An Overview of Digital Preservation Considerations for Production of "Preservable" e-Records: An Indian e-Government Case Study

Dinesh Katre Centre for Development of Advanced Computing (C-DAC) NSG IT Park, Aundh, Pune 411007, India 91-20-25503386

dinesh@cdac.in

ABSTRACT

In the Indian context, when the e-government records are received for archival purpose, it is observed that very often they are produced without proper compliances for long term digital preservation. This paper presents a case study of e-district Mission Mode Project which offers diverse citizen services and produces the e-records such as birth certificates, domicile certificates, marriage certificates, caste certificates, etc in very large volumes. Such born digital e-government records have to be retained and preserved for technological and legal reasons. The Centre of Excellence for Digital Preservation established at C-DAC, Pune, India has carried out the study of e-record production process in the e-district and the need analysis for its digital preservation. The digital preservation best practices are identified, which have to be incorporated in the production process of erecords, so that the final e-records are produced in "preservable" form with full compliance as per the requirements of OAIS.

Categories and Subject Descriptors

E.3 [Data Encryption] Public key cryptosystems

H.3.2 [Information Storage]: File organization, Record classification

I.7 [Document and Text Processing]: Document management, Document preparation, Format and notation, Markup Languages, Standards

J.1 [Administrative Data Processing]: Government

General Terms

Documentation, Design, Standardization, Theory, Legal Aspects.

Keywords

e-Government, Digital Preservation, Electronic Records, Fixed Digital Object, Significant Properties, Preservation Description Information (PDI), Submission Information Package (SIP), Open archival Information System (OAIS)

1. INTRODUCTION

1.1 Growth of e-records in India

The Indian government is spending more than 10 billion dollars on e-governance through its National e-Government Action Plan (NeGP) [2]. It has already launched 27 Mission Mode Projects which includes central, state and integrated MMPs such as Banking, Central Excise & Customs, Income Tax (IT),

Insurance, National Citizen Database, Passport, Immigration, Visa and Foreigners Registration & Tracking, Pension, e-Office, Agriculture, Commercial Taxes, e-District, Employment Exchange, Land Records, Municipalities, Police, Road Transport, Treasuries. Citizen Service Centres, e-Biz, e-Courts, e-Procurement, etc.

The forthcoming Electronic Service Delivery Bill which is awaiting to be passed by the Indian parliament will make it mandatory for all government organizations and departments to offer the citizen services through electronic media within next 5 years. Enlarging volumes of e-records, e-documents and digital information are anticipated to be produced through these initiatives by the Indian government.

1.2 Legal framework

The Indian laws which clearly spell out the legal obligation of government organizations to preserve the electronic records are briefly introduced in this section.

1.2.1 Information Technology Act 2008

As per the IT Act, conditions for retention of electronic records are specified as - "accessibility so as to be usable for a subsequent reference; retention in the format in which it was originally generated, to represent accurately the information originally generated, with the details, which will facilitate the identification of the origin, destination, date and time of dispatch or receipt of such electronic record" [9].

1.2.2 Public Records Act 1993

As per the Public Records Act, 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 [18].

1.2.3 Right To Information Act 2005

As per the Right To Information (RTI) Act, every public authority is obliged to maintain all its records duly catalogued and indexed and to 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 [20].

Apart from these, there are several other laws in the Indian constitution such as Copyright Act, Banker's Book Evidence Act, Indian Evidence Act (medico legal requirements) which also

emphasize the need to preserve the electronic records for various reasons.

1.3 India's National Digital Preservation Programme

The author of this paper was entrusted with the responsibility to prepare the National Study Report on Digital Preservation Requirements of India [16], as the first step towards formulating the Indian National Digital Preservation Programme of the Ministry of Information and Communications Technology, Government of India. The report included the recommendations of 30 experts from diverse domains across India. As per the study report, the Indian digital preservation scenario is observed as under-

- It is necessary to first establish what an e-record is in principle and how it can be recognized in electronic environment for preservation purpose [1, 3].
- The e-records are threatened by the continuing changes and obsolescence of computer hardware, software, file formats, storage media; and also the other dangers like data corruption, physical damage and disasters.
- There is lack of awareness about the need to preserve the erecords and the legal implications of failing to do so.
- There is absence of procedures and infrastructure for preserving the e-records.
- e-Government systems are being developed without incorporating the digital preservation consideration so that the e-records produced are preservable and comply with the minimum requirements of Open Archival Information System (OAIS).
- The present Departmental Record Officers (DROs), Record Keepers and Archivists working with the record producing agencies in India do not have the technical skills and knowledge of digital preservation [15, 17].

