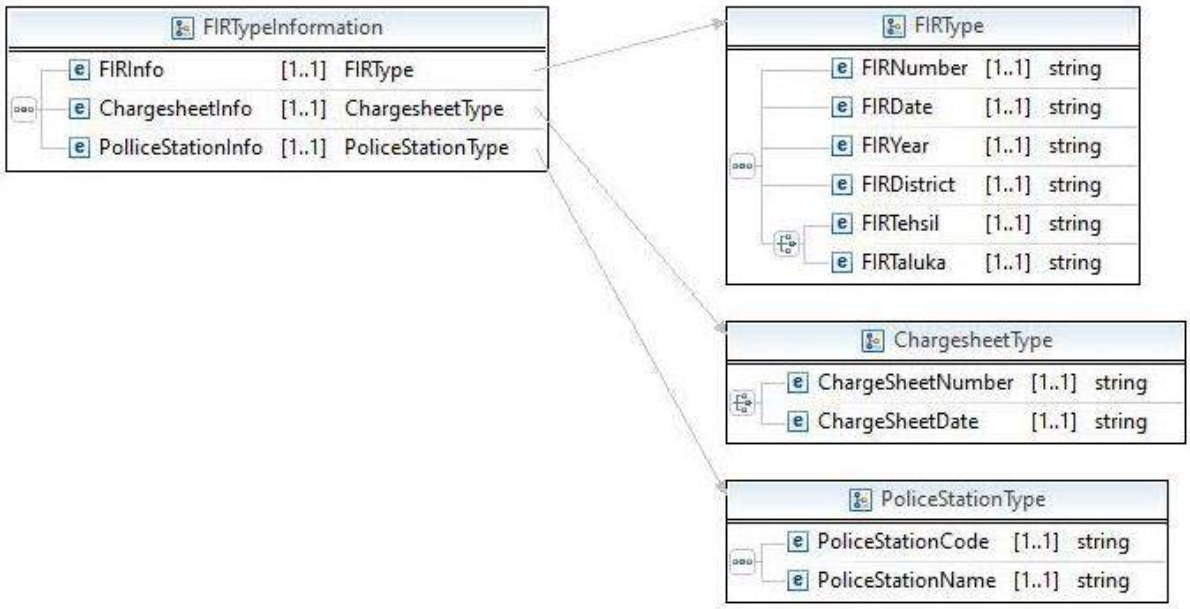
Document Information

Document No.	This information block can be used for identification of type of the document submitted by which user and what action has been taken on that document.
Date of Document	
Filed By Sr. No. of Petitioner/ Respondent / Advocate	
Nature Of Document	



4. FIR Information

FIR Information	
FIR District	This information block helps in identifying the location and time of the crime and the police station where the FIR was registered.
FIR No.	
FIR Year	
FIR Date	
FIR Tehsil/Taluka	
Chargesheet No./Date	
Police Station Code	
Police Station Name	



5. Act Information

Act Information (R)	
Act	This is an important information block which helps in identifying the nature of the case.
Section	

6. Judge Information

Judge Information (R)	
JO Code	This information block helps in getting the name, designation and code of the presiding officer/Judge who disposed of the case.
Judge Designation	
Judge Name	

7. Case Status Information

Case Status Information	
Purpose	This information block defines the nature of disposal of the case and its final purpose listed before the court.
Sub Purpose	

Date of disposal	
Disposal Type	

8. Order Information

Order Information (R)	
Order Type (Interim Order/ Zimini Order/ Final Judgment)	This information block indicates nature of order or final judgment along with date.
Order/Judgement Date	

9. Digitization Information

Digitization Information	
Status of Document (Original, Photocopy, Carbon Copy, True copy, Certified copy)	This information block picks up details from Digitization Verification Information (DVI) along with the incoming digitized records.
Date of Digitization	
Document Type	
No. of Pages	

10. Record Room Information

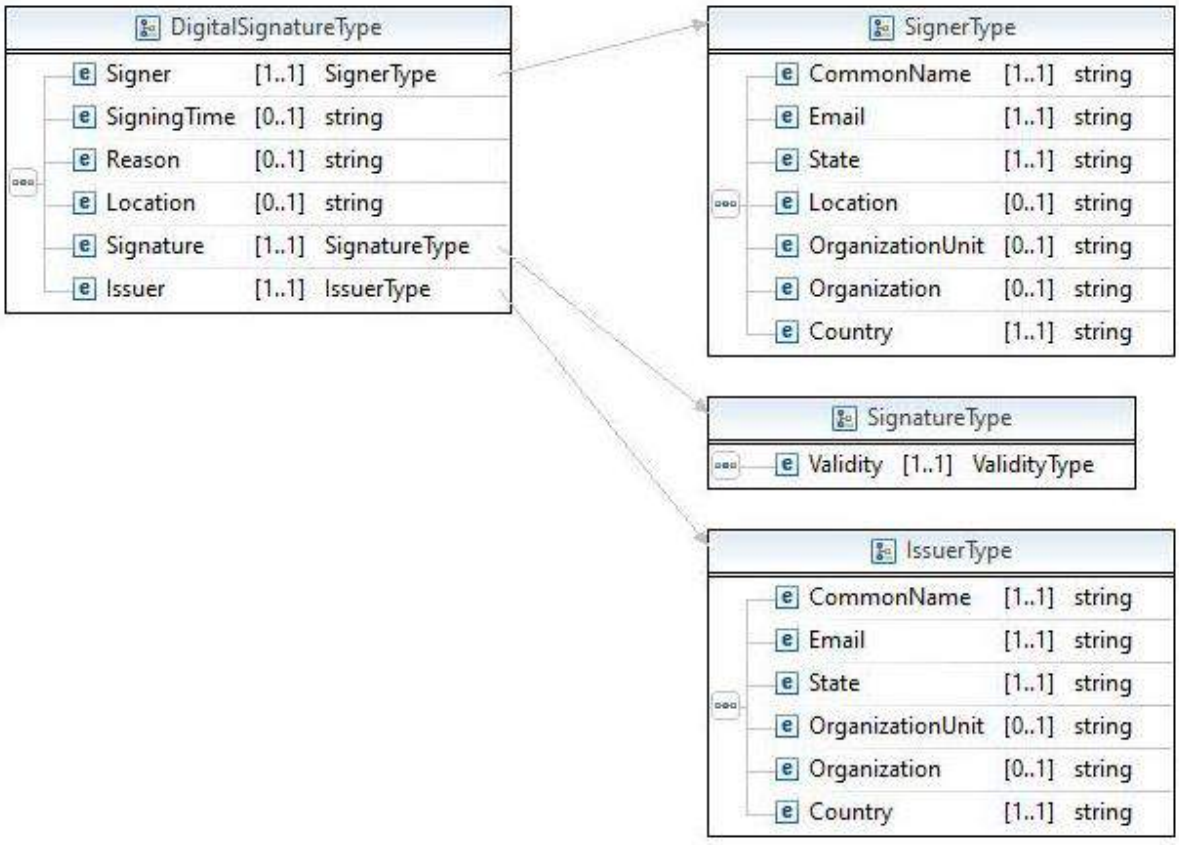
Record Room Information	
Date of Receipt	It helps in finding out the location of original hard copy of the file.
Block/Room No.	
Almira No.	
Rack No.	
Brief of Judgement	

11. Old Case Information

Old Case Information	
Old Case No.	This information block provides previous case number which was assigned manually before computerization started in the Court.
Old CNR No.	

