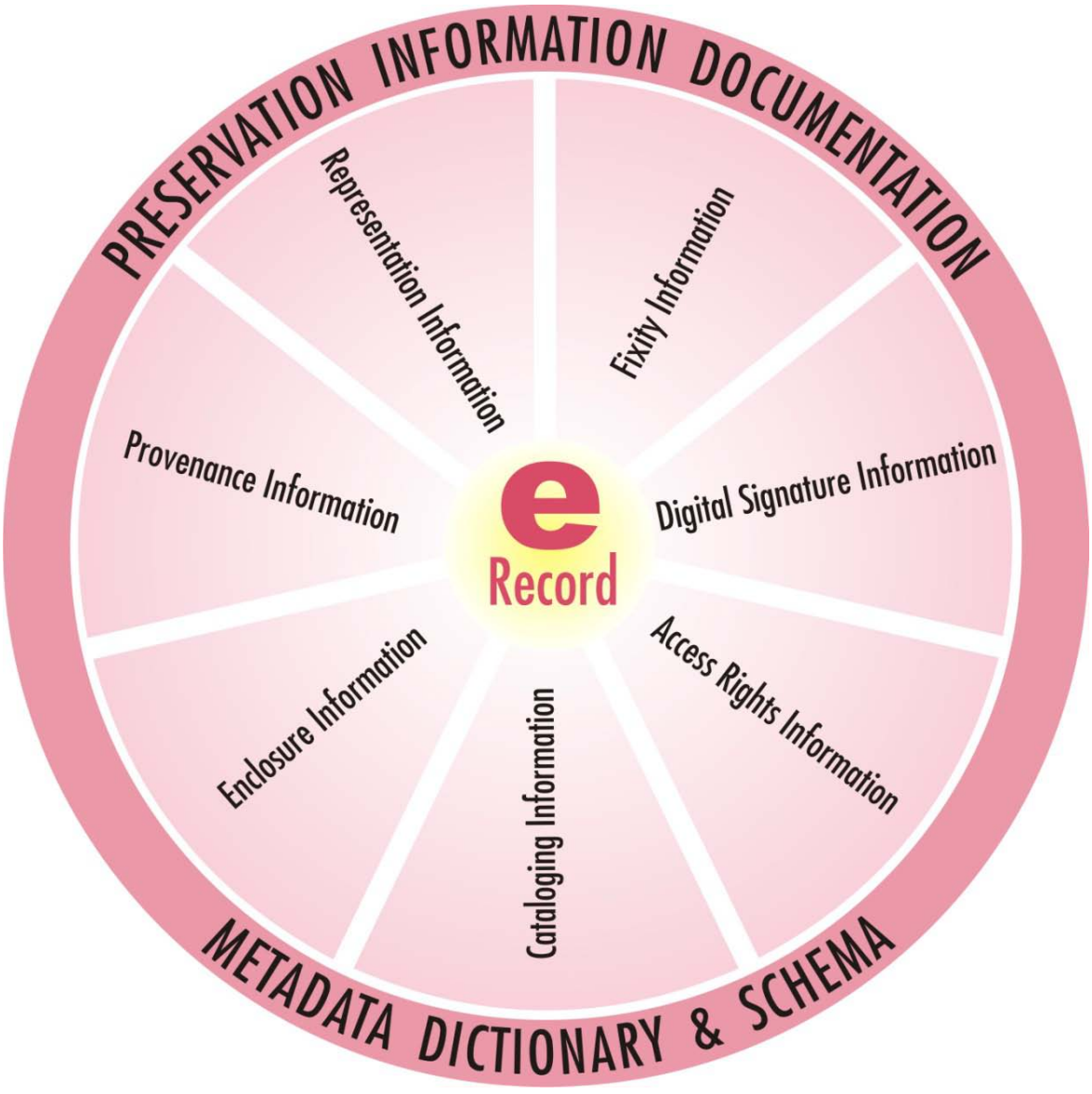
eGov.DP.01-02
Version: 1.0
December, 2013

eGOV-PID: Preservation Metadata & Schema

e-Governance Standard for Preservation Information Documentation (eGOV-PID) of Electronic Records



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

Metadata of the Document

S. No.	Data elements	Values
1.	**Title**	eGOV-PID: Preservation Metadata & Schema
2.	**Title Alternative**	e-Governance Standard for Preservation Information Documentation (eGOV-PID) of Electronic Records
3.	**Document Identifier** *(To be allocated at the time of release of final document)*	eGOV.DP.01-02
4.	**Document Version, month, year of release** *(To be allocated at the time of release of final document)*	Version 1.0 December 2013
5.	**Present Status**	Notified
6.	**Publisher**	Department of Electronics and Information Technology (DeitY), Ministry of Communications & Information Technology (MCIT), Government of India (GoI)
7.	**Date of Publishing**	06/12/2013
8.	**Type of Standard Document** *(Policy / Technical Specification/ Best Practice /Guideline/ Process)*	Technical Specification
9.	**Enforcement Category** *(Mandatory/ Recommended)*	Mandatory
10.	**Creator** ***(****An entity primarily responsible for making the resource)*	The Expert Committee for Digital Preservation Standards and Guidelines under the Chairmanship of Dr. Gautam Bose, Deputy Director General, National Informatics Centre (NIC)
11.	**Contributors** ***(****An entity responsible for making contributions to the resource)*	Centre of Excellence for Digital Preservation, Sponsored by DeitY, established at C-DAC Pune.
12.	**Brief Description**	The eGOV-PID provides a standardized metadata dictionary and schema for describing the "preservation metadata" of an electronic record. This standard proposes to capture most of the preservation information (metadata) automatically after the final e-record is created by the e-Government system. Such preservation information documentation is necessary only for those e-records that need to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. The implementation of this standard helps in producing the valid input i.e. Submission Information Package (SIP) for archival and preservation purpose as per the requirements specified in the ISO 14721 Open Archival Information Systems (OAIS)

S. No.	Data elements	Values
		Reference Model. The eGOV-PID allows to capture the preservation metadata in terms of cataloging information, enclosure information, provenance information, fixity information, representation information , digital signature information and access rights information.
13.	**Target Audience** *(Who would be referring / using the document)*	• E-record producers and data managers • Departmental Record Officers (DROs) record keepers, archivists and preservation officers • All stakeholders in central and state government, as well as public and private organizations involved in execution, design, development and implementation of e-Governance applications. • Central, state, district level archiving organizations
14.	**Owner of approved standard**	DeitY, MCIT, New Delhi
15.	**Subject** *(Major Area of Standardization)*	Digital Preservation
16.	**Subject. Category** *(Sub Area within major area)*	Preservation Metadata for Electronic Records
17.	**Coverage. Spatial**	INDIA
18.	**Format**	PDF
19.	**Language** *(To be translated in other Indian languages later)*	English
20.	**Copyrights**	DeitY, MCIT, New Delhi
21.	**Source** ***(****Reference to the resource from which present resource is derived)*	• ISO 15836:2009 Information and documentation -- The Dublin Core metadata elements • ISO/TR 15489-1 and 2 Information and Documentation - Records Management: 2001 • ISO 14721:2012 Open Archival Information Systems (OAIS) Reference Model • Metadata Encoding and Transmission Standard (METS), Library of Congress, 2010 • InterPARES 2, International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records, 2008 Adaptation of above sources is based on the research carried out by the team of Centre of Excellence for Digital Preservation Project at C-DAC Pune.
22.	**Relation** *(Related resources)*	This standard is to be used in conjunction with Best Practices & Guidelines for Production of Preservable Electronic Records (PRoPeR).

Table of Contents

Statement of Purpose ..... **iv**
Acronyms and definitions ..... **v**
Introduction **1**
1. Aim **1**
2. Scope **1**
3. Normative references **1**
4. Need for capturing the preservation information **2**
5. eGOV-PID for e-Records **3**
6. Guidelines for eGOV-PID metadata capture **3**
7. Principles and overview of preservation metadata **5**
7.1. Cataloging Information 5
7.2. Enclosures information 7
7.3. Provenance information 7
7.4. Representation information 8
7.5. Fixity information 8
7.6. Digital signature information 8
7.7. Access rights information 9
8. Metadata dictionary and schema **10**
8.1. Overview of schema definition for eGOV-PID 10
8.1.1. Schema definition for cataloging information 12
8.1.2. Schema definition for enclosure information 21
8.1.3. Schema definition for provenance information 23
8.1.4. Schema definition for representation information 27
8.1.5. Schema definition for fixity information 29
8.1.6. Schema definition for digital signature information 30
8.1.7. Schema definition for access rights information 38
9. Summary of best practices and guidelines **43**
10. References **44**
Annexure A. Implementation guidelines 45
Annexure B. Sample XML with preservation metadata 46
Acknowledgements **49**

Statement of Purpose

The e-Governance standard for Preservation Information Documentation (eGOV-PID) of Electronic Records provides standard metadata dictionary and schema for describing an electronic record. Most of the preservation information (metadata) can be automatically captured using this schema after the final e-record is created, as most of the required information is already present in an e-government system. Such preservation information documentation is necessary only for those e-records that need to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. The implementation of this standard helps in producing a valid input i.e. Submission Information Package (SIP) for archival and preservation purpose as per the requirements specified in the ISO 14721: 2012 Open Archival Information Systems (OAIS) Reference Model.

Acronyms and definitions

Archival

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

Certificate Authority (CA)

Certification Authority (CA) is an entity that issues digital certificates.

Content Information (CI)

A set of information that is the original target of preservation or that includes part or all of that information.

Data Dictionary

A formal repository of terms used to describe data.

Designated Community

An identified group of potential consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the archive and this definition may change over time.

DCMI

Dublin Core Metadata Initiative

Digital Object

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

Digital Signature

A digital signature is a mathematical scheme for demonstrating the authenticity of a digital message or document. A valid digital signature gives a recipient reason to believe that the message was created by a known sender such that they cannot deny sending it (authentication and non-repudiation) and that the message was not altered in transit (integrity).

e-Record

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

Extensible Markup Language (XML)

XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. It is defined in the XML 1.0 Specification produced by the W3C.

Long Term Digital Preservation (LTDP)

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

Metadata

The data which describes the e-record or digital object based on common parameters.

METS

Metadata Encoding & Transmission Standard

Open Archival Information System (OAIS)

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

Preservation Information Documentation (PID)

The information which is necessary for adequate preservation of the Content Information and which can be categorized as cataloging, enclosures, provenance, representation, fixity, authenticity and access rights. OAIS standard refers to this as preservation description information.