1.3.1 Centre of Excellence for Digital Preservation

Therefore, as per the recommendations given in the National Study Report on Digital Preservation Requirements of India, the Ministry of Information and Communications Technology, Government of India has funded the proposal of C-DAC Pune to establish the Centre of Excellence for Digital Preservation. This project aims at developing the standards, best practices, tools and systems for the preservation of electronic records. More information is available at http://www.ndpp.in/. The author of this paper is the chief investigator of this project.

2. RELATED WORK

Though there is limited guidance available on long term digital preservation of e-government records, we briefly discuss the most notable international projects related to this topic in this section.

The Canadian research project "International Research on Permanent Authentic Records in Electronic Systems (InterPARES) [8] offers the principles and guidance for the record creators and preservers both, so as to ensure the preservability of e-records when they are produced.

The following principles given by InterPARES are applied in our casestudy -

- The record creation process must be integrated with the recordkeeping rules with specific business processes [3].
- Digital objects must have a stable content and a fixed documentary form to be considered records and to be capable of being preserved over time.
- Preservation considerations should be embedded in all activities involved in record creation and maintenance if a creator wishes to maintain and preserve accurate and authentic records beyond its operational business needs.

National Archives Records Administration (NARA), USA provides record management guidance on digitally signed documents [19]. The following observation of NARA is particularly relevant to our case study –

• Since litigation will typically occur after the expiration of a public key certificate, it is important to take steps to ensure that pertinent records remain available after the certificate has expired. It is equally important that they be complete and understandable without the need for technical interpretation, to the extent possible.

Minnesota State Archives offers a broad strategy on E-records Management [4]. The National Archives of UK also provides the e-Government Policy Framework for Electronic Records Management.

However, the technical details and guidance provided by InterPARES and NARA were particularly helpful to us in understanding various aspects of e-records preservation. ISO/TR 15489 on Information and Documentation - Records Management is also very helpful in understanding the characteristics of records.

3. SCOPE

During our research on digital preservation of e-government records so far, we have come across following distinct categories of e-records -

E-records with fixed information content

A process which culminates into a final certificate or an official document with fixed information content of long term importance. The final e-record is to be retained and used as it is, without requiring any further processing or alteration.

Incrementally evolving e-records

A process in which new information is added into the e-record over a period e.g. banking transactions or change in the property ownership in land records. In such e-records the historical information of past transactions continues to be importance for preservation.

In this paper, we have focused on the digital preservation considerations for "final e-records with fixed information content" like birth certificate or domicile certificate issued to Indian citizens through the e-district Mission Mode Project (MMP).

4. CASE STUDY

We have chosen e-district Mission Mode Project (MMP) as a case study to build the pilot digital repository of e-records. The edistricts are offering following type of services to Indian citizens [21]-

• Creation and distribution of certificates for income, domicile, caste, Birth, Death etc.

- Arms Licenses, Driving Licenses, etc.
- Public Distribution System (PDS): Issue of Ration Card, etc.
- Social Welfare Schemes: Disbursement of old-age pensions, family pensions, widow pensions, etc.
- Marriage Registration, Land Records, etc.

Many services offered through e-district are producing large volumes of certificates which are authorized with digital signature. The certificates like birth certificate, marriage registration certificate, domicile certificate, caste certificate produced through electronic means need to be preserved as per the applicable retention rules and legal requirements. In this paper, we have focused on the digital preservation considerations related to certificates (birth, caste, marriage, domicile, etc) produced by e-districts.

After seeking due permissions, our team visited multiple edistricts, studied the system architecture and workflow, collected the sample database and certificates.

5. NEED ANALYSIS

The e-government systems should be designed to incorporate the following digital preservation considerations so as to produce the preservable e-records.