12. Digital Signature Information

Digital Signature Information (R)	
Signer	This information block is repeatable and it captures the signature details of all the digitally signed documents.
Signing Time	
Reason	
Location	
Signature	
Issuer	



13. Integrity Information

Integrity Information (R)	
Hash 1	This information block is repeatable and it captures the hash value of each document using 2 different algorithms.
Hash 2	

14. Access Control Information

Access Control Information	
Public	This information block is helpful in defining the access control for the digital record in terms of whether the record can be made online for public access or should be kept confidential.
Private	

13. SIP Preparation

It is important to transfer the digitized data in the form of Submission Information Package (SIP) as defined in the OAIS Reference Model.

13.1. Submission Information Package (SIP) Preparation

The SIP is an information package that is delivered to the repository and digital storage system for ingest. The valid SIP comprises of digital record in the specified format and preservation metadata to provide adequate understanding of the object being preserved. As shown in Fig. 8., there has to be an interoperability between Case Information System (CIS) and the Judicial Repository Management System (JRMS), which is compliant with OAIS Reference Model.

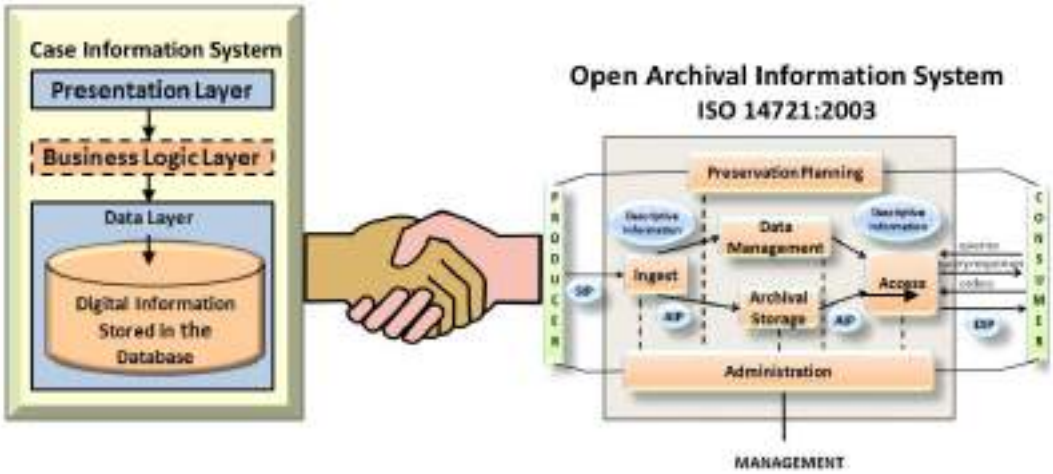


Fig. 9. Interoperability between CIS and OAIS

(Image Courtesy EGOV-PID Metadata Dictionary & Schema, eGov.DP.01-02 Version: 1.0 December, 2013, Published by MeitY)

13.2. Specification for Submission Information Package (SIP)

The Submission Information Package (SIP) is to contain following files:

CNR_Folder

- CNR_Case_Record.PDF
- CNR_FIR.PDF
- CNR_Order.PDF
- DVR.XML
- CNR.XML

13.3. Procedure for Integration of Digitized Legacy Records

- The courts have to organize the digitized files of legacy records case wise in separate folders by following standard naming conventions.
- The digitized records arranged as above are sent to High Court JDR on monthly basis by storing it on offline storage media.
- High Court JDR makes the digitized records available through Judicial Digital Preservation System (JDPS), which is accessible online.
- The court officials / data entry operators can log into JDPS and access their digitized records online.
- The data entry operators generate CNR number for the digitized record, enter the preservation metadata and link the digitized records using JDPS.
- The court officials can check the CNR number, metadata, linked digitized records and approve the ingest for preservation. They can also correct and revise the information or reject in case of gross inaccuracies.

13.4. Procedure for Integration of Records Received from CIS

- The courts prepare the metadata for digitized records using Case Information System (CIS).
- The digitized records are linked with the metadata.
- Backup of CIS databased along with linked digitized records is taken and sent to High Court JDR.
- Case wise SIPs are extracted from the database along with digitized records and stored in separate folders.
- Database records are extracted in the form of XML.
- Preservation metadata scheme is mapped with CIS databased record and relevant metadata is extracted.
- Final SIP is verified, validated and then ingested into Judicial Trustworthy Digital Repository (JDR).

13.5. Interoperability Requirements between JDPS and CIS

- JDPS to implement same logic and naming conventions for generating CNR as existing in CIS.
- It should be possible to extract metadata from JDPS for integration in CIS.
- It should be possible to extract metadata and digitized records from CIS for integration in JDPS.

14. Transfer of Digitized Records to JDR

14.1. Copy of Data on Storage Media

- The digitized records have to be organized with proper file naming and folder structure as per the guidelines given in the previous chapter on preparation of Submission Information Package (SIP).
- "Master Data" and "Access Quality Data" both should be copied in separate folders.
- SIP to include Access Quality Data.
- The prepared data as per the guidelines may be checked / verified and then encrypted.
- The encrypted data then can be copied on to a suitable storage media.
- The entire batch of the data should be copied on the storage media.
- Label the storage media with the proper information as per the following parameters.

14.2. Label for Storage Media

The label on the storage media should cover following details:

PROFORMA for Storage Media Label

Date:
Batch No.:
Media No.:
Name of Court:
Name of Officer:
Signature:
Mobile No:

14.3. Transferring the Digitized Records to JDR

- At the time of transferring of digitized data in LTO tapes or Blue Ray Disks or DVDs or Hard Discs, the responsible officer of the court should submit a CERTIFICATE along with the list of digitized records contained in the media to ascertain its evidentiary value.
- The proforma for CERTIFICATE is provided in chapter 16 of this SOP.

14.4. Frequency of Data Transfer to JDR

- The digitized records should be transferred to JDR at the High Court on monthly basis.

14.5. Online Data Tracking

There should be proper tracking and monitoring of the progress of transferred data to High Courts. The district courts would need a proper receipt of the transferred data. Also, the progress on the data validation, ingesting and publishing should be trackable by the concerned stakeholders. Each digital deposit of a record in the JDR must be trackable all times. The district court should also have a computer-generated list of digitized records that are transferred to High Court for verification to avoid any loss of data.

14.6. Backup of Digitized Records

Master data is the high resolution JP2K images along with access quality PDF/A documents of digitized records as described in the section on digitization specifications. This also includes digitally signed documents / e-records / e-files that require to be preserved as per the retention policies. The master data and access quality data should be copied on transferrable storage media after it is finalized, verified by designated officer and completed in all respects. Following are the guidelines for selection of storage media for backup and transfer of data to High Court.

14.7. Basic Guidelines for Selection of Storage Media

- The digitized master data / source data may be stored on Hard Disc or Network Attached Storage (NAS) at the time of digitization.
- In order to transfer the master data to High Court, the designated officers must consider the following aspects related to storage devices for selecting the storage media.
  - 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)
  - read / write speed
  - Write Once Read Many (WORM) storage
  - mature and established technology
  - cost of storage media and reading / writing device
- Ensure that you are able to copy the data on minimum number of offline storage media for transferring to High Court.
- The storage media containing the digitized records /e-records should be numbered,

classified, labeled and indexed in a register before transferring to High Court.

- Cost effective but reliable storage media may be chosen for data transfer.