Record(s) Officer

The officer nominated by the records creating agency for proper arrangement, maintenance and preservation of public records under his charge.

Submission Information Package (SIP)

The SIP is an Information Package that is delivered to the repository and digital storage system for ingest. The valid SIP comprises of CI and PID.

XML Schema Definition (XSD)

XSD language offers facilities for describing the structure and constraining the contents of XML documents

Introduction

1. Aim

The e-Governance standard for Preservation Information Documentation (eGOV-PID) of Electronic Records aims at automatically capturing the preservation information (metadata) of an e-record through an e-government system, following the creation of the final e-record. The implementation of this standard will help in producing a valid Submission Information Package (SIP) for archival and preservation, as per the requirements specified in the ISO 14721: 2012 Open Archival Information Systems (OAIS) Reference Model.

2. Scope

Such preservation information documentation is necessary only for those e-records that need to be retained for long durations (e.g. 10 years, 25 years, 50 years and beyond) and the e-records that need to be preserved permanently. The preservation information to be captured is broadly categorized in terms of cataloging, enclosures, provenance, representation, fixity, authenticity and access rights.

3. Normative references

- Information Technology Act, 2000, Government of India
- Information Technology Act Amendment (ITAA) 2008, Standing Committee Recommendations, Government of India
- IT Act Notifications GSR 582, 6th September, 2004, Published by Ministry of Communications and Information Technology, Government of India
- Central Secretariat e-Manual of Office Procedure (CSeMOP), DARPG, Government of India, 2012
- Public Records Act, 1993, Government of India
- Extensible Markup Language (XML), World Wide Web Consortium (W3C)
- ISO/TR 15489-1 and 2 Information and Documentation - Records Management
- ISO 14721:2012 Open Archival Information Systems (OAIS) Reference Model
- ISO 16363: 2012 Audit & Certification of Trustworthy Digital Repositories

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

4. Need for capturing the preservation information

In this document, it is well emphasized that the e-records have to be preserved in such way that it should be possible to find, read, represent, render and interpret them accurately as original along with all associated information necessary for its comprehension in distant future. The following questions are bound to arise if the e-records were to be used in the distant future–

- What is the unique identifier of an e-record?
- 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 e-record was issued?
- Which software was used for producing the e-record?
- In which file format the e-record was stored?
- How to know that the e-record available is the authentic one?
- What can be admissible as the proof or evidence of its authenticity?
- How to determine if the e-record has not been tampered?
- Does it require to be converted in the contemporary file format to be able to render it and read it?
- Who is authorized to access and read the e-record?
- Which are the other e-records related with it?
- How long the particular e-record should be retained?
- If the retention period of the given e-record is over then should it be disposed off or it should be reviewed again for extending its retention?

There are innumerable questions as mentioned above which may arise in future. The consequences of not capturing the preservation information are as under -

- the vital information associated with the e-record will remain scattered and untraceable
- it will not be possible to capture and link the evidences that would help in proving the origin, identity, integrity and authenticity of an e-record which are essential to fulfill the legal requirements

- it will not be possible to plan the migration or conversion activities if the file formats become obsolete
- it will be difficult to find the e-record in future
- it will not be possible to archive and preserve the valuable e-records
- the e-record will be lost forever
- bitter legal consequences will have to be faced on failing to reproduce the e-record

Therefore, it is necessary to capture all essential preservation information in the form of metadata while producing the e-record itself, as most of this information is available in the e-government system or e-records creation system. Refer the sample XML with preservation metadata in Annexure B.

5. eGOV-PID for e-Records

In the context of e-Government, the Preservation Information is categorized in terms of cataloging, enclosures, provenance, representation, fixity, authenticity and access rights. As per the OAIS standard, the Content Information (an e-record) along with Preservation Information Documentation (PID) constitutes a valid Submission Information Package (SIP) as shown in figure 1.

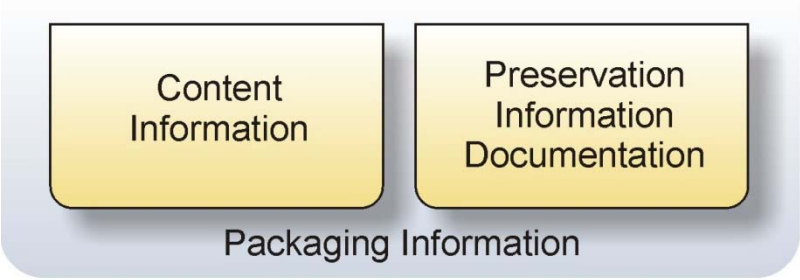


Figure 1. Submission Information Package (SIP)

Therefore, the e-governance systems must be designed to produce e-records in the form of digital object(s) along with Preservation Information Documentation (PID) so that it is ready to be accepted for preservation.

6. Guidelines for eGOV-PID metadata capture

- The e-record(s) should be captured as per the Guidelines for Production of Preservable e-Records (PRoPeR).

- Capture the preservation information using the eGOV-PID metadata schema in XML document form. Refer the sample XML with preservation metadata in Annexure B.
- The XML file containing the preservation information should be named as RECORD_IDENTIFIER_PID.XML (The record identifier is the unique accession number of the e-record). This is to help in distinguishing between the e-record and its preservation metadata.
- The preservation information in XML format should be stored along with the e-record in the same folder.
- XMLs, PDFs, Images, etc other than the main e-record should be named using its unique identifier with appropriate suffix separated by underscore character.

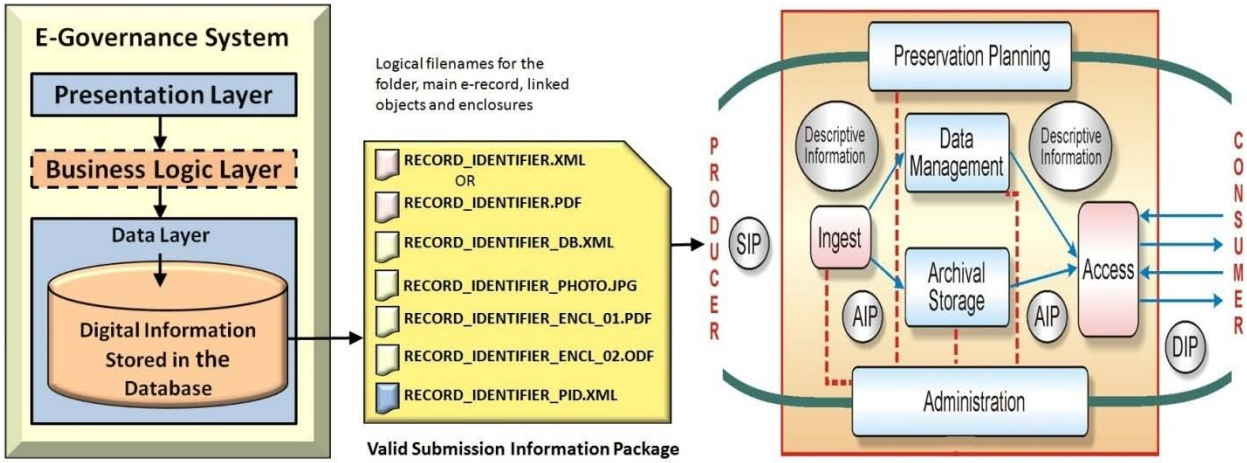


Figure 2: Production of valid Submission Information Package (SIP) for OAIS

As shown in figure 2, the e-government system or e-records creation system should be designed to enable capturing of e-records that need to be preserved for long durations.

7. Principles and overview of preservation metadata

This section presents the principles used for defining the various categories of metadata within the eGOV-PID Metadata Schema. The principles are evolved on the basis of minimum requirements of Registration Metadata as specified in the ISO/TR 15489- 2 Information and Documentation - Records Management guidelines. In the metadata schema, the cataloging, provenance, representation and fixity are the mandatory sections; and enclosures, authenticity and access rights are the optional sections (to be used if applicable).

7.1. Cataloging Information

The Paris Principles for Cataloging are adopted for defining the common cataloging parameters for electronic records (International Conference of Cataloging principles 1961). The Paris Principles primarily focus on how to find a single resource (e-record) and how to find sets of resources (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 cataloging parameters for e-records provide adequate access points for classification and retrieving the bibliographic data. The cataloging parameters are mandatory to be filled for the purpose of archival and access. This section incorporates basic elements defined by Dublin Core Metadata Initiative (DCMI).

Cataloging Information		
Label	**Definition**	**Obligation**
RecordIdentifier	An unambiguous reference to the e-record. It is the unique alphanumeric number assigned to the e-record.	Mandatory
Title	Name of the document.	Optional
Subject	The specific theme of the document or e-record.	Optional
Languages	State Recognized Official Language Code to be mentioned for describing the languages used in the e-record.	Mandatory
Type	The meta level classification of e-record in terms of whether it is a document, financial, human resource, legal, property, etc.	Optional
MainCategory	The higher level classification of the e-record.	Optional

SubCategory	The secondary level classification of the e-record.	Optional
DateTime	The official date and time on which the e-record got completed.	Mandatory
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	Optional
NameID	The name(s) of persons associated with the e-record e.g. name of property owner in case of property registration document. Mention the unique number associated with the person(s) such as UID No., PAN number, Election ID No., Passport No. Registration No. License No. or other ID numbers as applicable.	Optional
RecordProducer	The name of the records creating agency or the organization which produced the final e-record.	Mandatory
Owner	The name of the owner of e-record or the copyright associated with it.	Optional
Context	The background information that helps in knowing the circumstances in which the e-record is created.	Optional
Validity	A limited period for which the information in the e-record is applicable.	Optional
Retention	The duration for which the e-record must be preserved and the disposal action if necessary.	Mandatory
Relation	A related resource which defines the type of relation in terms of - - Renewal - Reference	Optional
Description	The explanation, comments or remarks or any other observations which the record producer may wish to write about the e-record.	Optional

7.2. Enclosures information

The final e-record is generated on the basis of various documents and digital objects which are sometimes linked with it. They are also helpful in establishing the context in which the e-record was produced. The accuracy of the final e-record can be verified and validated on the basis of the enclosed documents or digital objects.

Enclosure Information (if applicable)		
Label	**Definition**	**Obligation**
SerialNumber	Serial number of linked digital object or document	Mandatory
Title	The title of linked digital object or document.	Mandatory
MIMEType	The type of content.	Mandatory
FileName	Name of the file.	Mandatory

7.3. Provenance information

The provenance defines the information that documents the history of the e-record. This information tells the origin or source of the e-record, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. The archive is responsible for creating and preserving provenance information from the point of ingest; however, earlier provenance information should be provided by the record producer. Provenance information adds to the evidence to support authenticity.