5.1 Need of e-record objectification

We observed that the e-district maintains a database comprising of various information elements and images pertaining to millions of certificates issued to various citizens. In one of the e-districts, the size of the database file was close to 3 TB, which is inflating everyday with the addition of new certificates issued to the citizens. The final certificate is dynamically rendered in the browser as per the layout specifications. The final certificate is not given an object form with fixed information content.

					11.11	1141141	Date 1	1.00	141.02									
	and the second	1.44	101.04	and the other														
	(and be	1		1403-0010	nen lä													
13	j had	12.50	-															
		-	dente.	rection.	Applement, no	last, rev.	1000	141,78	Salar Jac Just	in the	hanne	profess	danet.	-			34	add Parts
		- 040	med.a.	0960033	ON Secold CODED	•	414 10	4		100.00	# 145	hold we	904	3452	100	MAL .	36.05.184	
0.4		882	mea.c.	0100043	9/9/Seeu08.CD0824		nue but			R 1874		-	-004	0052	1000	MILL.	10151078	
		and a	mail.	0.001	03-1-multi-Chillip	*	Good Sec.			411.00	84	table many for planes	- 10	982	100	mag.	01011000	
		1000	mild.	-	101 Sec. (31) (24) (31)	•	due but			No. of Lot, No.	241	Suppliers given			100	NULL.	(6-1-16)	
10		-	make	country)	08-Denoitk-C00027	•	Name Real.	4		Read to the	342	and an an an an	000	0005	-	MAL	30.061688	
1.4		102	med.L.	0100043	6/8-Seeuldt COMUS	*	ADS AND			new my	201234	abet	008	0052	100	MILL.	11-06-1079	
		100	MAL.	and and	0110-0403 (D4003	•	Good helds			1204 28 169		100	100	995	-	mat.	10111-1040	
		-	MAL.	-	011 Tarrol 0 (2001)	•	*** (***				****	built we	-	962	-	NULL	10.54 10.5	
			HALL.	concret/	098-DamailA.CD0002	•	and			which these		808	008	3000	-	- Section	12-10-1018	
	1	982	MAL.	0.0001	Stitutes (KCD0001)	*	211.06			10, 1014 Mg	#18	Soland.	308	0052	100	NULL	21/06/1081	
		1000	MAL.	-	BED-MARCHEN	•	Apre are				1414	dy 10 thurses	104	9962	100	MAL.	24151879	
	1	100	MAL.	-	083mold/2022		and su			Review My	347	seeper ner	104	942	-	MILL.	10000-004	Sealan aut
	s., en	-	MAL.	payors,	OR Density COURSE	F	916 1026			spend silts	\$28	shine wards mines not sugarat	909	3000	-	may.	224717684	the sta
	8, 105	- 012	MAL.	0300082	ATE Sweatte CORDAN	*	-		•	NAME AND	2.810248	car we	308	9922	1000	MILL.	11.05.1888	2011 101
	1. 105	882	MAL.	4000041	895 Decard (C) (2014)	•	one has			****	407	first give too spice 9	0.0	0002	-	MAL.	30.01.1981	1.6
		100	MAL.	0.000	010 Deno10 (2004)		100.000			B12 310	303	Bark and	100	0.02	100	Anth,	10.13107	21.41
	1.08	3452	MALL.	0000042	031 Garantile (2000)		124 24			where and	8.14	formed and	100	100	1000	MILL	14111903	101010
		102	weat.	(menolec)	05-0ww.0540064	448	gune .			feet and	8.98.80	State and speed and a	008	0002	-	NG41.	10.121104	40 ALL 10 ALL 10
		842	mill.	manut.)	019-5 www.ittli-CD02049		Append here			face allow	8.1048.20	4049 MIN	0.08	3952	100	mid.L	2112186	\$238 \$258ex me
0.		100	Mild.	smody.	BR Savutte COURSE		and here			1000		on only said support		1002	100	. 10.64	14/01/1882	onary wer
		380	MALL.	01000032	818-Down/18-C200947		partie an			\$1.350 MV	10	Box with	209	0112	1004	MISA.	010101015	
0.1	2.105	882	MAL.	0101012	818-Serval (5-C) 00048		ARD \$MS			second all		and also likes?	1016	0052	1002	NG41	21-21-1625	ALC: NO.
		100	mild.	search j	10 Deva10 (2004)	*	\$10, \$210 \$E			the second per	1947 #122	any tip: anti-maps	10	1992	100	mag.	2101488	Ret
53		100	mili.	second.	IN Local Column		See red			on sheat by red	117	-		1942		MILL	26121978	***
1.1		980	MALL.	0000042	094 Ganual 8 (2000)		men per			OWNER COST		1974 Roug when and against	0.09	0002	-	mail.	20111-1081	ABOR HOR STREET
	s. 105	462	NULL	0101012	019-Terror06-CD40552	*	une be			1981 199	1100 303	ten Bo't unterprovement and	008	0052	100	NILL	10/12/1467	Test .
	r	and a	mat	eners)	03-2-mail:0-(5-000)	*	and the				10	the age with a millioner	-	9953	-	mail	11-15-1465	attent.
	1	100	Mili.	contrary.	00 (and 0 (2004)		due bet			****		grant Mark				MILL.		danci gener
		040	MAG.	0000102	10% DownEll CO 0009		tion here			-		gives them	004	1002	-	1044		gree Mark
	1.109	882	mean.	6181042	015-Desu(08-C240096	*	othe and			should use	18.8.15	arden effent	008	0052	1002	mids.	2148-1877	16.8.72 extent
ШÌ		1000	mail	embral (09-1-mail@120827	*	annes .			atte	8.22	All shelf affent offent	100	9855	100	mitt.	Janita hiero	#11 weight
		100	Mild.	000042	Infanti Cilli		The R			date:		-			-	811	0.411976	48
15		382	mail.	country.	018-Donald CD0008		400 mill	4		are used rett	1.288	Thomas Mad	009	0000	-	mai.	26-011076	2100 1204
ь	1. 105	362	NULL	0000342	09-1mm/08-020000		anty serve			Helle Field	492.00	without .	008	0052	-	Mid.		सहित समय
Шŝ		1000	MAL.	manual	III (search COUNT		10.0			also the series	\$255,111	BU THELTON ARMINE	100	1000	100	mat	NUCLINES.	ardinese.
			8011	-	298.004030 (200001)		Aug			on summer the	12.000.07	project tota artificant	199			10011	10.11.190	souther not selling