15. Data Protection Strategy

Multiple copies stipulated by the 3-2-1 backup rule protect you from losing your primary data. Storing multiple copies ensures that there is no single point of failure and that your data is safe. 3-2-1 backup rule requires you to maintain at least 3 backups of your data (Primary copy + 2 copies on storage media + 1 copy at DR location). This strategy is strongly recommended by the information security professionals as well as US-CERT (United States Computer Emergency Readiness Team).

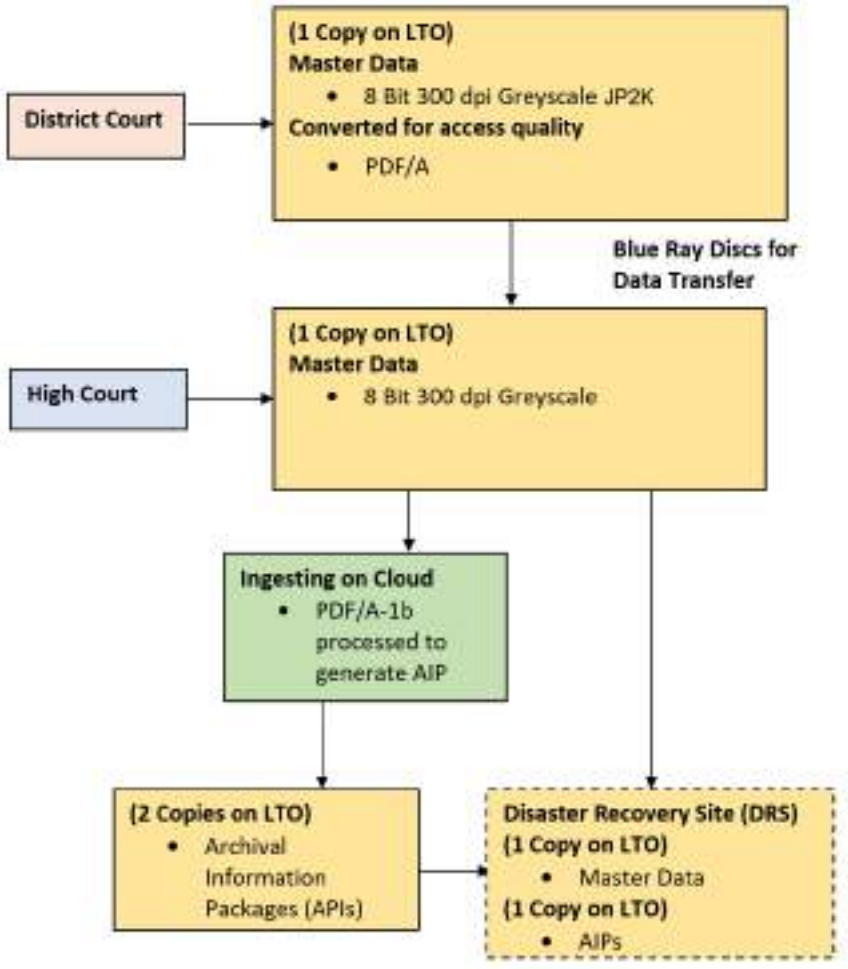


Fig. 10. Data Protection Strategy

15.1. Backup at District Courts

- District Courts are required to initially store the digitized data on re-usable NAS storage. They should keep a copy of master data and convert access quality data in LTO. The District Courts to maintain LTOs in e-record room.
- District Courts are required to transfer a copy of the master data and converted access quality data in separate folders on Blue Ray Discs or Hard Disks or Flash Drive (depending on the size of data) to the JDR at High Court on monthly basis.

15.2. Copies of Data at High Court and DR Site

- High Court JDR to make 2 copies of master data separately on 2 LTOs (One for the High Court and another for DR).
- The PDF/A documents are ingested using JDPS for producing Archival Information Packages (AIPs) on the cloud i.e. primary storage.
- Successfully processed AIPs to remain preserved on cloud.
- 2 copies of AIPs are maintained on 2 LTOs at the High Court JDR. 3rd copy of AIPs on LTO is maintained at the DR location.
- This ensures that High Courts will always have 3 copies of master data and 2 copies of AIPs in addition to the AIPs stored on cloud.
- DR Location to have a facility with a fire safe vaults, secure and temperature-controlled environment as specified by the manufacturer for keeping the LTOs.

15.3. Refreshing of Storage Media

- The digital records stored in the storage media need to be retained for longer duration than the lifetime of the media. Therefore, the digital data must be migrated into another media before its expiry.
- All storage media used for backup to be checked to ensure that they are readable and in working condition. The LTOs / NAS / Hard Discs to be refreshed by transferring the data on new storage media after certain period. Refreshing of the media means transferring the data from old media to latest storage media.

Data Protection Strategy			
District Court	Master data Access copy	LTO Copy 1	A copy of master / source data is kept at district court.
High Court JDR	Master data	LTO Copy 1	A copy of master data at JDR
	AIP	LTO Copy 1	A copy of AIPs on LTO for the High Court.
Disaster Recovery Site	Master data	LTO Copy 1	It is to be used only if the backup available at the High Court JDR is lost in disaster.
	AIP	LTO Copy 1	It is to be used only if the backup available at the High Court JDR is lost in disaster.

15.4. Cataloguing of storage media

All the storage media meant for long term digital backup must be properly labeled and catalogued at the High Court using a software with a facility to search and retrieve the media according to its contents. The software should facilitate batch numbering and generate the

label information. The software should also be able manage the updates the storage media at the time of refreshing.

16. Step-by Step Implementation

16.1. Overview of Digitization Process

Step-by Step Implementation	
High Court	**District Courts**
1 High Courts to establish separate Judicial Digital Repositories (JDR).	2 Courts to digitize the records.
	3 The district courts to transfer digitized records on monthly basis to the Judicial Trustworthy Digital Repositories (JDR) for preservation purpose. They can encrypt the data and transfer it through network (depending on the size of data and bandwidth) or through offline storage media.
4 JDR at High Court receives the data transferred by District Courts. The data is decrypted, verified and made available through the Judicial Digital Preservation System (JDPS) for online metadata creation.	
	5 The operators from district courts can log into the system to check the digital records through online portal and enter and submit the metadata.
	6 The officials from district courts to check the digital record along with its metadata for quality and then approve for ingest in the preservation system.

7 JDPS executes ingest process which culminates into Archival Information Package (AIP) for preservation purpose.	
8 High Court to have access to all district court records.	9 The district court officials and designated users are then able to search, retrieve and download the digital records through online portal.
10 JDR at High Court to preserve the digital records of district courts.	11 District courts can have exclusive access to their own digital records.
	12 District Courts may keep backup of all digitized data.
14 High Courts to maintain centralized digital repository infrastructure (Cloud) and Tape Library for backup.	13 District Courts continue transferring digitized records and born digital records to JDR for preservation as a regular practice.
15 High Courts to maintain Disaster Recovery Site.	
16 High Courts to develop digital preservation capabilities and competencies to manage the digital repository.	
17 High Courts to go through ISO 16363 Audit and achieve the certification of trustworthiness for the Judicial Digital Repository.	
18 High Courts to go through annual surveillance audits to retain ISO 16363 certification of trustworthiness for the Judicial Digital Repository.	

16.2. Training by the State Judicial Academy (SJA)

Training should be an essential component of the onboarding process of all personnel involved in this exercise. Training should also be given on an ongoing basis to disseminate learnings and resolve issues that are bound to arise during the digital preservation process. The State Judicial Academy (SJA) should organize programs for training the trainers and the training for the staff involved in digitization and digital preservation process as defined in the SOP.