Provenance Information		
Label	**Definition**	**Obligation**
Origin	The origin of e-record is documented in terms of addresses of the organization and the device which produced it.	Mandatory
Migration	It contains the relative path of an XML documenting the process of migrating the e-record from its original file format into another file format.	Optional

7.4. Representation information

Representation information (technological details) allows for the full interpretation of the data into meaningful information and can be helpful in reading the e-record in future.

Representation Information		
Label	**Definition**	**Obligation**
SoftwareList	The list of software(s) used for creating the e-record.	Mandatory
HardwareSpecification	The specifications of the hardware used for creating the e-record.	Mandatory

7.5. Fixity information

Fixity information provides the data integrity checks or validation/verification keys used to ensure that the particular e-record has not been altered in an undocumented manner.

Fixity Information		
Label	**Definition**	**Obligation**
Checksum	A checksum or hash sum is a fixed-size datum computed from an arbitrary block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage.	Mandatory

7.6. Digital signature information

The digital signature metadata needs to be captured so as to establish the authenticity of the e-record at a later date.

Digital Signature Information (if applicable)		
Label	**Definition**	**Obligation**
Signer	The name of person / authority who has signed the e-record.	Mandatory
SigningTime	Timestamp details of when the e-record is signed.	Optional

Reason	The context or purpose of digital certificate.	Optional
Location	The place where the e-record is signed.	Optional
MessageDigest	Names of hash algorithms and checksums.	Mandatory
PublicKey	It contains the names of encryption algorithms, key strength, public key value and message digest of the public key.	Mandatory
Signature	It contains the certificate data, version, serial number, validity and hash algorithms.	Mandatory
Issuer	Information regarding the CA that issued the certificate in terms of name, e-mail, location, state, organization unit, organization, country of the issuer.	Mandatory

7.7. Access rights information

Access rights information identifies the access restrictions pertaining to the e-record.

Access Rights Information (if applicable)		
Label	**Definition**	**Obligation**
RecordOfficer	Name and contact details of the record officer from the record producing organization.	Mandatory
Disclosure	Permission(s) to disclose the e-record in terms of "public" or "private". If it is a private e-record then confidentiality associated with the e-record is defined in terms of "secrete" or "top secrete".	Mandatory

8. Metadata dictionary and schema

The eGOV-PID metadata dictionary and schema are presented in this chapter.

8.1. Overview of schema definition for eGOV-PID

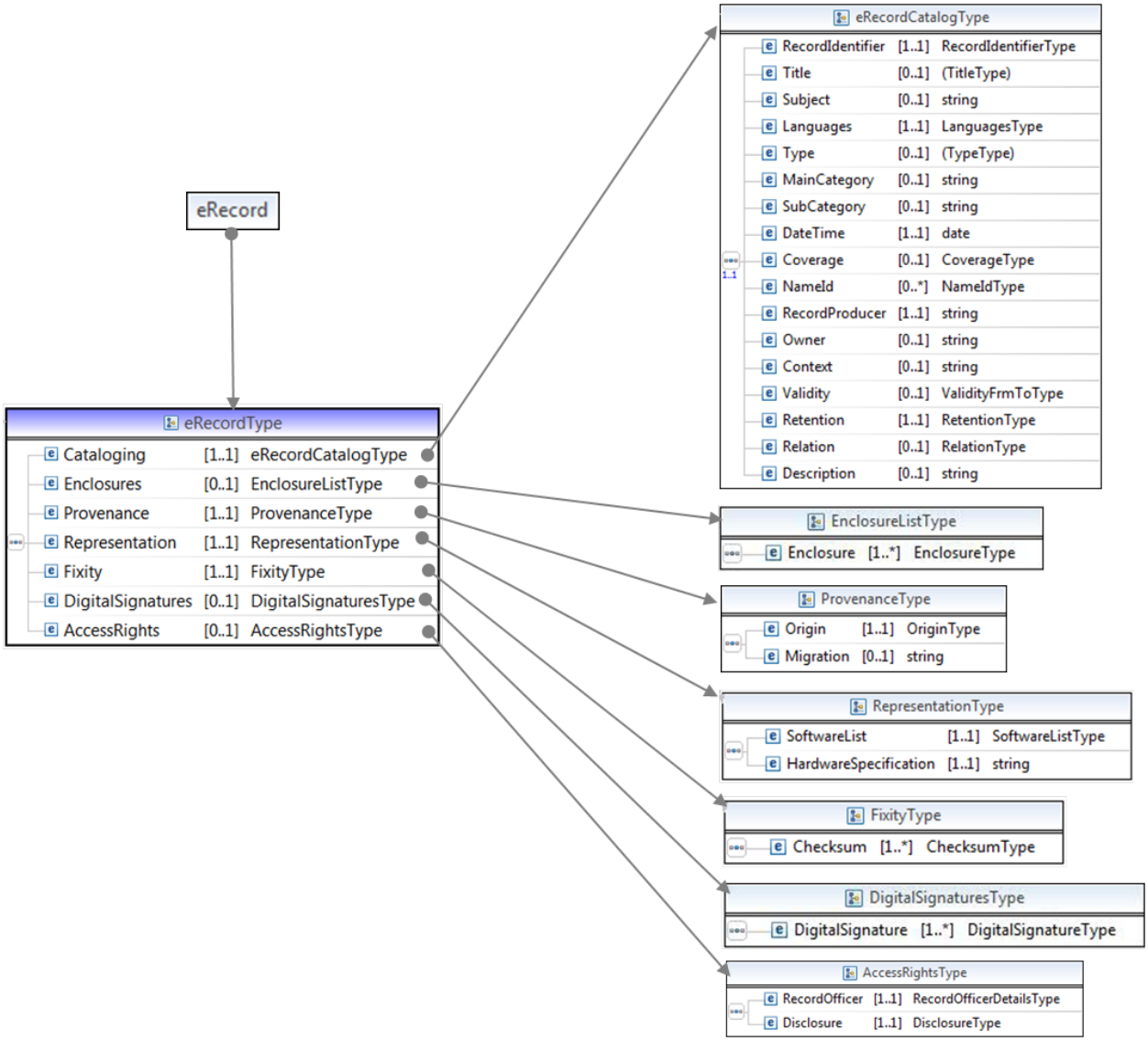


Figure 3: eGOV-PID XSD

Semantic Unit	**eRecord**
Semantic Components	**1. Cataloging** **2. Enclosures** **3. Provenance** **4. Representation** **5. Fixity** **6. DigitalSignatures** **7. AccessRights**
Definition	The e-record that needs to be preserved. It holds the different sections of metadata.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

8.1.1. Schema definition for cataloging information

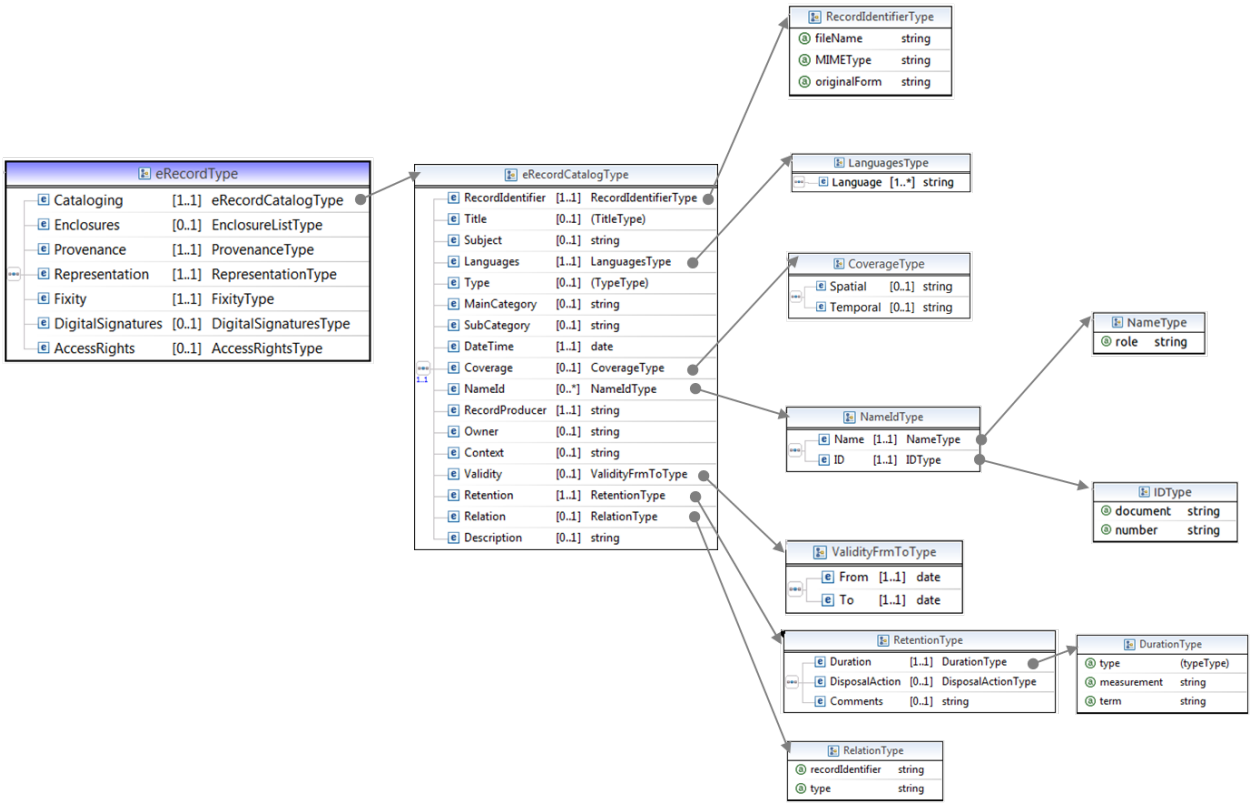


Figure 4. XSD overview of cataloging information

Semantic Unit	**1. Cataloging**
Semantic Components	**1.1 RecordIdentifier** **1.2 Title** **1.3 Subject** **1.4 Languages** **1.5 Type** **1.6 MainCategory** **1.7 SubCategory** **1.8 DateTime** **1.9 Coverage** **1.10 NameId** **1.11 RecordProducer** **1.12 Owner** **1.13 Context** **1.14 Validity** **1.15 Retention** **1.16 Relation** **1.17 Description**
Definition	It is a container to include semantic units defined external to e-record.