Figure 1. A database with data pertaining to millions of certificates

As per our assessment, the current approach poses following digital preservation challenges.

The digital information pertaining to certificates stored in the database is a result of the business logic which involves workflow, programme instructions, data structures, dependencies between values, formulas applied for calculated values and functions in force. Therefore any change in the business logic, representation logic and rendering logic can change the content of the certificate in an undesirable manner. Typically, the "current"

and transactional information should be maintained in the database. The final or "non-current" certificates should be given a fixed object form for long term preservation.

Refer figure 2 to understand the vulnerability to undesirable changes in e-records when they are under the influence of business logic.

Therefore, after the e-record is finalized, it is necessary to delink it from the business logic and fix it in the form of a self contained digital object for the purpose of preservation.



Figure 2. Vulnerability to undesirable changes in e-records

5.2 Need to digitally sign the entire certificate

Lets briefly understand the workflow of the e-district. The Citizen Service Centres (CSC) established in small towns and villages are connected to the e-district. The citizens are able to submit their application for certificate along with necessary documents and proofs with the CSC. The CSC operators digitize the applications and upload it for verification and issuing of certificate. The district authority verifies the documents and then grants the certificate (depending on the type of request) authorized by affixing the digital signature to "selected information values" (such as date of birth, name of person, etc) in the database. The digital signature is then stored in the database. The CSC is notified when the certificate is authorized. The approved information content is rendered in the browser as shown in figure 3 and then printed on standard stationary (paper) for issuing it to the citizen as shown in figure 4. It is a hybrid approach, in which the key contents of the certificate are born digital and digitally signed but the final certificate issued to the citizen is printed on paper.

In this process, many significant properties of the certificate such as layout, border, emblem, watermarked image, the authorization of state government regarding legal acceptance of digitally signed certificate are getting added only through the printed stationary. These significant properties are not part of the digitally signed information content of the certificate stored in the database. The affixing of digital signature to selected information values in the database ensures its integrity but it does not certify or authorize the final certificate as shown in figure 4.