17. Certificate for Transferring Digitized Records

As per the conditions referred to in sub-section (1) (2) and (3) in respect of a computer output in Section 65B of the Indian Evidence Act, for admissibility of electronic records in court, the responsible official in-charge of digitization should submit a certificate to ascertain evidentiary value of digitized records along with the storage media for transferring the digitized data to Judicial Digital Repository (JDR) at the High Court. It is possible to generate the proposed certificate using a software with digital signature.

Model CERTIFICATE Proforma

The following computer, digitization equipment and storage have been under lawful control of the office of this court from date …… to ……

Specification of Computer:

- Operating System
- Memory:
- Storage:
- Processor:
- MAC Address:

Specification for Scanner / Digitization Device

- Scanner Make:
- Scanner Type:
- Scanning Technology:
- Maximum Document Size:
- Maximum Colour Depth
- Maximum Resolution Supported:
- Device ID:
- Serial Number:

Specification of External Storage Device (e.g. NAS, SAN, Tape Drive, any other)

- Device Make:
- Type of Storage:
- Storage Capacity:
- Device ID:
- Serial No.:

The above Computer, Scanner and External Storage Device are being used regularly for the digitization activity during abovementioned period.

(Tick as applicable)

A) During the said period, the computer, scanner and storage device were operating properly.

OR

B) During the said period, the computer, scanner and storage device were operating properly except during date……..to ……. the Computer, Scanner, Storage (tick as applicable) was not operating properly or was out of operation during that part of the period. The equipment was repaired to ensure that this does not affect the electronic record or the accuracy of its contents.

A list of digitized records contained in the Storage Media Nos.

(Enclose a separate sheet in case of several storage media.)

I have checked and verified the original records and the corresponding digitized copies stored in the Tapes / Blue Ray Disks / DVDs/any other …… (Tick as applicable) to ensure that the information contained therein is captured accurately as in the source.

Enclosed is a list of digitized records along with Integrity Information and Digitization Verification Information (DVI) duly stored in the media.

Signature Seal:
Name of Officer in Charge – Digitization Project
Date:

Name of Court:

Address:

18. Certificate by JDR

As per the conditions referred to in sub-section (1) (2) and (3) in respect of a computer output in Section 65B, for admissibility of electronic records in court, a certificate should be produced by Repository Manager of the Judicial Digital Repository established at High Court, whenever any digital record is furnished as required by the High Court.

Model CERTIFICATE Proforma

The list of digital records enclosed herewith are preserved in the cloud-based Judicial Trustworthy Digital Repository under the lawful control of the High Court from date …… To ……

List of records

(Tick as applicable)

A) During the said period, the Judicial Digital Repository was operating properly.

OR

B) During the said period, the Judicial Digital Repository was operating properly except for the period from date…….. to ……. was not operating properly or was out of operation during that part of the period. The repository infrastructure has been repaired / maintained to ensure that this does not affect the electronic record or the accuracy of its contents.

I have checked and verified the digitized records retrieved from the Judicial Trustworthy Digital Repository to ensure its integrity. The computer-generated report is enclosed herewith.

Signature:

Date: Seal:

Name of Repository Manager:

Name of High Court:

Address:

19. Budget

Initially, we circulated the Survey Form I as per Annexure I among the High Courts to assess their digitization requirements. This form was helpful in getting a lot of valuable insights but provided mixed information on digital preservation requirements by the judiciary. Therefore, we designed another set of survey forms, in order to get the specific breakup of information on legacy records, new institution cases and pendency in order to project the growth of records for next 5 years.

The projections of digitization requirements received from various High Courts and district courts are used as the basis for budgeting. The overall estimation for legacy records, new institution of cases and pendency projected for next 5 years (after weeding out of records of short-term relevance) is given below.

19.1. Detailed Scope of digitization

Scope of Digitization						
Sr.No	High Court Location	Disposed (Legacy Records)	Pendency	Institution Projection for Next 5 years	Total Pages	Total Pages after Weeding out
1	Arunachal pradesh	67,00,000	21,70,000	1,94,67,675	2,83,37,675	2,56,57,675
2	Gauhati	7,98,00,000	4,85,00,000	40,13,07,333	52,96,07,333	49,76,87,333
3	Manipur	89,00,000	14,00,000	1,51,79,474	2,54,79,474	2,19,19,474
4	Meghalaya	67,00,000	68,00,000	3,41,53,816	4,76,53,816	4,49,73,816
5	Mizoram	49,80,000	5,60,000	1,41,35,885	1,96,75,885	1,76,83,885
6	Nagaland	1,72,00,000	28,00,000	2,27,69,210	4,27,69,210	3,58,89,210
7	Sikkim	38,00,000	16,20,000	1,00,56,401	1,54,76,401	1,39,56,401
8	Tripura	1,08,00,000	6,00,000	66,41,020	1,80,41,020	1,37,21,020
9	Orissa	37,34,00,000	12,41,00,000	30,16,92,038	79,91,92,038	64,98,32,038
10	Delhi	26,75,00,000	7,74,00,000	85,28,96,673	1,19,77,96,673	1,09,07,96,673
11	Punjab	94,34,00,000	16,31,00,000	14,42,04,999	1,25,07,04,999	87,33,44,999
12	Karnataka	1,17,19,00,000	31,50,00,000	75,73,60,861	2,24,42,60,861	1,77,55,00,861
13	Madhyapradesh	95,75,00,000	7,61,00,000	60,81,27,661	1,64,17,27,661	1,25,87,27,661
14	Uttarakhand	10,32,00,000	3,36,00,000	21,82,04,933	35,50,04,933	31,37,24,933
15	Uttarpradesh	1,60,12,00,000	44,68,00,000	1,36,61,52,624	3,41,41,52,624	2,77,36,72,624
16	Kerala	15,26,00,000	11,22,00,000	70,86,91,674	97,34,91,674	91,24,51,674
17	Rajasthan	1,44,54,00,000	20,04,00,000	1,56,25,37,064	3,20,83,37,064	2,63,01,77,064
18	Jharkhand	40,10,00,000	14,20,00,000	79,12,30,061	1,33,42,30,061	1,17,38,30,061
19	Chattisgarh	33,83,70,000	2,56,00,000	20,11,28,025	56,50,98,025	42,97,50,025
20	Madras	87,89,00,000	30,58,00,000	73,71,53,187	1,92,18,53,187	1,57,02,93,187
21	Andhra-Pradesh	1,02,90,00,000	7,86,00,000	1,06,06,65,718	2,16,82,65,718	1,75,66,65,718
22	Bihar	1,97,25,00,000	35,15,00,000	73,43,07,035	3,05,83,07,035	2,26,93,07,035
23	Calcutta	2,28,01,00,000	20,71,00,000	83,20,24,897	3,31,92,24,897	2,40,71,84,897
24	Telangana	71,95,00,000	5,59,00,000	70,58,45,522	1,48,12,45,522	1,19,34,45,522
25	Bombay	4,95,12,00,000	42,30,00,000	1,29,49,98,842	6,66,91,98,842	4,88,87,18,842
26	Himachalpradesh	20,13,00,000	10,14,00,000	28,08,20,262	58,35,20,262	50,30,00,262
27	Jammu & Kashmir	6,39,00,000	7,62,00,000	32,54,09,965	46,55,09,965	43,99,49,965
28	Gujarat	72,49,00,000	12,39,00,000	1,14,69,98,974	1,99,57,98,974	1,70,58,38,974
	Total Pages	20,71,56,50,000	3,50,41,50,000	15,15,41,61,828	39,37,39,61,828	31,08,77,01,828
	Total Pages after Weeding	12,42,93,90,000	3,50,41,50,000	15,15,41,61,828	31,08,77,01,828	

Note:-

The cost calculations, storage requirements, backup requirements are estimated based on the following assumptions in 19.2. It is important to note that JP2K format with lossless compression is considered for digitization of master copies and estimation of storage requirement. The storage requirements based on TIFF as the base format is likely to be much higher and very costly.