Rationale	It serves as the handle to an e-record. The basic information can be used for the purpose of identification, searching and sorting.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**1.1 RecordIdentifier**		
Semantic Components	None		
Definition	It provides an unambiguous reference to the e-record. It is a unique reference number or accession number by which the e-record is catalogued and identified.		
Rationale	The record identifier helps in searching and accessing the e-record. The record identifiers are also used in establishing relationships between two or more records.		
Data Constraint	String		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Example**
fileName	File name along with file extension of main e-record to be preserved.	Mandatory	record_identifier.pdf
MIMEType	Type of main e-record to be preserved.	Mandatory	text/xml image/jpeg application/pdf
originalForm	The form of the record when it is produced.	Mandatory	Born digital Reformatted digital
Usage notes	The unique record identifier is defined using a combination of alpha-numeric values, separators between the meaningful blocks and with optimal number of characters. Refer the guidelines for defining Unique Record Identifier provided in Best Practices and Guidelines for Production of Preservable e-Records (PRoPeR).		

Semantic Unit	**1.2 Title**
Semantic Components	None
Definition	A name given to the e-record.
Rationale	A human readable name by which the e-record is known.
Data Constraint	String
Examples	Property document, contract, officer order, user manual or booklet
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.3 Subject**
Semantic Components	None
Definition	The topic of e-record.

Rationale	Brief description of e-record given in 3 to 5 words.
Data Constraint	String
Examples	Any official letter has its subject stated in it.
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.4 Languages**
Semantic Components	**1.4.1 Language**
Definition	A list of languages used in the e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**1.4.1 Language**	
Semantic Components	None	
Definition	Directory of State Recognized Official Language Code defined by Office of Registrar General of India (ORGI) to be used for referring the languages used in the e-record. Refer G00.05-01, Data Element - Language Code, Metadata and Data Standards – Demographic, Version 1.1, November 2011, published by Department of Electronics and Information Technology, Government of India.	
Data Constraint	Official language code	
Repeatability	Repeatable	
Obligation	Mandatory	
Usage Notes	Refer the examples given below.	
	Recognized Official Language Code	**Language Name**
	6	Hindi
	21	Telugu
	13	Marathi
	99	Other Language (English)

Semantic Unit	**1.5 Type**
Semantic Components	None
Definition	Broad genre or nature of the e-record.
Rationale	To help in higher level classification of e-records and for performing operations like sorting, narrowing the scope of search, etc.
Data Constraint	String
Examples	Document, Financial, Human Resource, License, Permission, Contract, Property, Legal, etc. e-Records creating agencies can define different e-record types as applicable in their respective domains.

Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.6 MainCategory**
Semantic Components	None
Definition	The higher level classification of e-record.
Rationale	To help in grouping or classifying the e-records belonging to a particular type.
Data Constraint	String
Examples	If “Property” is defined as a higher level record type then “Immovable” or “Movable” could be the main category. e-Records creating agencies can define the main categories as applicable in their respective domain.
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.7 SubCategory**
Semantic Components	None
Definition	Secondary level classification of e-records based on some common attributes.
Rationale	To help in grouping or classifying the e-records belonging to a particular type.
Data Constraint	String
Example	If “Immovable” is defined as the main category then “Property Registration” is a sub-category. e-Records creating agencies can define the sub-categories as applicable in their respective domain.
Repeatability	Not repeatable
Obligation	Optional
Usage notes	It is applicable only if the main category is mentioned.

Semantic Unit	**1.8 DateTime**
Semantic Components	None
Definition	The official date and time on which the e-record got completed.
Rationale	For calculation of validity and retention duration.
Data Constraint	Date in (dd/mm/yyyy hh:mm:ss) format
Repeatability	Not repeatable
Obligation	Date is mandatory.
Example	07/08/2013 16:15:59

Semantic Unit	**1.9 Coverage**
Semantic Components	**1.9.1 Spatial** **1.9.2 Temporal**
Definition	The spatial or temporal topic of e-record / resource, the spatial applicability of the e-record / resource, or the jurisdiction under which the e-record / resource is relevant.
Rationale	It helps to know the span of duration or geographical region associated with e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.9.1 Spatial**
Semantic Components	None
Definition	Spatial characteristics of e-record.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional
Usage notes	Comma separated values in terms of height, width, length or x, y, z coordinates in corresponding units or name of cities, districts defining a region.

Semantic Unit	**1.9.2 Temporal**
Semantic Components	None
Definition	Temporal characteristics of e-record.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional
Usage notes	Period definition

Semantic Unit	**1.10 NameID**
Semantic Components	**1.10.1 Name** **1.10.2 ID**
Definition	The name(s) of persons associated with the e-record e.g. name of property owner in case of property registration document.
Rationale	The names of persons and IDs associated with e-record.
Data Constraint	Container
Repeatability	Repeatable
Obligation	Optional

Semantic Unit	**1.10.1 Name**
Semantic Components	None

Definition	The name of the person associated with e-record.		
Data Constraint	String		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
role	The role of person(s) associated with e-record.	Optional	Citizen, passport officer, judge, petitioner, respondent, licenser, buyer, etc.
Usage notes	If the value of the attribute *role* is not given, it will take *citizen* as default value.		

Semantic Unit	**1.10.2 ID**		
Semantic Components	None		
Definition	The details of the type of ID proof and associated number.		
Rationale	To authenticate the identity of the person or organization and to separate the persons with same names.		
Data Constraint	None		
Repeatability	Not repeatable		
Obligation	Optional		
Attributes	**Definition**	**Obligation**	**Examples**
document	Name of identity document declared by the user.	Mandatory	UID, PAN Card, Employee ID Card, Passport, Registration Card
number	The number provided in the ID document.	Mandatory	

Semantic Unit	**1.11 RecordProducer**
Semantic Components	None
Definition	The name of e-record creating agency or the organization which produced the final e-record.
Rationale	It mentions the name of organization which produced the final e-record.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**1.12 Owner**
Semantic Components	None
Definition	Name of the owner of e-record or copyright or intellectual property.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	1.13 Context
Semantic Components	None
Definition	The background information which helps in knowing the circumstances in which the e-record is created.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	1.14 Validity
Semantic Components	1.14.1 From 1.14.2 To
Definition	A limited period for which the e-record is consider to be valid.
Rationale	It helps in know the validity period of e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	1.14.1 From
Semantic Components	None
Definition	Start date of validity period.
Data Constraint	Date in (dd/mm/yyyy) format.
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	1.14.2 To
Semantic Components	None
Definition	End date of validity period.
Data Constraint	Date in (dd/mm/yyyy) format.
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	1.15 Retention
Semantic Components	1.15.1 Duration 1.15.2 Disposal Action 1.15.3 Comments
Definition	The retention and disposition requirements for the given e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**1.15.1 Duration**		
Semantic Components	None		
Definition	The duration for which the e-record is required to be retained.		
Data Constraint	Container		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attribute	**Definition**	**Obligation**	**Example**
type	The retention duration can be declared in terms of – - Permanent - Period	Mandatory	Period
term	Positive Number	Optional	25
measurement	Measurement Unit	Optional	Years
Usage notes	If the retention type is “Permanent” then the duration need not be mentioned. The value "Permanent" maps with the **Category I** type of records as defined in the Central Secretariat Manual of e-Office Procedure (e-Manual) by DARPG. If the retention type is “Period” then the duration needs to be mentioned in terms of number of years. The value "Period" maps with the **Category II** type of records which have to be retained for a limited duration as defined in the Central Secretariat Manual of e-Office Procedure (e-Manual) by DARPG. If the retention type is “Period” then the retention guideline should be mentioned as comments (refer 1.15.3).		

Semantic Unit	**1.15.2 DisposalAction**
Semantic Components	None
Definition	The guidance in terms of whether the e-record is to be reviewed for extension of retention period or disposed after the stipulated retention period is over.
Data Constraint	Controlled vocabulary
Repeatability	Not repeatable
Obligation	Optional
Usage notes	The values should be either of the following - - Review - Dispose

Semantic Unit	**1.15.3 Comments**
Semantic Components	None
Definition	The e-record retention guidelines or rules should be mentioned as comments.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**1.16 Relation**		
Semantic Components	None		
Definition	A related resource which defines the type of relation in terms of renewal, reference, etc.		
Data Constraint	None		
Repeatability	Repeatable		
Obligation	Optional		
Attributes	**Definition**	**Obligation**	**Examples**
recordIdentifier	It is the reference or record identifier.	Mandatory	
type	The type of relation between the records.	Mandatory	Basis, Reference, Renewal, Other

Semantic Unit	**1.17 Description**
Semantic Components	None
Definition	Supplementary information related with e-record.
Rationale	It can include the comments, reasons and other useful information related to e-record which is not captured through other parameters.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

8.1.2. Schema definition for enclosure information

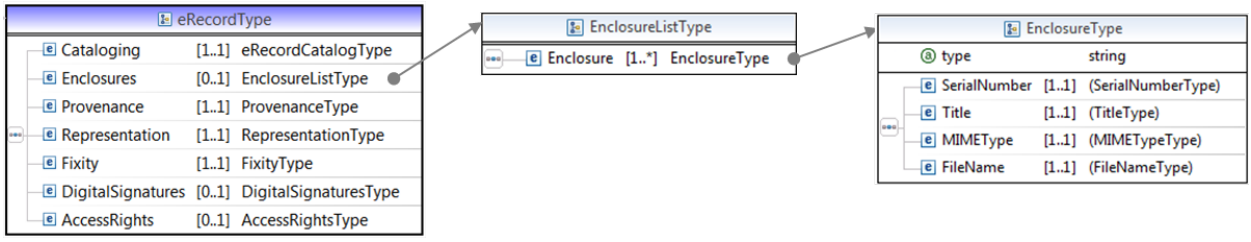


Figure 5. XSD Overview of enclosure information

Semantic Unit	**2 Enclosures**
Semantic Components	**2.1 Enclosure**
Definition	It provides a list of supplementary documents, images, digital objects linked with the main e-record.
Rationale	It helps in establishing the context and authenticity of the main e-record for verification purpose.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**2.1 Enclosure**		
Semantic Components	**2.1.1 SerialNumber** **2.1.2 Title** **2.1.3 MIMEType** **2.1.4 FileName**		
Definition	It includes the details of enclosure.		
Data Constraint	Container		
Repeatability	Repeatable		
Obligation	Mandatory		
Attribute	**Definition**	**Obligation**	**Example**
type	The domain specific metadata not covered in this dictionary can be linked as a separate XML.	Optional	OtherDescriptiveMetadata
Usage Notes	The attribute 'type' will appear only in case of OtherDescriptiveMetadata XML files.		

