

PREMIS RECOMMENDATIONS
Based on PREMIS Data Dictionary Version 1.0

OBJECT ENTITY

Objects are discrete units of information in digital form. Objects are what the repository preserves. Examples include representation, file, bitstream. At the representation level, the following are mandatory:

Mandatory Semantic Unit	Repeatable ?	Definition	Example	Usage notes	Recommendation
objectIdentifier (container)	y	Designation used to uniquely identify the object within the preservation repository	n/a		Value is already recorded as OBJID in the METS root element so ignore.
objectIdentifierType	n	Designation of the domain within which the object identifier is unique	<ul style="list-style-type: none"> • DLC • FCLA Digital Archive 	Type of identifier may be implicit within the repository as long as it can be explicitly communicated when the digital object is disseminated outside of it	Implicit so ignore
objectIdentifierValue	n	Value of the objectIdentifier	<ul style="list-style-type: none"> • 000003112 • 1001/dig/pres/2004-024 		Value already recorded as OBJID in the METS root element so ignore.
preservationLevel	n	A value indicating the set of preservation functions expected to be applied to the object	<ul style="list-style-type: none"> • Bit-level • Full • 0 • 1 • 2 • fully supported with future migrations 	If the repository offers only a single preservation level, this does not need to be explicitly recorded within the repository	Include in METS amdSec in techMD mdWrap MDTYPE="PREMIS" <premis:object> <premis:preservationLevel>1</premis:preservationLevel> This should be taken from a controlled vocabulary – we can use same codes as Deep Blue – 1 is highest (we only have one level at this point)
objectCategory	n	The category of object to which the metadata applies	<ul style="list-style-type: none"> • representation • file • bitstream 	Repositories likely to treat different categories of objects differently in terms of metadata and data management functions	Object category = "representation" is implied if metadata is not recorded within <file> element in mets:amdSec – so ignore

At the file level, the following are mandatory:

Semantic unit		Definition	Example	Usage notes	Recommendation
objectIdentifier (container)	y	Designation used to uniquely identify the object within the preservation repository	n/a		n/a
objectIdentifierType	n	Designation of the domain within which the object identifier is unique	<ul style="list-style-type: none"> • DLC • FCLA Digital Archive 	Type of identifier may be implicit within the repository as long as it can be explicitly communicated when the digital object is disseminated outside of it	n/a
objectIdentifierValue	n	Value of the objectIdentifier	<ul style="list-style-type: none"> • 000003112 • 1001/dig/pres/2004-024 		File group plus sequence number already ensures uniqueness at file level - ignore
preservationLevel	n	A value indicating the set of preservation functions expected to be applied to the object	<ul style="list-style-type: none"> • Bit-level • Full • 0 • 1 • 2 • fully supported with future migrations 	If the repository offers only a single preservation level, this does not need to be explicitly recorded within the repository	We currently only offer one level of preservation so ignore for now
objectCategory	n	The category of object to which the metadata applies	<ul style="list-style-type: none"> • representation • file • bitstream 	Repositories likely to treat different categories of objects differently in terms of metadata and data management functions	Object category "file" is implied for metadata contained within the <file> element - ignore
objectCharacteristics (container)	y	Technical properties of a file or bitstream that are applicable to all or most formats	n/a	Components = Mandatory: compositionLevel, format Optional: Fixity (container for messageDigest info) Size significantProperties, inhibitors	n/a

Semantic unit		Definition	Example	Usage notes	Recommendation
compositionLevel	n	Indication of whether the object is subject to one or more processes of decoding or unbundling	<ul style="list-style-type: none"> • 0 • 1 • 2 	compositionLevel of zero indicates that the object is a base object not subject to further encoding, level of 1 or higher indicates that one or more decodings must be applied. Numbering goes lowest to highest – 0 is base, 1-n are subsequent. Use 0 as the default if there is only one compositionLevel.	All our files are 0 so can ignore
fixity (container)	n	Information used to verify whether an object has been altered in an undocumented or unauthorized way	n/a	The act of performing a fixity check and the date it occurred would be recorded as an Event. The result of the check would be recorded as the eventOutcome. Therefore, only the messageDigestAlgorithm and messageDigest need to be recorded as objectCharacteristics for future comparison	n/a
messageDigest Algorithm	y (if include fixity)	Algorithm used to construct the message digest for the digital object	<ul style="list-style-type: none"> • MD5 • Adler-23 • SHA-1 		Included as CHECKSUMTYPE in file ELEMENT in mets fileSec so ignore
messageDigest	y (if include fixity)	Output of the message digest algorithm			Included as CHECKSUM in file ELEMENT in mets fileSec so ignore
messageDigest Originator	n	The agent that created the original message digest that is compared in a fixity check	<ul style="list-style-type: none"> • Google • University of Michigan Digital Conversion Unit • Trigonix 	If a repository ingests files that have had message digests calculated by the submitter, checking ensures that the file s received is the same as the file sent. Or repository may ingest files that do not have message digests and so must calculate the initial value upon ingest.	Recorded as linkingAgentIdentifierValue for premis:eventType = message digest calculation so ignore
format (container)	n	Identification of the format of a file or bitstream where format is the organization of digital information according to preset specifications	n/a		n/a

Semantic unit		Definition	Example	Usage notes	Recommendation
formatDesignation (container)	n	Identification of the format of the object	n/a		n/a
formatName	n	Designation of the format of the file or bitstream	<ul style="list-style-type: none"> • text/sgml • Image/tiff/geotiff 		This is already recorded in MIMETYPE in METS fileSec so ignore
storage (container)	y	Information about how and where a file is stored in the storage system		Normally there would be single storage location and medium. The storage composite should be repeated if there are two or more copies that are identical bit-wise and managed as a unit except for the medium on which they are stored. Must have a single objectIdentifier and be managed as a single object by the repository	n/a
storageMedium	n	The physical medium on which the object is stored (magnetic tape, hard disk, etc)	<ul style="list-style-type: none"> • magnetic tape • hard disk • TSM 	In some cases can be masked from direct repository management by storage management systems but the underlying assumption is that the repository ultimately is in control and needs to manage for technical obsolescence. In some cases may not be the specific medium, but the system that knows the medium (e.g. Tivoli Storage Manager)	All on hard disk so ignore.