19.2. Approximate Rates and Assumptions

The rates for digitization, percentage of weeding out, data sizes etc., are estimated based on certain assumptions for preparing the budget. Therefore, the judiciary need not take them as absolute rates but negotiate appropriate and optimal rates with concerned vendors without compromising the quality parameters.

Assumptions		
Item Head	Rates	Unit
Weeding Out Rate [w]	40	%
Approx. Average Pages per record [p]	100	Pages
Digitization Per Page Rate [s]	0.60	Rupees
1 Bit Bitonal 400 DPI JP2K	35	% of legacy Record Pages
8 Bit GreyScale 300 DPI JP2K	60	% of legacy Record Pages
24 Bit Colour 300 DPI JP2K	5	% of legacy Record Pages
Pendency : 1 Bit Bitonal 300 DPI JP2K		
New Pendency : 1 Bit Bitonal 300 DPI JP2K or Born-Digital		
New Pendency will not go to DR Site		

19.3. Digital Preservation Budget for Judiciary

The budget estimate for all High Courts and District Courts combined is distributed over 5 years as given on next page. The estimate is evolved on the basis of survey inputs received from the High Courts.

Budget for High Courts + District Courts							
	Budgetry Heads	Year-1	Year-2	Year-3	Year-4	Year-5	Total (Crore Rs.)
Digitization	Digitization	375.05	375.05	375.05	375.05	375.05	1865.28
	LTO Backup + DR	7.88	7.88	7.88	7.88	7.88	39.39
	Blue Ray Disks for data transfer	7.47	7.47	7.47	7.47	7.47	37.34
Judicial Digital Repository	JDR Cloud IIW & Upgradation	171.00	16.50	2.10	10.50	47.00	241.10
	Contingency + Maintenance for Cloud	64.00	64.00	64.00	64.00	64.00	320.80
	Data Management	12.82	14.11	15.52	17.07	18.78	78.29
Digital Preservation Tools and Software Solutions to enable compliance as per ISO 16363	Testbed Cloud	10.00	1.75	2.75	2.75	3.50	21.75
	• ISO 14721 - Judicial Digital Preservation System (JDPS) • Portal with Distributed Search & Retrieval • Multilingual support • Data Tracking System • Digitization Quality Verification Tool • Tool for integrity assurance with a digitally signed certificate • Process Optimization and Automation • LTO Cataloguing Software • Interoperability measures as applicable • Customization • Training the Master Trainers • Documentation • Software Deployment • Software Upgradation • Audit support	12.57	13.18	7.08	3.89	4.25	39.76
	Software Support & Maintenances	2.50	3.00	8.40	9.99	10.96	34.86
	Grand Total (Crore Rs.)	661.09	494.03	488.24	496.60	536.89	2,677.76

Total Pages	31,08,77,01,628
Total Cost	26,77,75,62,064
Per Page Cost :	0.86

The separate sections of the budget are elaborated hereafter.

19.3.1. Budget for Digitization

The digitization budget includes the costing for scanning, file format conversion, metadata entry, LTO backup and Blue Ray Disks for data transfer by district courts, data backup at JDR, and DR Site, and the management of overall digitization activity by the District Courts and High Courts. The courts are responsible for weeding out of records as per the record retention policies. The selection of records for digitization and weeding out of unwanted records has to be performed simultaneously.

Estimated Data Storage Requirements			
Sr.No	High Court Location	AIP (Disposed case Digitization) on Cloud (TB)	DR Site DataSize (TB)
1	NorthEast States	191.14	582.36
2	Orissa	511.19	1,563.06
3	Delhi	440.14	1,193.69
4	Punjab	1,287.74	3,945.30
5	Karnataka	1,599.95	4,901.19
6	Madhyapradesh	1,396.99	4,094.27
7	Uttarakhand	141.27	431.98
8	Uttarpradesh	2,227.64	6,738.22
9	Kerala	208.30	638.17
10	Rajasthan	1,975.47	6,047.16
11	Jharkhand	553.27	1,682.89
12	Chattisgarh	503.88	1,457.07
13	Madras	1,200.28	3,676.14
14	Andhra-Pradesh	1,404.59	4,303.28
15	Bihar	2,692.46	8,248.99
16	Calcutta	3,132.34	9,555.38
17	Telangana	984.12	3,010.95
18	Bombay	6,758.39	20,705.92
19	Himachalpradesh	275.57	842.63
20	Jammu & Kashmir	87.22	267.23
21	Gujarat	1,859.37	5,761.54
	Total	**29,431.32**	**89,647.42**

19.3.2. Budget for Judicial Digital Repository (JDR) Cloud

The JDR cost includes budget for scalable cloud infrastructure with a basic storage capacity of 2 petabytes (usable), 5 servers with 384 GB RAM, cooling system, UPS, Diesel Generator, audit, technical support and maintenance for 5 years. However, the actual cloud configuration or server with storage configuration is to be evolved separately for each High Court as per the estimated data size for next 5 years. As per the estimated projections in the table on data storage requirements, almost every High Court, along with the District Courts under its administrative control, is bound to require 1 to 7 petabytes of cloud storage with efficient search and retrieval in the near future. The High Court with smaller storage requirements from 25 to 300 TB may be provided with basic server + storage solution.

Additional storage space over and above the estimated data size needs to be provided considering that High Courts may require some working space plus they may have other types of data and applications, which is not considered in scope of this SOP.

The budget also includes the cost for creating the data centre environment, networking, fire safe vaults for keeping the LTO tapes and facility management and operating costs for the cloud hardware, data centre including the network, UPS, DG set. It may be possible to procure a higher configuration of cloud than the suggested configuration in the provided budget. The upgradation of the cloud for additional storage and servers is anticipated in the 5th year depending on the growth of data. The contingency and maintenance budget includes the budget for co-location of servers and facility management, audit and certification, surveillance audits and management of unanticipated incident handling.

Refer to the cloud specification given in Annexure II, which is suggestive. The actual cloud specification will have to be defined on the basis of a fresh study of the market, technology and costing trends.

19.3.3. Budget for JDR Data Management

The JDR data management salary budget includes the cost for following manpower:

- Digital Repository Manager – 1 Nos.
- Senior System Administrators - 2 Nos.
- System Administrators / desktop support engineers – 2 Nos.
- Digital Curator – 1 Nos

The High Courts are required to provide the following additional manpower for managing the activities of processing of the data from district courts.

- Director (Information Technology) – 1 Nos.
- Assistant Digital Repository Manager – 1 Nos.
- Digital Curators - (10 Nos)

19.3.4. Options for Cloud Services

The SOP proposes 3 options, which may be considered by the High Courts for availing the cloud services for digital preservation purpose.