Semantic Unit	**2.1.1 SerialNumber**
Semantic Components	None
Definition	It is the sequence number of the attached document.
Rationale	To know how many enclosures are attached.
Data Constraint	Number
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**2.1.2 Title**
Semantic Components	None
Definition	Title of the attached document or digital object.
Rationale	Helps in knowing the subject of enclosure.
Data Constraint	String
Examples	- ID Card - Address Proof
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**2.1.3 MIMEType**
Semantic Components	None
Definition	MIME types form a standard way of classifying file types.
Rationale	It helps in knowing the file format of the enclosure.
Data Constraint	String
Examples	- image/jpg - image/png - application/pdf
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**2.1.4 FileName**
Semantic Components	None
Definition	Name of the enclosure file along with its relative path.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

8.1.3. Schema definition for provenance information

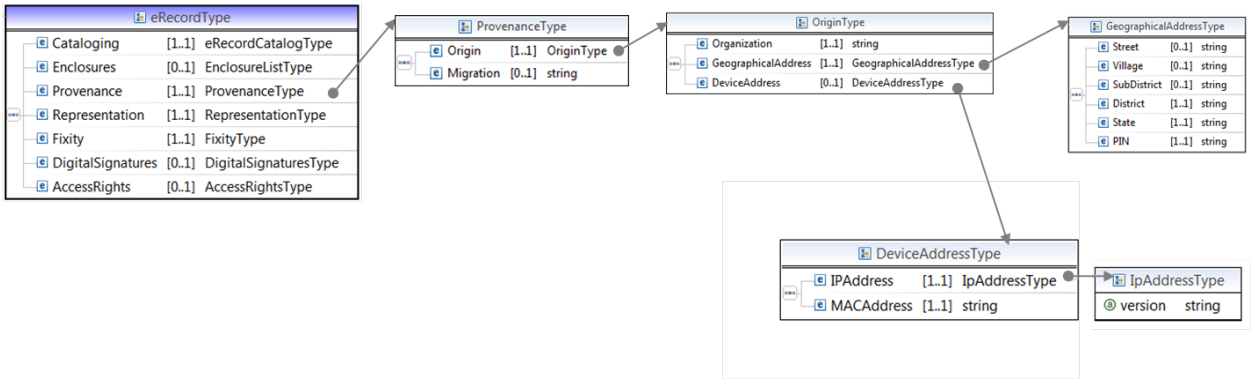


Figure 6: XSD overview of provenance information

Semantic Unit	**3. Provenance**
Semantic Components	**3.1 Origin** **3.2 Migration**
Definition	It describes the origin or the source of e-record.
Rationale	It helps in authenticating the e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**3.1 Origin**
Semantic Components	**3.1.1 Organization** **3.1.2 GeographicalAddress** **3.1.3 DeviceAddress**
Definition	It describes the source of e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**3.1.1 Organization**
Semantic Components	None
Definition	The name of e-records creating agency which created the final e-record.
Rationale	Helps to know the name of organization which produced the e-record.
Data Constraint	String
Examples	- Name of e-district - Name of e-records creating agency
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	3.1.2 GeographicalAddress
Semantic Components	3.1.2.1 Street 3.1.2.2 Village 3.1.2.3 SubDistrict 3.1.2.4 District 3.1.2.5 State 3.1.2.6 PIN
Definition	It provides the postal address of the e-records creating agency.
Rationale	It allows one to contact the concerned.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	3.1.2.1 Street
Semantic Components	None
Definition	Street information
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	3.1.2.2 Village
Semantic Components	None
Definition	Name of village
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	3.1.2.3 SubDistrict
Semantic Components	None
Definition	Name of Taluk or Tehsil or Town, etc
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	3.1.2.4 District
Semantic Components	None
Definition	Name of district
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	3.1.2.5 State
Semantic Components	None

Definition	The code or name as per the Office of Registrar General of India for the Indian state or union territory where the e-record is produced.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**3.1.2.6 PIN**
Semantic Components	None
Definition	Postal Index Number (PIN) code is the post office numbering or post code system used by the Indian Postal Service.
Data Constraint	Number
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**3.1.3 DeviceAddress**
Semantic Components	**3.1.3.1 IPAddress** **3.1.3.2 MACAddress**
Definition	The identification details of the machine which produced the e-record.
Rationale	It helps in tracing the source of e-record for authentication purpose as required in IT Act 2000.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**3.1.3.1 IPAddress**		
Semantic Components	None		
Definition	Internet Protocol address assigned to the machine.		
Rationale	IP address of the machine or device at the time (date and time) of finalizing the e-record helps in tracing back the origin of the e-record over network.		
Data Constraint	String		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attribute	**Definition**	**Obligation**	**Example**
version	Version of IP Address	Mandatory	V4 or V6

Semantic Unit	**3.1.3.2 MACAddress**
Semantic Components	None
Definition	A Media Access Control address (MAC address) is a unique identifier assigned to network interfaces for communications on the physical network segment.

Rationale	Helps in tracing back the origin of the e-record.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**3.2 Migration**
Semantic Components	None
Definition	It contains the relative path of an XML documenting the process of migrating the e-record from its original file format into another file format.
Rationale	It helps in authenticating the source of the digital information contained in the migrated e-record.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

8.1.4. Schema definition for representation information

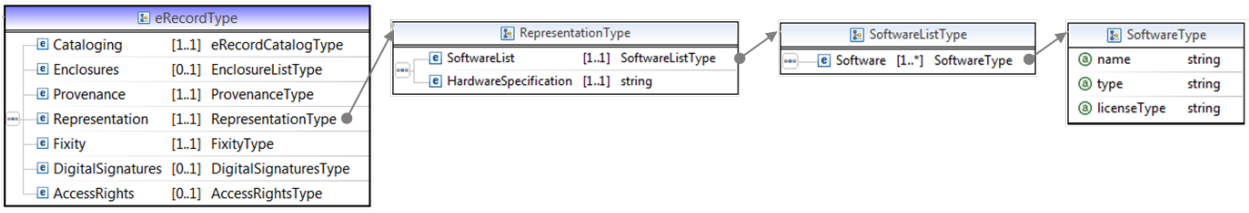


Figure 7: XSD Overview of representation information

Semantic Unit	**4. Representation**
Semantic Components	**4.1 SoftwareList** **4.2 HardwareSpecification**
Definition	Representation Information allows for the full interpretation of the data into meaningful information and can be helpful in reading the e-record in future.
Rationale	Representation information helps to identify the software packages, operating system platforms or hardware specifications which are needed to read, render and interpret the e-record in its original form.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**4.1 SoftwareList**
Semantic Components	**4.1.1 Software**
Definition	A list of software(s) used for creating the e-record.
Rationale	Helps to identify the details of software used for creating the e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**4.1.1 Software**		
Semantic Components	None		
Definition	The name, type and license type of software used for creating the e-record.		
Rationale	Helps to know the software environment necessary for viewing the e-record in future.		
Data Constraint	None		
Repeatability	Repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
name	The name and	Mandatory	Windows 7, Apache FOP 1.1,

	version of software.		etc.
type	The type of software.	Mandatory	Creator, reader, server, database, operating system, compiler, API Library, application, tool, web browser, version
licenseType	The terms of using the software.	Mandatory	Open source, General public license, Proprietary, etc.

Semantic Unit	**4.2 HardwareSpecification**
Semantic Components	None
Definition	The hardware specification of the machine which is used to create the e-record.
Rationale	Helps to know the hardware environment necessary for viewing the e-record in future.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory
Usage notes	A detailed specification, exact statement of particulars of motherboard, system model, system type, processor, memory, display device specification, etc.

8.1.5. Schema definition for fixity information

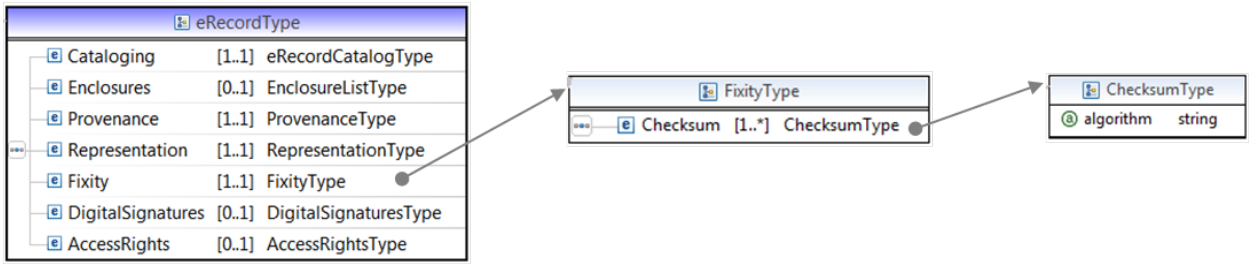


Figure 8: XSD Overview of fixity information

Semantic Unit	**5. Fixity**
Semantic Components	**5.1 Checksum**
Definition	Fixity information provides the data integrity checks or validation/verification keys used to ensure that the particular e-record has not been altered or tampered.
Rationale	It helps in ensuring the integrity of the main e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**5.1 Checksum**		
Semantic Components	None		
Definition	A checksum or hash sum is a fixed-size datum computed from an arbitrary block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage.		
Rationale	The integrity of the data can be checked at any later time by re-computing the checksum and comparing it with the stored one.		
Data Constraint	String		
Repeatability	Repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
algorithm	Name of the digest algorithm for generating checksum	Mandatory	MD5, SHA1

