

National Library of New Zealand

**Metadata Standards Framework –
Metadata Implementation Schema**

July 2003

Preservation Metadata

Table of Contents

1. Executive Summary	2
1.1. Introduction.....	2
1.2. Background.....	2
1.3. The Data Dictionary and a Metadata Repository - Introduction.....	2
1.4. Field Reference Number and Name.....	3
1.5. Structure of Paper.....	4
2. Entity List	5
2.1. Entity 1 - Object.....	5
2.2. Entity 2 - Process.....	10
2.3. Entity 3 - File.....	13
2.4. Entity 4 - Metadata Modification.....	24
Appendices	26
Appendix 1 – Schema Map.....	26
Appendix 2 – Preservation Metadata Model.....	27
Appendix 3 - Logical / Design Data Comparison.....	28
Appendix 4 – Associated Documents.....	33

1. Executive Summary

1.1. Introduction

The National Library of New Zealand (NLNZ) is committed to the establishment of a Digital Archive for the management and preservation of the digital materials that are increasingly being brought into its collections. The Digital Archive will enhance access to the Library's digital resources for all New Zealanders now and in the future.

The incentive to actively pursue this work has increased greatly since the passing of the National Library of New Zealand Act 2003 which mandated the legal deposit of electronic resources to the National Library of New Zealand.¹

1.2. Background

In November 2002 the National Library of New Zealand released the first version of a preservation metadata schema designed to provide a working tool for the collection of preservation metadata to be held against the material held in the Library's digital collections.

In June 2003 a revised edition of the Preservation Metadata Schema was released incorporating some advances in NLNZ thinking based on further experience, for example the new revised version is more closely aligned with NISO.Z39.87

Since the release of the schema the Library has continued work on extending the logical model to this implementable data model and on the development of a tool to automatically extract preservation metadata from digital files, based on available file header information and covering a wide range of file types.

The current document presents a data dictionary model for implementation of the Library's logical preservation metadata schema. It details the data elements needed to support the preservation of digital objects and forms the basis for the design of a database repository and input systems for collecting and storing preservation metadata. It incorporates a number of data elements needed to manage the metadata in addition to metadata relating to the digital object itself. The aim has been to produce a document that will serve as an implementation template while at the same time remaining consistent with standards being developed globally around preservation metadata.

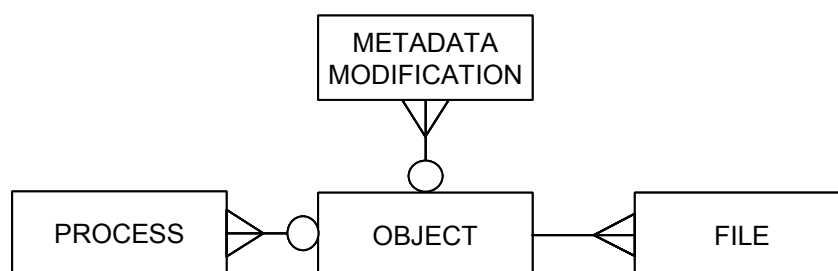
1.3. The Data Dictionary and a Metadata Repository - Introduction

This data dictionary defines the preservation metadata fields from a design or implementation perspective. The dictionary is based on the logical preservation metadata model (Revised version²) and maintains the overall structure and data relationships contained there. Individual fields however have been adjusted to facilitate their population with readily available structured data that conforms to recognized standards. In some instances this has resulted in the logically defined fields being split into subfields to enable multiple dimensions or sub elements to be recorded.

¹ The National Library of New Zealand (Te Puna Matauranga o Aotearoa) Act 2003.
<http://www.natlib.govt.nz/en/about/1keypolnlact.html> (Accessed 16 July 2003)

² National Library of New Zealand. 2003. *Metadata Standards Framework – Preservation Metadata (Revised)*.
<http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta> (Accessed 16 July 2003)

Data fields are based on the following structure:



The above model specifies the following relationship rules:

- An Object **may** have one or more Processes associated with it
- An Object **may** have one or more Metadata Modifications associated with it
- An Object **must** have one or more Files associated with it
- A Process **must** always be associated with a single Object
- A Metadata Modification **must** always be associated with a single Object
- A File **must** always be associated with a single Object

Data dictionary information for each field contains the following details:

1.4. Field Reference Number and Name

Definition	A textual description of the data field
Standards Used	The name or a reference to any external or internal standard used to define the data values.
Data Type	Classification of the field. Options are: String Date/Time Positive Integer Enumerated Type (restricted to list) Enumerated Type (restricted to external standard) Reference (pointer to another object)
Required	Shows the degree to which this field is compulsory. Options are: M = mandatory MA = mandatory if applicable R = recommended O = optional
Values	Where data type is ‘Enumerated Type (restricted to list)’ the actual permissible values and their meanings are listed. For example the field Video Sound may be limited to: yes no
Notes	Any additional information e.g. usage notes, references to additional documentation or examples.