Option 1:

The private cloud infrastructure and a data centre is established at the High Court which ensures that entire Judicial Digital Repository (JDR), data centre facility is under its administrative and technical control. The private cloud infrastructure is provisioned for

exclusive use by a single organization comprising multiple consumers (e.g., subordinate courts). It may be owned, managed, and operated by the organization, a third party.

A suitable agency may be engaged on contractual basis for supply, installation, integration, security audit compliances, maintenance, management and operation of the cloud infrastructure and data centre established at the High Court.

Many national banks are functioning as per the 1st option wherein the data centres are established within their premise and the facility management is contracted to an external agency.

Option 2:

The cloud hardware is procured by the High Court and housed / co-located in the external Data Centre Service Provider (DCSP) with biometric authentication and a lockable cage. This approach has many advantages as the ownership of the cloud hardware remains with the High Courts, which makes it possible for them to change the DCSP as necessary in the future. It also ensures ISO 27001 Information Security compliances, which may be already available with the DCSP. The DSCP takes care of the cloud operations and management under the supervision of a High Court official deputed in the DCSP.

Option 3:

The High Courts may consider availing the public cloud services offered by NIC (Meghraj) or the Cloud Service Providers (CSPs) empaneled by MeitY or the respective State Governments. The public cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business or government organization. The public cloud facility exists on the premises of the CSP.

In case of CSPs, the High Courts should negotiate appropriate policies to ensure full control over the data and the freedom to migrate the repository to another CSP if necessary.

19.3.5. Digital Preservation Tools and Software Solutions

The digital preservation tools and software solutions should enable achievement of technical compliance as per ISO 16363. The cost of software development is one time and independent of the number of courts. The scope of software development activity is defined as below:

- ISO 14721 - Judicial Digital Preservation System (JDPS)

- Portal with Distributed Search & Retrieval with Multilingual support depending on language of records
- Data Tracking System
- Digitization Quality Verification Tool
- Tool for integrity assurance with a digitally signed certificate
- Process Optimization and Automation
- LTO Cataloguing Software
- Interoperability measures as applicable
- Customization of software as per High Court specific requirements
- Training the Master Trainers
- Documentation
- Software Deployment
- Software Upgradation
- Technical Support and Maintenance

A separate system needs to be developed for managing the digitized pending records comprising of active and passive data.

It includes the cost of training the master trainers of State Judicial Academy (SJA). Later, the SJA should train the concerned staff which will be involved in managing the digital preservation activity.

Note:-

The per page cost derived at the end of budget cannot be used as a standard rule of thumb for measuring the overall cost, as the budget involves dissimilar items e.g. infrastructure cost, which is not proportional to page count. However, in hindsight, it provides a simplified understanding of the complex budget and therefore the entire budget is reduced to derive the cost of digitization on per page basis.

20. AI / ML based Applications Leveraging on JDRs

Judicial Digital Repositories (JDR) will ensure the reliability of the source, integrity and authenticity of the judicial records. The quality of metadata, organized cataloging of digital records, uniformity of file formats, full text indexing will support easy retrievability. Implementation of this SOP across all the High Courts and Supreme Court can ensure interoperability between different JDRs. Interoperable JDRs serve as the foundation for building AI/ML based applications for the judiciary as they leverage on massive volumes of data / digital records.

20.1. Intelligent Decision Support

The eCourts mission mode projects are aimed to make the justice delivery system affordable, accessible, cost effective, transparent and accountable.

JDRs established at the High Courts will prove to be a primary source of information for building intelligent decision support capabilities using Artificial Intelligence (AI) and Machine Learning (ML). The interoperability between multiple JDRs across different High Courts can widen scope of digital information resources with diversity, which are essential for any machine learning activity. But it is a fundamental requirement for the Supreme Court, High Courts and District Courts to agree on uniform standards and homogenous systems. Fragmented and heterogenous approaches may be counter-productive.

If the digitization efforts are invested with proper foresight and vision then in the near future AI and ML techniques can be evolved to automate certain aspects of decision-making process with efficiency and accuracy to accelerate the justice delivery.

20.2. Big Data Analytics for Identification of Similar Cases

Judiciary also requires to search the precedence of the judgments delivered in similar cases in the past in order to ensure parity and consistency. Simple keyword searches over the text can produce hundreds of search results and make the task very challenging to analyze each and every case to find similarity in content. Therefore, advanced Natural Language Processing (NLP) based techniques have to be applied to perform intelligent content analysis over the subject matter of the cases to identify similar cases. Such AI tools and techniques can be helpful in ensuring uniformity across judgements. Machine Learning (ML) tools are heavily

dependent on training models built using vast amounts of cleaned data which becomes available through JDRs for processing.

20.3. Machine Translation for Court Case Records

High Courts frequently require translation of court cases in English from the regional languages. India has around 22 official languages. Therefore, it is extremely critical to develop and use the solutions for machine translation from regional languages to English and vice versa.

20.4. Cross-lingual Search & Retrieval and Analytics

Cross-lingual search & retrieval and cross-lingual analytics across different Judicial Digital Repositories (JDRs) based on machine translation of court cases in different regional languages can also supplement enhancement of efficiency of justice delivery system.

20.5. Knowledge Modelling, Reasoning & Semantic Linking

In addition, data mining, knowledge modelling, reasoning, semantic linking and information extraction techniques can be also applied for improved comprehensibility, analytics and visualization capabilities for enhancing the judicial intelligence.

20.6. Long Term Sustenance

Digital preservation, infrastructure development and management, preservation of electronic evidence, cloud computing, cyber forensics, e-discovery, intelligent decision support, AI / ML based intelligent applications and multi-lingual analytics for acceleration of justice are extremely critical for meeting the daunting challenges of Indian Judiciary. The various areas of technologies for the judiciary mentioned here are interlinked, interdependent and they have to function together in an integrated manner.

Readymade/ off-the-shelf-solutions are unavailable to meet the emerging and future technological requirements of the judiciary considering the massive volume of pending court cases, freshly filed court cases, insufficient staff and the complexities posed by the diverse and exploding population of India. Therefore, the Indian Judiciary needs to collaborate with a technology partner with comprehensive strengths in abovementioned areas of technology on a long-term basis for sustaining the digital preservation infrastructure and development of new technological solutions.

21. References

[1] Preservation of Electronic Legal Materials, UELMA Preservation Group, 2018

Accessed Date	17 July 2020
Access URL	https://www.aallnet.org/wp-content/uploads/2018/04/Preservation-of-Electronic-Legal-Materials-White-Paper.pdf

[2] Marlene Coir, Virginia C. Thomas, UELMA: The Uniform Electronic Legal Material Act, Libraries and Legal Research, Michigan Bar Journal, October 2014, 50-53

Accessed Date	17 July 2020
Access URL	https://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1092&context=libsp

[3] JTC Resource Bulletin, Developing an Electronic Records Preservation and Disposition Plan, Joint Technology Committee, National Centre for State Courts December 2014

Accessed Date	17 July 2020
Access URL	https://www.ncsc.org/__data/assets/pdf_file/0027/17694/6jtc-e-records-10-final.pdf

[4] Report of the Advisory Committee to Develop Policies for Retention, Destruction and Access to Electronic Court Records, Supreme Court Arizona, December 2013

Accessed Date	17 July 2020
Access URL	https://www.ncsc.org/__data/assets/pdf_file/0023/19535/handout-4-az-electronic-records-committee-report.pdf

[5] Fostering Innovation in U.S. Court System, Published by RAND Corporation, Santa Monic, California, Supported by National Institute of Justice (NIJ), U.S. Department of Justice, 2016