8.1.6. Schema definition for digital signature information

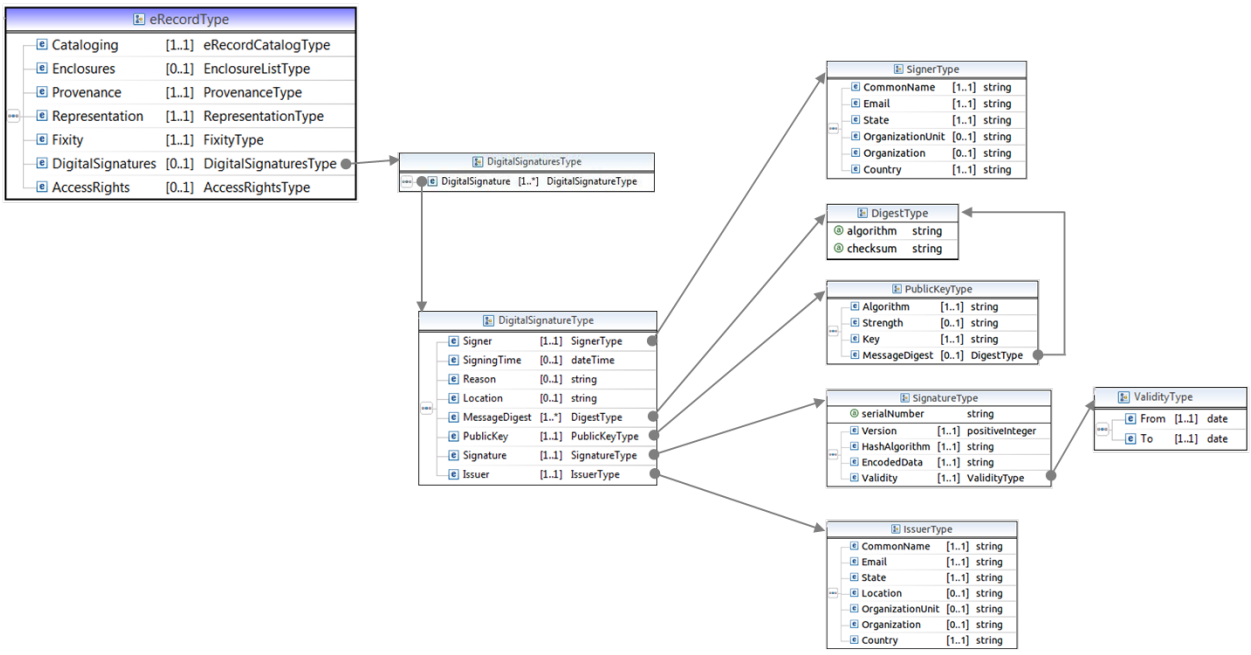


Figure 9. XSD overview of Digital Signature Information

Semantic Unit	**6. DigitalSignatures**
Semantic Components	**6.1 DigitalSignature**
Definition	A digital signature is a mathematical scheme for demonstrating the authenticity of a digital message or document. A valid digital signature gives a recipient reason to believe that the message was created by a known sender, such that the sender cannot deny having sent the message (authentication and non-repudiation) and that the message was not altered in transit.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1 DigitalSignature**
Semantic Components	**6.1.1 Signer** **6.1.2 SigningTime** **6.1.3 Reason** **6.1.4 Location** **6.1.5 MessageDigest** **6.1.6 PublicKey** **6.1.7 Signature** **6.1.8 Issuer**
Definition	It contains the details of digital signature.
Data Constraint	Container

Repeatability	Repeatable
Obligation	Mandatory

Semantic Unit	**6.1.1 Signer**
Semantic Components	**6.1.1.1 CommonName** **6.1.1.2 Email** **6.1.1.3 State** **6.1.1.4 OrganizationUnit** **6.1.1.5 Organization** **6.1.1.6 Country**
Definition	The details about the signer who digitally signed the e-record.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.1.1 CommonName**
Semantic Components	None
Definition	Name of the signer as registered with Certificate Authority (CA).
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.1.2 Email**
Semantic Components	None
Definition	The email ID of the signer as registered with CA.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.1.3 State**
Semantic Components	None
Definition	The state of the signer as registered in CA.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.1.4 OrganizationUnit**
Semantic Components	None
Definition	It refers to the functional department of signer's organization.
Rationale	Organization units model the specific organizational groups inside of an organization.

Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.1.5 Organization**
Semantic Components	None
Definition	Name of the signer's organization.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.1.6 Country**
Semantic Components	None
Definition	Country name of the signer as registered with CA.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.2 SigningTime**
Semantic Components	None
Definition	The time stamp of digital signature attachment to e-record.
Rationale	For calculation of validity and retention duration.
Data Constraint	Standard date format to be followed for date and time.
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.3 Reason**
Semantic Components	None
Definition	The reason for signing.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.4 Location**
Semantic Components	None
Definition	The location name in which the signer signing e-record digitally.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.5 MessageDigest**		
Semantic Components	None		
Definition	The message digest of signed data.		
Data Constraint	None		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
algorithm	Name of the digest algorithm for generating checksum.	Mandatory	MD5, SHA1
checksum	The checksum generated using the algorithm specified.	Mandatory	

Semantic Unit	**6.1.6 PublicKey**
Semantic Components	**6.1.6.1 Algorithm** **6.1.6.2 Strength** **6.1.6.3 Key** **6.1.6.4 MessageDigest**
Definition	It is the public key distributed along with digital signature. It can be used in the verfication process.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.6.1 Algorithm**
Semantic Components	None
Definition	The name of algorithm using which the public key is generated.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.6.2 Strength**
Semantic Components	None
Definition	Bit strength of public key algorithm e.g. 1024 bits
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.6.3 Key**
Semantic Components	None
Definition	The public key
Data Constraint	String

Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.6.4 MessageDigest**
Refer 6.1.5 for use at this location.	

Semantic Unit	**6.1.7 Signature**	
Semantic Components	**6.1.7.1 Version** **6.1.7.2 HashAlgorithm** **6.1.7.3 EncodedData** **6.1.7.4 Validity**	
Definition	A digital signature that can be used to authenticate the identity of the sender of a message or the signer of a document, and to ensure that the original content of the message or document is unchanged.	
Data Constraint	Container	
Repeatability	Not repeatable	
Obligation	Mandatory	
Attribute	**Definition**	**Obligation**
serialNumber	The serial number of the signature from issuer's database.	Mandatory

Semantic Unit	**6.1.7.1 Version**
Semantic Components	None
Definition	Represents version number of digital signature.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.7.2 HashAlgorithm**
Semantic Components	None
Definition	The name of the hash algorithm used to generate the digital signature.
Data Constraint	String
Examples	- SHA1 RSA - MD5WithRSAEncryption
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	6.1.7.3 EncodedData
Semantic Components	None
Definition	The encoded value of e-record and digital signature details.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	6.1.7.4 Validity
Semantic Components	6.1.7.4.1 Starts 6.1.7.4.2 Ends
Definition	The validity period of digital signature.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	6.1.7.4.1 Starts
Semantic Components	None
Definition	The official date and time on which the validity of digital signature begins.
Rationale	For calculation of validity of digital signature.
Data Constraint	Standard date format to be followed for date and time.
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	6.1.7.4.2 Ends
Semantic Components	None
Definition	The official date and time on which the validity of digital signature ends.
Rationale	For calculation of validity of digital signature.
Data Constraint	Standard date format to be followed for date and time.
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	6.1.8 Issuer
Semantic Components	6.1.8.1 CommonName 6.1.8.2 Email 6.1.8.3 State 6.1.8.4 Location 6.1.8.5 OrganizationUnit 6.1.8.6 Organization 6.1.8.7 Country
Definition	It provides the details of the person or organization who has issued the digital signature.
Rationale	It provides the details of the person or organization who has

	issued the digital signature.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.8.1 CommonName**
Semantic Components	None
Definition	Name of the issuer i.e. certificate authority.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.8.2 Email**
Semantic Components	None
Definition	Email of the issuer i.e. certificate authority.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.8.3 State**
Semantic Components	None
Definition	The state where certificate authority office or organization is residing.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.8.4 Location**
Semantic Components	None
Definition	Area where certificate authority office or organization is residing.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**6.1.8.5 OrganizationUnit**
Semantic Components	None
Definition	Organization unit refers to the functional department of issuer's organization.
Rationale	Organization units model the specific organizational groups inside of an organization.
Data Constraint	String

Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.8.6 Organization**
Semantic Components	None
Definition	Name of the issuer's organization.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**6.1.8.7 Country**
Semantic Components	None
Definition	Country name of the issuer.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

8.1.7. Schema definition for access rights information

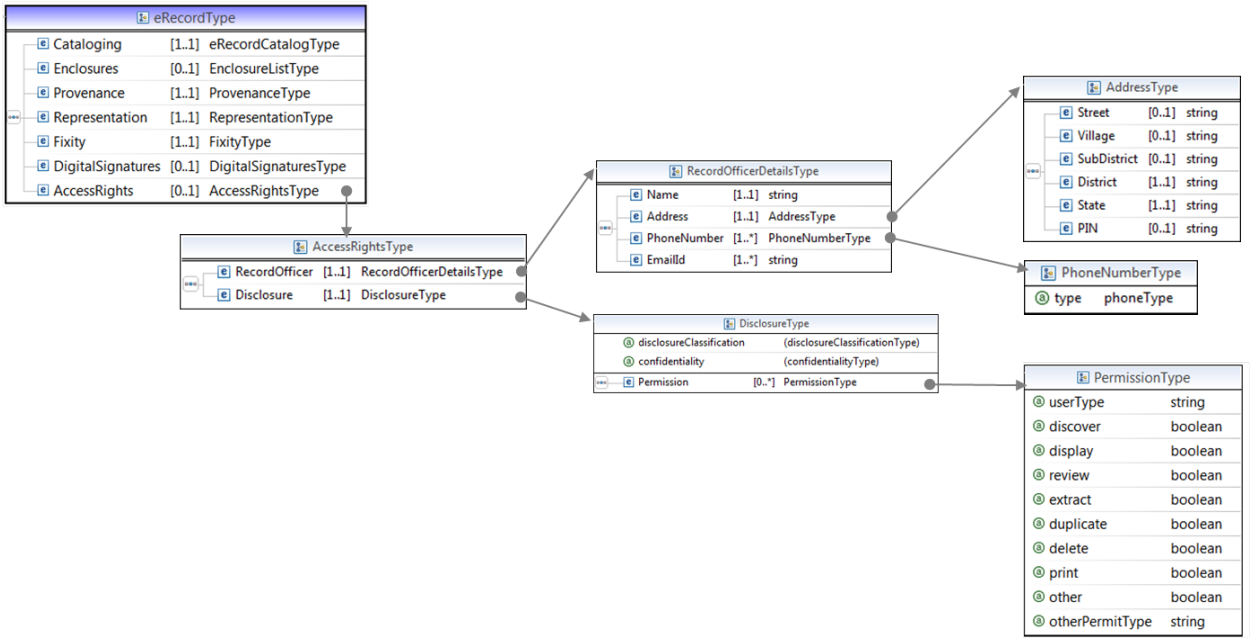


Figure 10.:XSD overview of access rights information

Semantic Unit	**7. AccessRights**
Semantic Components	**7.1 RecordOfficer** **7.2 Disclosure**
Definition	User wise permissions pertaining to access of e-record.
Rationale	Legal compliances
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**7.1 RecordOfficer**
Semantic Components	**7.1.1 Name** **7.1.2 Address** **7.1.3 PhoneNumber** **7.1.4 EmailId**
Definition	The person nominated and responsible for e-records in the records creating agency or organization as required in the Public Records Act.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**7.1.1 Name**
Semantic Components	None