Figure 3. Rendering of a demo certificate with digital signature in browser



Figure 4. A digitally signed certificate printed on stationary paper

Although, our intention is not to comment on the process of authorization but from the digital preservation perspective, the difference between what remains in the database (refer figure 1) and what is issued as the final certificate (refer figure 3) is notable. Ideally the certificate issued to the citizen and the certificate retained for preservation must be exactly the same in terms of its logical and conceptual representations. To further substantiate this point, as per Duranti et al the form of transmission of a record is the physical and intellectual form that the record has when it is received; and the authenticity is best ensured by guaranteeing that a record maintains the same form through transmission, both across space and through time [3].

5.3 Need of significant properties

The significant properties are those characteristics [technical, intellectual, and aesthetic] agreed by archive or by the collection manager to be the most important features to preserve over time [5]. In case of the certificates as shown in figure 4, the significant properties such as layout, border, emblem of the state government, font style for logo and color scheme are added only through the printed stationary. The dynamic on-screen rendering of certificate is dependent on browser and display settings. It may render differently on different computers. Therefore, the minimum essential significant properties of the certificates must be purposefully designed and embedded in its digital rendering and given a fixed form for long term preservation. The significant properties are helpful to the curators in asserting or demonstrating the continued authenticity of objects over time, or across transformation processes [7].

5.4 Need of file naming policy

It is observed that the file names generated by the e-government systems follow some type of incremental numbering system but such filenames are not adequate to be consistent, meaningful, unique and parseable. For example, the files pertaining to birth certificates, domicile certificates, marriage certificates, etc can be categorized by pre-fixing a standard code or short forms such as BC, DC and MC in the file name. It will be so helpful in categorizing the certificates based on file names.

5.5 Need of Preservation Description Information (PDI) along with certificates

It is observed that most of the e-government projects are focused on offering the citizen services but no consideration is given to how the e-records produced by the e-government systems will be preserved for future. It is possible to capture some parts of the Preservation Description Information (PDI) through e-government system itself while producing the final e-record or the certificate. The final digital object must accompany the PDI with minimum essential metadata for it to be acceptable as a valid Submission Information Package (SIP) for the Open Archival Information System (OAIS) [12]. If we consider the huge volumes of erecords it is not practically possible to generate the PDI in a post facto mode at the time of archival and therefore, we suggest that it should be automatically captured when the e-record is produced.



Figure 5. Need of Preservation Description Information (PDI) as per OAIS

If the certificates produced through e-district were to be discovered, read and understood in future then we will need to know the following – $\,$

- What is the identifier of certificate?
- To whom was it issued?
- When, where and who had produced it?
- What was the context in which it was produced?
- What was the basis on which the certificate was issued?
- Which software was used for producing the certificate?
- In which file format was it stored?
- How to know that the certificate available in the archive is the authentic one?
- What can be given as the proof or evidence of its authenticity?
- How to know if the certificate has not been modified?
- Does it require to be converted in the latest file format to be able to render it and read it?
- Who is authorized to access and read the certificate?

The answers to these questions are normally found in the Preservation Description Information (PDI).

Due to unavailability of the standard vocabulary and metadata schema, the present e-government systems are not able to produce the PDI along with the final e-record.

In this context, while exploring the ways in which PDI could be described for e-records, we came across the application of DSpace for cataloging of court case records which were described using Dublin Core Metadata Elements as shown in figure 6.

One can notice that the names appearing in front of dc.contributor.author are the names of judges who passed the final judgment on the court case.

The names appearing in front of dc.contributor.editor and dc.contributor.illustrator are the names of Petitioner and Respondent in the particular court case.

Full metadata record										
DC Field		Value			Language					
dc.contributor.author	N.K.PATIL AND K	.N.KESH	IAVANARA'	YANA	en_US					
dc.contributor.editor	S B KRISHNAMU	RTHY S/	O BOREGO	WDA	en_US					
dc.contributor.illustrator	K N NARASIMHA	IAH S/O	NARASANI	NA	en_US					
dc.coverage.spatial	Bangalore				en_US					
dc.date.accessioned	2009-10-23T08:	14:33Z			-					
dc.date.available	2009-10-23T08:	14:33Z			-					
dc.date.created	2009-10-05				-					
dc.date.issued	2009-10-23T08:	14:33Z			-					
dc.identifier.isbn	2001				en_US					
dc.identifier.issn	2006				en_US					
dc.identifier.sici	8				en_US					
dc.identifier.ismn	MFA				en_US					
dc.identifier.uri	http://hdl.handle	e.net/12	3456789/2	09929	-					
dc.description.abstract	MFA 9233/2005	en_US								
dc.title			en_US							
Appears in Collections: Miscl. First Appeal - MFA										
Files in This Item:										
File	Description	Size	Format							
MFA2001-06-05-10-2009.	odf 34	46.45 kB	Adobe PDF	View/O	pen					
	Show simple item re	ecord								