Accessed Date	22 July 2020
Access URL	https://www.rand.org/content/dam/rand/pubs/research_reports/RR1200/RR1255/RAND_RR1255.pdf

[6] National Study Report on Digital Preservation Requirements of India, Recommendations for National Digital Preservation Program (NDPP) – Volume -I, Ministry of Electronics & Information Technology, Government of India, 2010

Accessed Date	17 July 2020
Access URL	http://ndpp.in/index.php/national-study-report

[7] eGOV-PID Preservation Metadata Dictionary & Schema, Notified by MeitY, Notification No. 1(2)/2010-II December 13, 2013

Accessed Date	17 July 2020
Access URL	http://egovstandards.gov.in/sites/default/files/Metadata_of_Document_Sample.pdf

[8] Production of Preservable e-Records (PROPeR) Guidelines, eGOV-PID Preservation Metadata Dictionary & Schema, Notified by MeitY, Notification No. 1(2)/2010-II December 13, 2013

Accessed Date	17 July 2020
Access URL	http://egovstandards.gov.in/sites/default/files/Best%20Practices%20and%20Guidelines%20for%20Production%20of%20Preservable%20e-Records%20Ver1.0.pdf

[9] Jason R. Baron, Law in the age of exabytes: Some further thoughts on 'information inflation' and current issues in e-discovery search, Richmond Journal of Law and Technology, Vol. XVII, Issue 3, 2011

Accessed Date	22 July 2020
Access URL	http://jolt.richmond.edu/jolt-archive/v17i3/article9.pdf

[10] Proceedings of Digital Preservation and Development of Trustworthy Digital Repositories, C-DAC / APA International Conference, Published by Excel India Publishers, 2014

Accessed Date	11 October 2020
Access URL	http://www.ndpp.in/APA-DPDTR-2014/

[11] Sartor, Giovanni, Branting, Luther (Eds.), Judicial Applications of Artificial Intelligence, 1998, Springer Netherlands

[12] Indian Evidence Act 1872

[13] Information Technology Act 2000/2008

[14] IT Act Notifications GSR 582
[15] Right to Information Act 2005
[16] Public Records Act 1993
[17] ISO 16363 Audit & Certification of Trustworthy Digital Repositories
[18] ISO 14721 Open Archival Information System (OAIS) Reference Model
[19] ISO 13008 Digital Records Conversion & Migration Process
[20] ISO 13028 Implementation Guidelines for Digitization of Records
[21] BS 10008 Evidential Weight & Legal Admissibility of Electronically Stored Information (ESI)
[22] ISO 27001 Information Security Management
[23] ISO 19005-1:2005 Use of PDF 1.4 (PDF/A-1)
[24] ISO 19005-2:2011 Use of ISO 32000-1 (PDF/A-2)

Annexure I – Survey Forms

High Court Digitization Survey Form (I)

The following survey form was circulated among the High Courts for collecting information on the total volume of legacy records that are required to be digitized and the volume of existing digitized records.

High Court Digitization Survey form

This survey is being conducted in order to estimate the digitization requirements in High Courts.

Name of High Court:_____
Location: _____

1. Specify the Record Types:_____
2. Specify the Data size: _____ MB / GB / TB
3. What is the total volume of legacy records (total number of pages) that needs to be digitized?

   Total number of pages (approx.): _____
4. What are the languages in the legacy records?
   Languages :_____
5. Have you already digitized some records?

   ☐ Yes ☐ No

If yes, please specify the digitization specifications as below:

Master File format:

☐Uncompressed Tiff / ☐ JPEG2000 / ☐ Any other format_____
DPI:_____

Access Copy File Format:

☐PDF / ☐PDFA
DPI:_____

6. Have you performed OCR on digitized documents?

☐ Yes ☐ No

7. What is the total number of digitized pages?
   Total number of pages (approx.): _____
8. What is the total size of digitized documents?

   Total Size:_____☐MB / ☐GB / ☐TB

9. Do you have audio / video data that requires to be digitized?
Indicate the volume:_____

10. Do you have any other types of data that requires digitization?
Please specify:_____

11. What storage media is used for storing the data?

☐Hard Disc ☐LTO Tapes ☐DVDs

☐CDs ☐NAS ☐SAN

☐External HDD ☐Other

12. Whether the digital media are kept in e-records room?

☐ Yes ☐ No

13. Are you using any document management software?

☐ Yes ☐ No

If yes please specify name of the software: _____

Name of the Officer:_____

Designation:_____

Signature:_____

High Court Digitization Survey Form (II)

During the 1st round of survey, it was not clear whether the projections given by the High Courts are inclusive of the legacy records from district courts. Therefore, the following survey form was circulated again to get the clarity. In addition to the storage requirement for legacy records, we also need to get the future projection of storage requirement for the e-filing and born-digital records. Therefore, the following survey form was circulated.

High Court Digital Preservation Requirement - Survey form

Note: *This survey form is based on the new requirement generated out of the previous survey form*

Name of High Court :
Location :

New records at the High Court (Annual Projection)

- What is the volume of **new records** generated (to be digitized) by the High Court every year? (Approx. no of pages)
  - **Pendency till 2018: _____ (approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
  - **Institution in 2019: _____ (approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
    No of pages received electronically through e-filing software or electronic mode:_____
  - **Disposal in 2019: _____ (approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
  - **Disposed Cases: _____ (approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____

    Pages digitized in which format : - TIFF / JPEG
    Resolution of scanning (DPI) : -

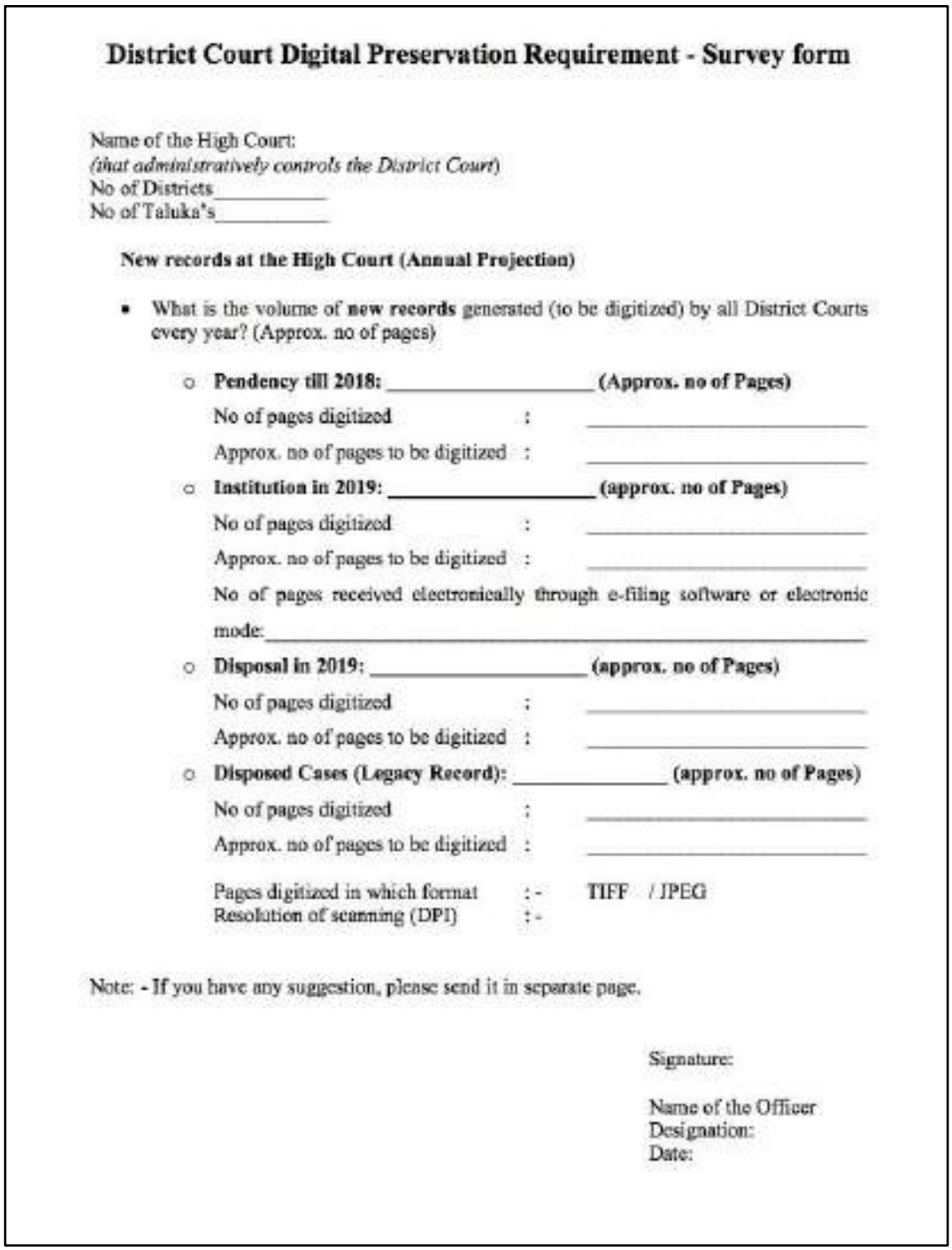
- What is the volume of the **new disposed cases** that require to be **permanently preserved** at the High Court with reference to above question? (like historical record)