Definition	The name of Record Officer
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**7.1.2 Address**
Semantic Components	**7.1.2.1 Street** **7.1.2.2 Village** **7.1.2.3 SubDistrict** **7.1.2.4 District** **7.1.2.5 State** **7.1.2.6 PIN**
Definition	It provides the official postal address of record officer.
Data Constraint	Container
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	**7.1.2.1 Street**
Semantic Components	None
Definition	Name or number of the street.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**7.1.2.2 Village**
Semantic Components	None
Definition	Name of the village.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**7.1.2.3 SubDistrict**
Semantic Components	None
Definition	Name of Taluk or Tehsil or Town, etc
Data Constraint	String
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	**7.1.2.4 District**
Semantic Components	None
Definition	A division of an area for administrative purposes.
Data constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	7.1.2.5 State
Semantic Components	None
Definition	The code or name as per the Office of Registrar General of India for the Indian state or union territory where the e-record is produced.
Data Constraint	String
Repeatability	Not repeatable
Obligation	Mandatory

Semantic Unit	7.1.2.6 PIN
Semantic Components	None
Definition	Postal Index Number (PIN) code is the post office numbering or post code system used by the Indian Postal Service.
Rationale	It helps to identify a place or location in India.
Data Constraint	Number
Repeatability	Not repeatable
Obligation	Optional

Semantic Unit	7.1.3 PhoneNumber		
Semantic Components	None		
Definition	The phone number of record officer.		
Rationale	It is to help in contacting the record officer for record review.		
Data Constraint	Number		
Repeatability	Repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
type	Mode of communication	Mandatory	Landline or Fax or Mobile

Semantic Unit	7.1.4 EmailId
Semantic Components	None
Definition	The official e-mail address of Record Officer.
Rationale	It is to help in contacting the record officer for record review.
Data Constraint	String
Repeatability	Repeatable
Obligation	Mandatory

Semantic Unit	**7.2 Disclosure**		
Semantic Components	**7.2.1 Permissions**		
Definition	Permission(s) to disclose the e-record.		
Rationale	Legal compliances		
Data Constraint	Container		
Repeatability	Not repeatable		
Obligation	Mandatory		
Attributes	**Definition**	**Obligation**	**Examples**
disclosureClassification	The classification of disclosure in terms of "public" or "private" e-record.	Mandatory	Private
confidentiality	The degree of secrecy associated with the e-record which is defined as: - Secrete - Top Secrete	Optional	Secrete
Usage Notes	If the disclosure classification of e-record is public then confidentiality attribute is not applicable. If the disclosure classification of e-record is "private" then the confidentiality attribute can be applicable.		

Semantic Unit	**7.2.1 Permissions**		
Semantic Components	None		
Definition	It describes various user categories and the authorizations for access to the given e-record.		
Rationale	Enables in protecting the access concerns of the owners of e-records.		
Data Constraint	None		
Repeatability	Repeatable		
Obligation	Optional		
Attributes	**Definition**	**Obligation**	**Examples**
userType	Broad category of user is described by means of userType.	Mandatory	individual
discover	The e-record is available for searching or other discovery.	Mandatory	true or false
display	Viewing, rendering, playing, and executing an e-record.	Mandatory	true or false
review	This permission allows the user to review the e-record.	Mandatory	true or false
extract	Extract a portion of information from the e-record for reuse.	Mandatory	true or false

duplicate	Make an exact copy of e-record for repository management purposes.	Mandatory	true or false
delete	Remove or destroy the e-record from repository.	Mandatory	true or false
print	Rendering the resource onto paper or hard copy.	Mandatory	true or false
other	This allows the user to add custom permissions.	Mandatory	true or false
otherPermitType	Specifies the custom permissions.	Optional	modify
Usage Notes	If the other permission type is true then otherPermitType should be mentioned.		

9. Summary of best practices and guidelines

1.	Design the e-government system or e-records creation system to enable capturing of e-records that need to be preserved for long durations.
2.	Capture the e-record(s) as per the Guidelines for Production of Preservable e-Records (PRoPeR).
3.	Capture the preservation information using the eGOV-PID metadata schema in XML document form.
4.	The XML file containing the preservation information should be named as RECORD_IDENTIFIER_PID.XML (The unique record identifier is the accession number of the e-record). This is to help in distinguishing between the e-record and its preservation metadata.
5.	The preservation information (metadata) in XML format should be stored along with the e-record in the same folder.
6.	XMLs and PDFs other than the main e-record should be named using its unique identifier with appropriate suffix separated by underscore character.
7.	The organizations must define access rules / policy for e-records, as same is reflected in the section on Access Rights Information of eGOV-PID schema.

10. References

- Information Technology Act, 2000, Government of India
- Information Technology Act Amendment (ITAA) 2008, Standing Committee Recommendations, Government of India
- IT Act Notifications GSR 582, 6th September, 2004, Published by Ministry of Communications and Information Technology, Government of India
- Public Records Act, 1993, Government of India
- ISO/TR 15489-1 and 2 Information and Documentation - Records Management, 2001
- ISO 14721:2012 Open Archival Information System (OAIS) Reference Model
- ISO 15836:2009 Information and documentation -- The Dublin Core metadata elements
- Extensible Markup Language (XML), World Wide Web Consortium (W3C)
- ISO 16363: 2012 Audit & Certification of Trustworthy Digital Repositories
- Metadata and Data Standards – Demographic, Version 1.1, published by Ministry of Communications and Information Technology, Department of Electronics and Information Technology, Government of India, November 2011
- ISO 639-3 for Language Codes
- Paris Cataloging Principles, The International Conference on Cataloguing Principles, Paris, 1961
- PREMIS Data Dictionary for Preservation Metadata, Version 2.0, 2008
- Metadata Encoding and Transmission Standard (METS), Library of Congress, 2010
- Archivi, International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2, Edited by Luciana Duranti and Randy Preston, Published by Padova, Italy, 2008
- InterPARES 2, International Research on Permanent Authentic Records, A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records, 2008
- Electronic Records Management: An Audit Guide, EUROSAI IT Working Group, Version 0.8

Annexure A. Implementation guidelines

Ideally, the e-government system should be designed and developed to enable capturing of e-records and preservation metadata as per the eGOV-PID metadata schema.

However, in case of dealing with legacy e-government systems, the software developers can implement the following steps for generating the Submission Information Packages from the database-

- Analyze the producer database.
- Identify the tables which contain basic cataloging fields such as record identifier, date time, names, address, etc as required in the eGOV-PID.
- Create a database view which consolidates the cataloging fields.
- Map the view fields along with the cataloging elements provided in the eGOV-PID Metadata Schema. It should be ensured that at least all mandatory elements are mapped properly.
- In a similar way, the metadata should be mapped as applicable for other sections of eGOV-PID such as Enclosure Information, Provenance Information, etc.
- Appropriate values for Representation Information and Access Rights Information may be provided externally (if this information is not available in the database).
- Check if the digital signature is stored in the database. If it is so then the metadata pertaining to digital signature can be extracted and mapped into the eGOV-PID metadata schema.
- In case the main e-record is stored in the database then it should be extracted in its original format and stored in the file system. It should be named as per its unique record identifier.
- The fixity information should be calculated and incorporated in the eGOV-PID metadata schema.

The eGOV-PID XSD is readily available at the following URL –

http://www.ndpp.in/digital-preservation-standards

Refer the sample XML with preservation metadata in Annexure B.