Figure 6. DCMES applied through DSpace to describe a court case record

It is obvious that the Dublin Core Metadata Elements are more suitable for describing the resources like books and not suitable for court cases or certificates or e-government records. It is also very misleading, as the judges are mapped as the authors, petitioner is mapped as the editor, and the respondent is mapped as the illustrator. Also, the court cases do not have ISBN.

Therefore, a suitable metadata schema with appropriate vocabulary (which represents the local understanding) is needed for the description of certificates and e-government records in Indian context.

The requirements identified so far are part of the packaging process involved in the making of a Trustworthy Digital Object (TDO) [6].

6. BEST PRACTICES AND GUIDELINES

Based on the study of workflow and characteristics of egovernment records (certificates) produced through e-district, the Centre of Excellence for Digital Preservation has identified following best practices and guidelines for production of preservable e-records.

6.1 The final certificate as a fixed digital object

As per the findings of Canadian InterPARES 2 (International Research on Permanent Authentic Records in Electronic Systems) project, the preservation considerations should be embedded in all activities involved in record creation and maintenance if a creator wishes to maintain and preserve accurate and authentic records beyond its operational business needs. ISO/TR 15489-2 for Information Documentation - Records Management Guidelines also specifies the need to capture the e-record with fixed representation of actions [13]. Therefore, the final contents (information + images + significant properties) of the certificate produced by e-district should be given a composite and fixed object form.

Selection criteria for objectification of e-record

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

- The e-record is meant to be used as a certificate or a final statement proof
- The legal obligations and implications of failing to reproduce such e-record in its original and authentic form in future
- The value of information contained in the e-record
- The e-record forming a basis or dependency for other transactions
- The historical significance of the e-record
- The retention rules pertaining to such e-records
- The record keeping and preservation policy of the record producing organization

Typically the e-records like birth certificate, domicile certificate, marriage registration certificate, death certificate, senior citizen certificate, insurance policy, ration card, passport, income tax return, mark sheet, service record or documents such as MoU, contract, agreement, parliamentary bills / acts, court case judgments along with proceedings, user manuals which need to be retained for various reasons (like legal, value of information, historical importance) can be considered to be produced in the form of a digital object with fixed information content.

6.1.1 The criteria for not giving a fixed object form to e-records-

The following type of e-records need not be given an object form based on following criteria-

- The e-record has temporary significance
- There are no legal obligations or implications for not maintaining such e-record beyond its purpose of use
- As per the retention rules such e-record is not required for more than 5 years (in that case it can be maintained in the database)

6.2 The PDF for Archival (PDFA) format specification for final certificate

As per our study, some e-government systems are producing the proprietary Adobe PDF output which is not recommended for preservation. Therefore, the final e-records like certificates should be objectified in the form of PDF for Archival format specified as under-

ISO 19005 PDFA-1a is recommended for archival of "born digital documents" [10].

ISO 19005 PDFA-1b is recommended for archival of "reformatted digital documents" (for example composite PDF comprising of TIFF images).

PDFA-2a [11] can also be used but PDFA-1a and PDFA-1b is adequate in the present context.

6.3 Conceptual representation of certificate

An e-record in the database is nothing else but digital information distributed in various tables of the database. It forms the logical representation of the given e-record. The conceptual representation of e-record covers the rendering attributes and visual appearance which are essential for human sensorial understanding of the e-record. Most e-government systems are designed to store the logical representation of e-records. Such systems do not address the requirements of the conceptual representation of e-records which is necessary to be captured while producing the final digital object in the PDF/A format.



Figure 7. Need to specify the conceptual representation of e-record

We need to specify the significant properties of e-record so that its conceptual representation or the rendering aspects get properly addressed.