  Approx. no of pages: _____

Signature:
Name of the Officer
Designation:
Date:

District Court Digitization Survey Form (III)

During the 1st round of survey, it was not clear whether the projections given by the High Courts are inclusive of the legacy records from district courts. Therefore, the following survey form was circulated again to get clarity. In addition to the storage requirement for legacy records, we also need to get the future projection of storage requirement for the e-filing and born-digital records. Therefore, the following survey form was circulated.



District Court Digital Preservation Requirement - Survey form

Name of the High Court:
(that administratively controls the District Court)
No of Districts_____
No of Taluka's_____

New records at the High Court (Annual Projection)

- What is the volume of **new records** generated (to be digitized) by all District Courts every year? (Approx. no of pages)
  - **Pendency till 2018:** _____ **(Approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
  - **Institution in 2019:** _____ **(approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
    No of pages received electronically through e-filing software or electronic mode: _____
  - **Disposal in 2019:** _____ **(approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____
  - **Disposed Cases (Legacy Record):** _____ **(approx. no of Pages)**
    No of pages digitized : _____
    Approx. no of pages to be digitized : _____

Pages digitized in which format : - TIFF / JPEG
Resolution of scanning (DPI) : -

Note: - If you have any suggestion, please send it in separate page.

Signature:

Name of the Officer
Designation:
Date:

High Court Data Centre Survey Form (IV)

The following survey form was designed and circulated among the High Courts to study the availability of cyberinfrastructure.

High Court Data Centre Survey form

(This survey form is regarding the available of Data Centre at the High Court)

Name of High Court:
Location:

(Definition: Data Centre means a specially designed environment for keeping servers, computer systems, digital storage, network)

A. Data Centre Environment

1. Does the High Court have a dedicated Data Centre Environment?

   ☐ Yes ☐ No

2. If Yes, is any cooling system present to maintain the necessary temperature for the Data Centre?

   ☐ Yes ☐ No

3. Does the Data Centre have Uninterrupted Power Supply (UPS) support for battery backup?

   ☐ Yes ☐ No

   a. If Yes, please specify the capacity of the UPS (kVA)._____

4. Does the Data Centre have Diesel Generator (DG) backup power supply?

   ☐ Yes ☐ No

   a. If Yes, please specify the capacity of the DG set (kW / kVA)._____

5. Please specify the dimensions of the Data Centre Room. (Not included UPS Room)

   Length (ft) _____ X Breadth (ft) _____

B. Rack Information

1. Please specify the number of the **server racks** present in the data centre:_____
2. How many more server racks can be accommodated in the data centre?_____

C. Server Information

1. Brief details of the server using in the Data Center:-

S.No	Name, Model and year of purchase of the server	Technical Specification of the server.(like Storage, RAM, Core etc)	Name and brief description of the Application running in the server. Also mention the software the application was developed	No of users access this application per day or data replicated to the other server on day to day basis.

Page 1 of 2

2. Please specify the storage capacity of the SAN or NAS, if available.

_____

3. Availability of Precision Air Conditioner in the Data Center.

☐ Yes ☐ No

a. If Yes, please specify the Number and Capacity of the Precision Air Conditioner

_____

4. Any provision of Disaster Recovery Center.

☐ Yes ☐ No

a. If Yes, brief note on Disaster Recovery methods using in Data Center.

_____

5. Please specify the applications running on the Server.

_____

D. Software:-

S.No	Name of the software like RHEL, MySQL, Windows server, Virtulaization etc	No of the licenses	Purpose

E. Connectivity: - Brief description of Internet or lease line connection in the Data Center including District Court.

S.No	Network Type	Service Provider	Bandwidth	Expenditure, if bear by the High Court.

F. Technical Manpower of the High Court

S.No	Designation	No of Post	High Court or District Court	Duties assigned.

****

Page 2 of 2

Annexure II – Cloud Specification

The basic cloud specification for the High Courts is provided below. This is suggestive only. The cloud specifications will need to be redefined again at the time of tendering based on fresh study of market, technology trends and prices. Basic and scalable cloud infrastructure is proposed while considering the fact that the High Courts will also use this cloud for storing other data (AV and e-evidence) and variety of other software applications including the upcoming big data analytics, AI/ML based intelligent tools on top of the judicial digital repository.

Sr. No	Item	Unit
1.	**Server: Compute Cluster** Up to two 2nd Generation Intel® Xeon® Platinum 8276L Processor or Equivalent or Better (# of Cores 28, # of Threads 56, Processor Base Frequency, 2.20 GHz, Max Turbo Frequency 4.00 GHz), 384GB RAM, 480GB SSD SATA HDD	5
2.	**Server: Management nodes** CPU 2 x Intel® Xeon® Processor E7-4850 v4 or Equivalent or Better (Cores 16, Threads 32, Processor Base Frequency 2.10 GHz, Cache 40 MB), 96GB RAM	2
3.	**Storage** 2PB SAN Storage Model, scalable up to 10PB	1
4.	**SAN Switch** Brocade 6510 48-Port 16 Gbps Fiber Channel SAN Switch, Front-to-Back Airflow	2
5.	**L3 Switch** Broadcam, 48 Port with 10G ports	2
6.	**Firewall** (1 GB) with unified threat management (UTM) capabilities	2
7.	**Tape Library** LTO-8 Tape drive ,12TB native capacity	1
8.	**Server Rack** With cooling, biometric security, Camera Based Surveillance, Analogue / Digital KVM	1

Software

Sr. No	Item	Unit
1	VMware vCloud suite or Equivalent or Better	14
2	Backup software	1
3	Web Application Firewall	1