Annexure B. Sample XML with preservation metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<ndpp:erecord xmlns:ndpp="http://www.ndpp.in/coe-dp/2013/eRecordSchema_Consolidated" xmlns:xsi=
"http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation=
"http://www.ndpp.in/coe-dp/2013/eRecordSchema_Consolidated eRecordSchema_Consolidated.xsd ">
 <ndpp:Cataloging>
  <ndpp:RecordIdentifier MIMEType="application/pdf" fileName="1610_275_1_2011_477.pdf"
  originalForm="Reformatted digital">1610_275_1_2011_477</ndpp:RecordIdentifier>
 <ndpp:Title>SALE DEED</ndpp:Title>
    <ndpp:Languages>
       <ndpp:Language>99</ndpp:Language>
       <ndpp:Language>6</ndpp:Language>
    </ndpp:Languages>
 <ndpp:Type>Registered Document</ndpp:Type>
    <ndpp:MainCategory>Property</ndpp:MainCategory>
    <ndpp:SubCategory>Sale Deed</ndpp:SubCategory>
    <ndpp:DateTime>2011-02-15+05:30</ndpp:DateTime>
    <ndpp:NameId>
       <ndpp:Name role="Buyer">GULAMMUS SAQLAIN</ndpp:Name>
       <ndpp:ID document="PAN Card" number="ANVPM8418L"/>
    </ndpp:NameId>
    <ndpp:NameId>
       <ndpp:Name role="Seller">SHAIK ABDUI WADOOD</ndpp:Name>
       <ndpp:ID document="Driving License" number="9814/HV/2002"/>
    </ndpp:NameId>
 <ndpp:Owner>GULAMMUS SAQLAIN</ndpp:Owner>
 <ndpp:RecordProducer>Sub-Registrar Office</ndpp:RecordProducer>
 <ndpp:Retention>
  <ndpp:Duration type="Permanent"/>
  <ndpp:DisposalAction/>
 </ndpp:Retention>
</ndpp:Cataloging>
<ndpp:Enclosures>
 <ndpp:Enclosure type="OtherDescriptiveMetadata">
       <ndpp:SerialNumber>1</ndpp:SerialNumber>
       <ndpp:Title>Database Record</ndpp:Title>
       <ndpp:MIMEType>text/xml</ndpp:MIMEType>
       <ndpp:FileName>1610_275_1_2011_477_db.xml</ndpp:FileName>
    </ndpp:Enclosure>
 <ndpp:Enclosure>
       <ndpp:SerialNoOfEnclosure>2</ndpp:SerialNoOfEnclosure>
       <ndpp:TitleOfEnclosure>Property Plan</ndpp:TitleOfEnclosure>
       <ndpp:MIMEType>image/jpeg</ndpp:MIMEType>
       <ndpp:FileName>1610_275_1_2011_477_photo</ndpp:FileName>
    </ndpp:Enclosure>
    <ndpp:Enclosure>
       <ndpp:SerialNoOfEnclosure>3</ndpp:SerialNoOfEnclosure>
       <ndpp:TitleOfEnclosure>Passport</ndpp:TitleOfEnclosure>
       <ndpp:MIMEType>image/jpeg</ndpp:MIMEType>
       <ndpp:FileName>1610_275_1_2011_477-5</ndpp:FileName>
    </ndpp:Enclosure>
    <ndpp:Enclosure>
       <ndpp:SerialNoOfEnclosure>4</ndpp:SerialNoOfEnclosure>
       <ndpp:TitleOfEnclosure>PAN Card</ndpp:TitleOfEnclosure>
       <ndpp:MIMEType>image/jpeg</ndpp:MIMEType>
       <ndpp:FileName>1610_275_1_2011_477-57</ndpp:FileName>
    </ndpp:Enclosure>
</ndpp:Enclosures>
```

Original format, **identifier** and title of e-record

Multilevel classification for **access and retrieval**

The names of individuals, their IDs and roles pertaining to e-record

Name of **e-record producer, ownership**

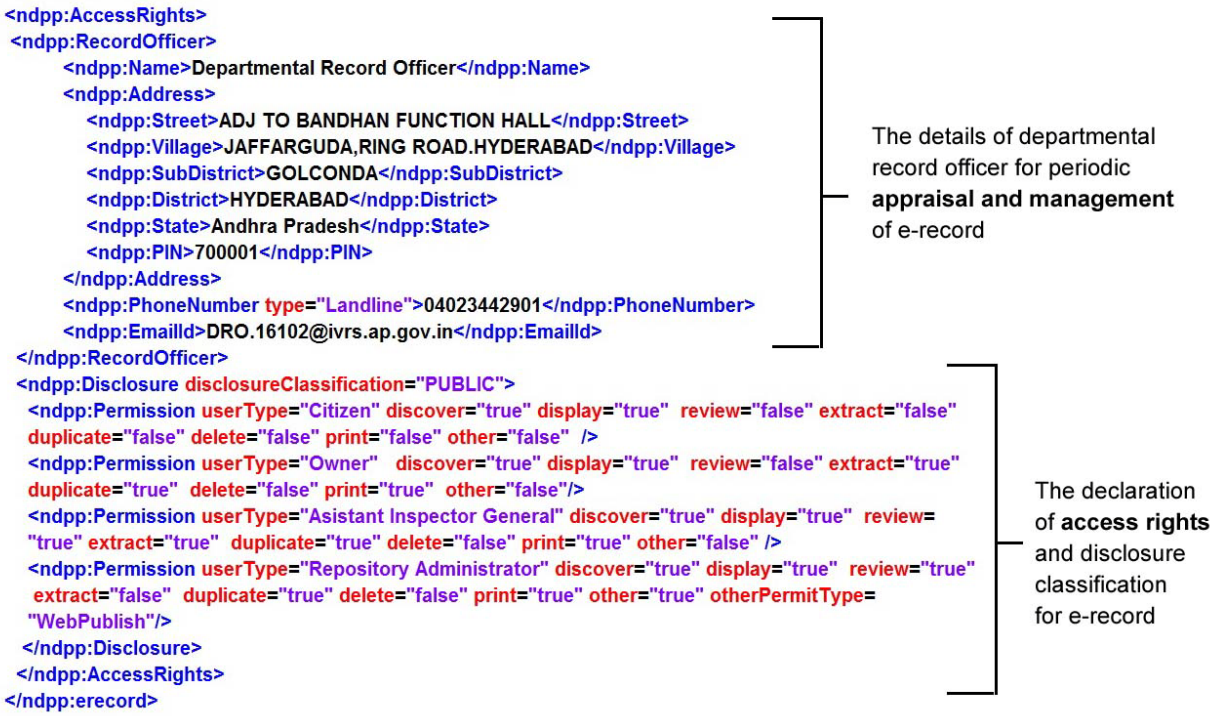
Retention duration of e-record

The **reliability** of e-record can be established on the basis of enclosed documents

(continued on next page)


```
<ndpp:Provenance>
 <ndpp:Origin>
  <ndpp:Organization>Sub-Registrar Office GOLCONDA</ndpp:Organization>
 <ndpp:GeographicalAddress>
      <ndpp:District>HYDERABAD</ndpp:District>
      <ndpp:State>Andhra Pradesh</ndpp:State>
      <ndpp:PIN>700001</ndpp:PIN>
   </ndpp:GeographicalAddress>
 <ndpp:DeviceAddress>
      <ndpp:IPAddress version="V4">10.208.28.94</ndpp:IPAddress>
      <ndpp:MACAddress>35-62-76-F0-F9-45</ndpp:MACAddress>
   </ndpp:DeviceAddress>
 </ndpp:Origin>
 <ndpp:Migration/>
</ndpp:Provenance>
<ndpp:RepInfo>
 <ndpp:SoftwareList>
  <ndpp:Software licenseType="LGPL" name="Appache Fop 1.0" type="Application Software"/>
  <ndpp:Software licenseType="GNU GPL" name="Ubuntu 11.04" type="Operting System"/>
  <ndpp:Software licenseType="GNU Public License 2.0" name="Timmana Regular" type="True Type
  Font"/>
 </ndpp:SoftwareList>
 <ndpp:HardwareSpecification>Intel(R) Core TM i3 CPU,64 bit</ndpp:HardwareSpecification>
</ndpp:RepInfo>
 <ndpp:Fixity>
   <ndpp:Checksum algorithm="MD5" >086d6f77f2faa09382497c8e4f203814</ndpp:Checksum>
   <ndpp:Checksum algorithm="SHA-1" >79d935c885eba5cab21c0d0c3248dc61719bab85
   </ndpp:Checksum>
 </ndpp:Fixity>
<ndpp:DigitalSignatures>
    <ndpp:DigitalSignature>
      <ndpp:Signer>
        <ndpp:CommonName>P.SUBRAMANYA SARMA</ndpp:CommonName>
        <ndpp:Email>pssarma@ap.nic.in</ndpp:Email>
        <ndpp:State>AP</ndpp:State>
        <ndpp:OrganizationUnit>APSC</ndpp:OrganizationUnit>
        <ndpp:Organization>NIC</ndpp:Organization>
        <ndpp:Country>IN</ndpp:Country>
      </ndpp:Signer>
      <ndpp:SigningTime>2011-02-015T18:23:43+05:30</ndpp:SigningTime>
      <ndpp:Reason>CCA Signing </ndpp:Reason>
      <ndpp:Location>GOLCONDA</ndpp:Location>
      <ndpp:MessageDigest algorithm="MD5"
        checksum="95c34ff0ed9b39c1bb15b291fb380cd8" />
      <ndpp:MessageDigest algorithm="SHA-1"
        checksum="d57c151d898fc3722dad85b3931850e0b9f1521" />
      <ndpp:PublicKey>
        <ndpp:Algorithm>RSA</ndpp:Algorithm>
        <ndpp:Strength>1024</ndpp:Strength>
        <ndpp:Key>30 81 9F 30 0D 06 09 2A 86 48 86 F7 0D 01 01 01 05 00 03 81 8D 00 30 81 89 02 81
        81 00 92 55 0D 4B 8F A6 98 05 E8 93 F6 31 EA 60 66 BB 51 CA 9B E4 B2 9F E2 A9 5A 4D 7A 7E
        32 0E 42 CF 32 26 5F 4F BA 94 8D F7 C1 C2 DD F2 B2 4C A0 39 AE 94 6C 49 BB DC B8 7C 16 A7
        F3 E0 58 16 AF 5F 93 13 16 55 02 03 01 00 01 </ndpp:Key>
        <ndpp:MessageDigest algorithm="SHA-1"
          checksum="173e9c72b26f8302cf3246ffe234d5d5a6d1f496" />
      </ndpp:PublicKey>
      <ndpp:Signature serialNumber="07 DB 10 21 0B 10 73 18 72 9F ">
        <ndpp:Version>3</ndpp:Version>
        <ndpp:HashAlgorithm>SHA1withRSA</ndpp:HashAlgorithm>
        <ndpp:EncodedData>30 82 04 E5 30 82 03 CD A0 03 02 01 02 02 0A 07 DB 10 21 0B 10 73 18 72
        9F 30 0D 06 09 2A 86 48 86 F7 0D 01 01 05 05 00 30 81 B0 31 0B 30 09 06 03 55 04 06 13 02 49
        4E 31 24 30 22 06 03 55 04 0A 13 1B 4E 61 74 69 6F 6E 61 6C 20 49 6E 66 6F 72 6D 61 74 69 63
        73 20 43 65 6E 74 72 65 31 0E 30 0C 06 03 55 04 0B 13 05 4E 49 43 43 41 31 21 30 1F 06 03 55 04
        05 44 65 6C 68 69 30 1E 17 0D 31 31 30 32 31 31 30 37 32 34 30 36 5A 17 0D 31 33 30 32 31 30
        E9 FD FD A1 EA 5A 58 08 D6 33 E3 E9 14 09 7A FB B4 35 C3 46 82 EB 84 D9 A0 82 57 7D C6 E7
        </ndpp:EncodedData>
        <ndpp:Validity>
          <ndpp:From>2010-02-11+05:30</ndpp:From>
          <ndpp:To>2012-02-10+05:30</ndpp:To>
        </ndpp:Validity>
      </ndpp:Signature>
      <ndpp:Issuer>
        <ndpp:CommonName>NIC Certifying Authority</ndpp:CommonName>
        <ndpp:Email>support@camail.nic.in</ndpp:Email>
        <ndpp:State>Delhi</ndpp:State>
        <ndpp:Location>New Delhi</ndpp:Location>
        <ndpp:OrganizationUnit>NICCA</ndpp:OrganizationUnit>
        <ndpp:Organization>National Informatics Centre</ndpp:Organization>
        <ndpp:Country>IN</ndpp:Country>
      </ndpp:Issuer>
    </ndpp:DigitalSignature>
</ndpp:DigitalSignatures>
```

(continued on next page)



The XML example of an e-record which is to be kept for 10 years and then reviewed before disposal as under-

```
<ndpp:Retention>
  <ndpp:Duration measurement="year" term="10" type="Period"/>
  <ndpp:Comments>Upgrade to Category I if the e-record is needed for legal purpose beyond 10 years.
  </ndpp:Comments>
  <ndpp:DisposalAction>Review</ndpp:DisposalAction>
</ndpp:Retention>
```

Acknowledgements

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

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

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

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