The numbering within this data dictionary corresponds to the numbers used in the logical model. In some instances fields have been expanded from the logical model to cater for multiple dimensions associated with that field. For example Image Dimension (element 3.12.2 in the schema) has been expanded into Width (3.12.2.1) and Length (3.12.2.2) in the data model.

1.5. Structure of Paper

This document comprises the following sections:

Section 1	Executive Summary and overview
Section 2	The Detailed Entity List
Appendices	The Schema Map
	A Logical / Design Model Data Comparison table
	Preservation Metadata Model
	Associated Documents

2. Entity List

2.1. Entity 1 - Object

DEFINITION: A file or aggregation of files for which a Preservation Master has been created and for which preservation metadata is created and maintained. It is derived from a digital original or a digitised version of non-digital material acquired or created by the Library.

PURPOSE: The Object entity contains information about a logical object that exists as a file or aggregation of associated files. The object is held in the Digital Archive and will also have a range of other metadata relating to it including descriptive and collection management records. This entity contains information that identifies the Preservation Master and describes those characteristics relevant to preservation management, including the processes undertaken to generate the Preservation Master from the original object acquired by the Library.

1.1 Name of Object

Definition	Plain text name given to material by the creator/curator/selector and used to identify the object for users.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Expected to be consistent with name held as part of descriptive metadata and name held in collection management system. This field may not be required as a separate preservation metadata element once an integrated metadata structure is implemented.

1.2 Reference Number

Definition	Reference number or identification number used as the object identifier by the Library's core bibliographic/description or collection management applications.
Standards Used	Conforms to data structures set by other NLNZ systems
Data Type	Reference
Required	M
Values	
Notes	This field may not be required as a separate preservation metadata element once an integrated metadata structure is implemented.

1.3 Object Identifier

Definition	An internal identifier assigned to digital objects within NLNZ. This is a non-intelligent numeric number allocated sequentially to digital objects.
Standards Used	
Data Type	Positive Integer
Required	M
Values	
Notes	This field may not be required as a separate preservation metadata element once an integrated metadata structure is implemented and a fully functional persistent identifier system is in place. It is a shared data requirement with other metadata categories and will eventually be accessible through use of the persistent identifier mechanism.

1.4 Group Identifier

Definition	An internal identifier assigned to the objects that comprise an object group within NLNZ.
Standards Used	
Data Type	Positive Integer
Required	MA
Values	
Notes	An object group is an intellectual construct that links together a number of objects. The identifier may be used as an internal management tool to associate the objects within an object group. This identifier will be mandatory when an object is assigned to a group (see 1.8, 1.9 below).

1.5 Original Identifier

Definition	An indicator showing whether a digital original exists for this object.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	M
Values	Yes No
Notes	

1.6 Persistent Identifier

Definition	A published international identifier for digital objects. For persistent identifiers assigned by NLNZ, this incorporates the Object Identifier (see 1.3 Object Identifier above).
Standards Used	
Data Type	String
Required	M
Values	
Notes	<p>The Persistent Identifier can be either assigned by an external organisation, in which instance the format is dependant on the scheme adopted, or it is assigned by NLNZ and will be based on the Object Identifier (see 1.3 above) and will use the Library's selected persistent identifier scheme.</p> <p>For example, in the latter case the Persistent Identifier might comprise the following three components:</p> <p>1727 The code assigned to identify the object as belonging to NLNZ (there is potential to implement this level code as a national, ie. New Zealand level, code).</p> <p>10 – 19 Use of the code within this range is at the discretion of NLNZ. Current proposal is that 10 is used for archived objects and 11 for non-archived objects.</p> <p>IID The Object Identifier (Element 1.3) is used to ensure the PID is unique.</p> <p>In this example a persistent identifier based on the Handle System might look like 1727.10/1234</p> <p>It is assumed that access to the physical preservation master files will be through use of the persistent identifier system. If for some reason the persistent identifier system is not available, access will still be possible via a file search based on file name/directory.</p>

1.7 Preservation Master Creation Date

Definition	The date when the first Preservation Master for this object was created.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	8 digit numeric field in the format <code>yyyymmddthhmmssddd</code> Example is: 20030226t171136044 All date and time fields will be held in the above format with seconds recorded and also three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software.