EVENT ENTITY

An event is an action that involves or impacts at least one Object or Agent associated with or known by the repository. Recording events helps document digital provenance and can track history of object through chain of events that occur during Object lifecycle. Determining which events are in scope, which should be recorded and at what level of granularity is up to the repository. Note: objectID + eventID = unique event within the repository.

Mandatory Semantic Unit	Repeatable?	Definition	Example	Usage notes	Recommendation
eventIdentifier (container)	n	A designation used to uniquely identify the event within the preservation repository system	n/a	No global scheme/standard; likely to be system generated. Whether or not to record an event depends upon the importance of the event. Actions that modify objects should always be recorded.	Record events in METS amdSec in digiproVMD <mets:digiproVMD ID="(required xml id value., e.g. event1?)"> <mets:mdWrap MDTYPE="PREMIS"> <mets:xmlData> <premis:event> <premis:eventIdentifier>
eventIdentifierType	n	A designation of the domain within which the event identifier is unique	<ul style="list-style-type: none"> FDA Stanford Repository Event ID UUID 	Most repositories will use their own internal numbering system. It can be implicit within the system and provided explicitly only if the data is exported	Not necessary – always SDR so ignore
eventIdentifierValue	n	The value of the eventIdentifier	<ul style="list-style-type: none"> [a binary integer] E-2004-11-13-000119 	Can make up numbering system/codes. Event ID must be unique for the object e001.1 (Unique id within repository is doi for the volume plus event id. Event could be at object level or file level).	Include, but decision on numbering system/codes can be delayed until implementation/left up to the Programmer. <premis:eventIdentifierValue>e001.1</premis:eventIdentifierValue> </premis:eventIdentifier>
eventType	n	A categorization of the nature of the event	<ul style="list-style-type: none"> Ingest 	Value should be taken from a controlled vocabulary. Each repository should define its own controlled vocabulary.	Include - see terms listed in chart below – e.g., <premis:eventType>validation</premis:eventType>
eventDateTime	n	Single date and time or date and time range at or during which the event occurred		Any date/time convention as long as can be translated into ISO8601	Include - e.g. <premis:eventDateTime>2007-10-25T12:45:35</premis:eventDateTime>

Mandatory Semantic Unit	Repeatable?	Definition	Example	Usage notes	Recommendation
eventOutcomeInformation (container)	y	Information about the outcome of an event	n/a		Include ONLY for eventType fixity check
eventOutcomeDetail	n	A non-coded detailed description of the result or product of the event.	<ul style="list-style-type: none"> pass fail 		Include ONLY for eventType fixity check – use pass and fail for possible outcomes
linkingAgentIdentifier (container) ONLY MANDATORY IF APPLICABLE repeatable	y	Designation used to uniquely identify the agent within a preservation repository system	n/a	Include when necessary for any eventType defined in the table in the eventType controlled vocabulary above where possible Agents have been identified (agents can be persons, organizations or software)	Include – all eventTypes we have defined should specify an agent.
linkingAgentIdentifierType	n	Designation of the domain in which the agent identifier is unique	<ul style="list-style-type: none"> LCNAF SAN MARC Organization Codes URI 	Must be taken from a controlled vocabulary –but examples I've looked at people define their own	Include - e.g. <premis:linkingAgentIdentifier> <premis:linkingAgentIdentifierType> SDR</premis:linkingAgentIdentifierType>
linkingAgentIdentifierValue	n	Value of the agentIdentifier	92-79917 Owens, Erik C. MH-CS Info:lccn/n788099 0	May be a unique key or a controlled textual form of name	Include - Extract from Artist tag in XMP metadata e.g. <premis:linkingAgentIdentifierValue> Google</premis:linkingAgentIdentifierValue> </premis:linkingAgentIdentifier>

eventType controlled vocabulary (record all events at the Object level):

Events	Definition	Materials downloaded from GRIN: source of eventDateTime	Materials not downloaded from GRIN: source of eventDateTime	Source of linkingAgentIdentifierValue*
capture	process whereby repository actively obtains object	GRIN date scanned. Note: we will need Google to take date from TIFF header for items they did not scan.	Based on file timestamp but check w/Chris on how this info is embedded in METS	Extract from Artist tag in XMP metadata
compression	process of coding data to save storage space or transmission time	GRIN date converted – use date scanned for Google rejects that we/vendor scan	Same as capture date	Extract from Artist tag in XMP metadata
decryption	process of converting encrypted data to plaintext (better definition?)	GROOVE	n/a	Always UM
fixity check	process of verifying that an object has not been changed in a given period	GROOVE	GROOVE	Always UM
ingestion	process of adding objects to a preservation repository	GROOVE	GROOVE	Always UM
message digest calculation	process by which a message digest (hash) is created	GRIN date converted – use date scanned for Google rejects that we/vendor scan	Same as capture date	Extract from Artist tag in XMP metadata
validation	process of comparing an object with a standard and noting compliance or exceptions	GROOVE	GROOVE	Always UM

***Note: We may not have this information for some legacy DLPS materials**