6.3.1 Significant properties for certificates

Following type of significant properties should be embedded in the final e-record at the time of objectification.

Proper page layout (page size, orientation, margins)

- Tables with specifically defined columns, rows and cell spacing
- Emblem / logo of the organization with proper color specification / color code
- Header and footer information
- Font specifications, style settings for titles and the textual information
- Bar code
- QR code
- Images with specific DPI, dimensions and format
- Watermarked image
- Fixed location coordinates for images
- Fixed location coordinates for digital signature



Figure 8. Significant properties of a certificate

6.3.2 Why significant properties are important for preservation of e-records?

The significant properties are extremely helpful in fulfilling the requirements of usage, authentication, preservation and several legal obligations which are enlisted below-

- Legal obligation as stated in IT ACT 2000 (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" [9].
- ISO 14721:2003 OAIS [12] specifies and refers this requirement as Information Properties needed for preservation.
- Meaningful understanding and usage of information
- Verification of originality and authenticity of e-record
- Renderability of contents exactly as original in future, even if the present document format or software becomes obsolete
- Reconstructability of the digital object by using its elements

6.4 Consistent and logical file naming policy

The record producer (e-district) can select at least 3 to 4 relevant file name elements as per the examples given in this section for defining the logical and consistent file naming policy. Appropriate abbreviations / short forms can be used along with separators and incremental serial numbers. We must avoid using the controlled characters and empty spaces in filenames. The filename / length / character sets should be compatible across operating systems / file systems. Examples of file name elements are given below-

- Type of certificate
- Service code
- Reference number / accession number
- Place
- Date of creation
- Name of creator / organization
- Title of content
- Department number
- Name of organization
- Records series

6.5 Affix the digital signature to final e-record in PDFA format

- After completing all information processing the final erecord is produced in the form of PDFA document and then it should be digitally signed by the competent authority for authorization and non-repudiation.
- The PDFA document could be printed on the standard stationary paper for issuing to the citizen.
- The PDFA document is then submitted for archival and preservation, which has the required significant properties.

6.6 Capture the Preservation Description Information (PDI) of final e-record during its production process

The Centre of Excellence for Digital Preservation has defined a comprehensive metadata schema titled as "E-governance Standard for Preservation Information Documentation of E-records (E-Gov SPIDeR) based on the types of e-records produced in the Indian context.

We have studied the existing metadata schemas like Dublin Core, MODs, METS and PREMIS. The designers of these metadata schemas have considered wide range of objects and it reflects the state-of-the-art and maturity of archiving practices in the developed countries. As per our assessment, the existing metadata schemas are too exhaustive and not perfectly fitting in the context of Indian e-government records.

We needed something smaller, simpler and yet comprehensive which could capture the minimum essential preservation information at the time of record production itself. Therefore, we have defined our own metadata schema for the description of erecords which reflects our local understanding and requirements. It is a hybrid metadata schema which includes our own contributions in addition to the selected metadata elements from the established schemas. The major sections of the e-Gov SPIDeR metadata schema are briefly explained here as it is not possible to reproduce the entire schema due to space limitation.

Cataloging Information

The cataloging metadata for e-records retains some of the Dublin Core metadata elements with new additions like RecordIdentifier, RecordType, MainCategory, SubCategory, NameID, OfficeType, Validity and RetentionDuration. The Paris Principles for cataloging [14] are adopted for defining the common cataloging parameters for electronic records [13].

Enclosure Information

The final e-record (e.g. the certificates issued by e-district) is generated on the basis of various documents, proofs and correspondence which are enclosed with it. The enclosure information is needed for establishing the context in which the erecord was produced. The list of enclosures can be included in the PDI if applicable. The accuracy of the final e-record can be verified and validated on the basis of the enclosed documents.

Provenance Information

It includes the address of Citizen Service Centre (CSC) that received the application for certificate, the office address of edistrict which issued the final certificate and the device IDs of the servers where the request was processed and final certificate was issued.

Representation Information

It includes the names and version information of software, operating system, compiler, API Library, application, tools, web browser, database, etc which was used for creating the final erecord and the software necessary for reading it.

Fixity Information

It includes the checksums of the final e-record (certificate) and its enclosures.

Digital Signature Information

Digital signature metadata portion is adopted from PREMIS.

Access Rights Information

The access rights metadata portion is adopted from METS.

It is ensured that the E-Gov SPIDeR metadata schema can be mapped with the established metadata schemas like Dublin Core.

As per our study, major portion of the metadata can be captured at the time of record production itself as the required descriptive information is either getting generated through the process or it is available in the database.

If this information is not captured during the record production then it is likely to remain scattered in e-government systems, and eventually it may be lost forever. Also, in the post facto mode it is difficult to gather the descriptive metadata for ingest and ensure its accuracy and authenticity.



Valid Submission Information Package (SIP)

Figure 9. Final e-record produced with basic digital preservation consideration

7. CONCLUSION

In case of e-Government Records, it is necessary to incorporate the basic digital preservation considerations throughout the erecord production. It is important to ensure that the e-government systems are designed and developed in such a way that the final erecords produced by them are "preservable" enough and comply with the requirements of the OAIS standard [12].

8. ACKNOWLEDGMENTS

The encouragement and support received from Department of Electronics and Information Technology, Government of India and C-DAC Pune is acknowledged with gratitude and thankfulness.

9. REFERENCES

- Acland G. I., 1996. Electronic Records: The View From Beyond OZ, Australian Society of Archivists Conference, Alice Springs, (May 1996).
- [2] All Govt. services to go online by 2014, 2010. Deccan Herald, Bangalore, Saturday, (Jul. 2010) 8.

http://www.mit.gov.in/sites/upload_files/dit/files/ApexMeeting_23072010_0.pdf

- [3] Duranti L. and MacNeil H., 1996. The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project, Archivaria 42 (Fall 1996): 46-67.
- [4] Electronic Records Management Guidelines, Minnesota State Archives http://www.mnhs.org/preserve/records/electronicrecords/erm s.html
- [5] Giaretta D., 2011. Advanced Digital Preservation, 1st Edition, Published by Springer Verlag, Berlin Heidelberg (June 2011), ISBN-10: 3642168086.
- [6] Gladney H. M., 2010. Preserving Digital Information, Published by Spinger Verlag, Berlin Heidelberg (2010), ISBN 978-3-642-07239-0.
- [7] Grace S., Knight G. and Montague L. 2009. Final Report on Investigating the Significant Properties of Electronic Content over Time (InSPECT), The National Archives of UK, (December 2009).

http://www.significantproperties.org.uk/inspect-finalreport.pdf

[8] InterPARES 2, 2008. International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records (March 2008)

http://www.interpares.org/public_documents/ip2(pub)policy _framework_document.pdf

- [9] Information Technology Act, 2008.
- [10] ISO 19005-1:2005 PDF/A-1
- [11] ISO 19005-2:2011 PDF/A-2
- [12] ISO 14721:2003 Open Archival Information Systems (OAIS)
- [13] ISO/TR 15489-1 and 2 Information and Documentation -Records Management
- [14] International Conference on Cataloguing Principles (Paris : 1961). Report. – London : International Federation of Library Associations, 1963, p. 91-96.
- [15] Katre D. S., 2009. Ecosystem for digital preservation in Indian context: A proposal for sustainable and iterative lifecycle model. In Proceedings of Indo-US Workshop on International Trends in Digital Preservation, (March 2009), Pune, India, 137–141.

http://ndpp.in/download/Indo-US-DP-Proceedings-C-DAC-2009.pdf

- [16] Katre D. S., 2010. National Study Report on Digital Preservation, Requirements of India. Volume I: Recommendations for National Digital Preservation Programme, Published by C-DAC, India, (2010).
- [17] Katre D. S., 2011. Digital preservation: converging and diverging factors of libraries, archives and museums - An Indian perspective, IFLA Journal, Vol. 37, no. 3, (October 2011), Sage Publications, London, UK, 195-203.
- [18] Public Records Act, 1993.
- [19] Records Management Guidance For PKI Digital Signature Authenticated and Secured Transaction Records by Federal Public Key Infrastructure Steering Committee Legal/Policy Working Group, National Archives and Records Administration, (March 2005).

http://www.archives.gov/records-mgmt/policy/pki.html#4-5

[20] Right To Information Act, 2005.