1.8 Object Classification

Definition	A classification of the object based on the relationships between the files that make it up.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	Simple Object Complex Object
Notes	A simple object comprises a single file A complex object comprises a number of dependent files

1.9 Structural Type

Definition	The type of object being described in the Preservation Metadata record.
Standards Used	http://dublincore.org/documents/2002/07/13/dcmi-type-vocabulary/
Data Type	Reference
Required:	Y
Values	
Notes	Examples are, Image, Interactive Resource Choice of appropriate preservation strategy may depend on knowing the structural type

1.10 Hardware Environment

Definition	A description of the hardware environment required to access or use the object for preservation activities
Standards Used	
Data Type	String
Required	O
Values	
Notes	Data in this field will be manually entered and only supplied where specialised hardware is required

1.11 Software Environment

Definition	A description of the software environment required to access or use the object for preservation activities.
Standards Used	
Data Type	String
Required	O
Values	
Notes	Data in this field will be manually entered and only supplied where specialised software is required

1.12 Installation Requirements

Definition	Details of any specialised requirements needed to install or initiate access to the object for preservation activities.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

1.13 Access Inhibitors

Definition	Details of any method used to restrict access to the object that may impact on preservation activities.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

1.14 Access Facilitators

Definition	Any systems or methods used to enhance access to the object and needed to be maintained or considered in preservation activities.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

1.15 Quirks

Definition	Any characteristic of the original that may appear as a loss in functionality or change in the look and feel of a Preservation Master.
Standards Used	
Data Type	String
Required	O
Values	
Notes	To prevent time and effort being spent trying to solve problems that were inherent in the original. Examples are, Eg. For all Web documents in the object produced prior to HTML 4, the text format tag is no longer supported Eg. The Shockwave files could not be captured from the source document at the time of web harvesting

1.16 Metadata Record Creator

Definition	The name(s) of the individual, business unit and/or agency responsible for the creation of this preservation metadata information.
Standards Used	
Data Type	String
Required	M
Values	
Notes	It is anticipated that this information will be obtained from the logon Id of the individual initiating the change. This logon Id will be expanded to the person's name before being stored in this data field.

1.17 Date of Metadata Record Creation

Definition	Date that preservation metadata was first collected for this object.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	8 digit numeric field in the format yyymmddthhmmssddd Example is: 20030226t171136044 Where possible, this information will be system derived rather than manually entered. All date and time fields will be held in the above format with seconds recorded and also three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software.

1.18 Structural Composition

Definition	The name of a file, which for complex objects and object groups contains details of the individual files that comprise the object, including the overall directory structure and the hierarchical position of each file.
Standards Used	
Data Type	String
Required	MA
Values	
Notes	This field is mandatory for complex objects and object groups. The filename consists of the Object's Identifier followed by _ss.txt. For example 123456_ss.txt. This file describes for complex objects and object groups details of the individual files that comprise the object, including the overall directory structure and the hierarchical position of each file to enable component files of a Preservation Master to be reassembled in their correct structure. It is anticipated that the structural type file for complex objects will be held at the root level within the directory established for that object. For object groups the structural context file will exist in the object directory even if individual components of the group may be in other directories, for example where there is both simple and complex objects within a group.

1.19 Comments

Definition	Any other information relevant to the object's preservation.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.2. Entity 2 - Process

DEFINITION: Any action performed on a Preservation Master whether that Preservation Master is changed or not (including the processes undertaken to generate the Preservation Master from the original object acquired by the Library). A process is carried out by an individual or agency, it has an objective, a result and takes place at a specific date and time.

PURPOSE: The Process entity records a complete history of actions performed on all Preservation Master objects.

The Process entity contains information about all relevant details of any processes applied to a Preservation Master throughout its life cycle. It includes the software used, specific settings or actions, details of all critical equipment and responsible persons and/or agencies. This entity records only processes carried out whilst the Preservation Master is the responsibility of NLNZ.

2.1 Object Identifier

Definition	An internal identifier assigned to digital objects within NLNZ. This is a non-intelligent numeric number allocated sequentially to digital objects.
Standards Used	
Data Type	Positive Integer
Required	M
Values	
Notes	This is the key field that associates process details with the appropriate object (see 1.3 Object Identifier above).

2.2 Process Type

Definition	A classification of a preservation processes.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	Creation of first Preservation Master from Original Creation of subsequent Preservation Master through migration Creation of subsequent Preservation Master through emulation Creation of new Object from file(s) within this Object Data validation Creation of dissemination copy
Notes	This list is likely to expand as additional classifications are determined.

2.3 Purpose

Definition	The reason why this Process has been undertaken.
Standards Used	
Data Type	String
Required	O
Values	
Notes	This field is only completed if the process classification in Field 2.2 is insufficient to provide necessary information regarding the reason for performing the process e.g. why the migration was necessary.

2.4 Person / Agency Performing Process

Definition	The individual(s), business unit and / or agency that carried out the process.
Standards Used	
Data Type	String
Required	M
Values	
Notes	

2.5 Permission

Definition	The individual(s), business unit and / or agency who approved the process.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Permission may have a functional dependency with 2.3 Process Type in that some processes may require specific permissions.

2.6 Permission Date

Definition	The date when permission was given to perform the process.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	8 digit numeric field in the format <code>yyymmddthhmmssddd</code> Example is: 20030226t171136044 All date and time fields will be held in the above format with provision for seconds and three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software. It is unlikely that time information will be recorded for Permission Date unless the data is system generated.

2.7 Hardware Used

Definition	A description of any noteworthy hardware used in performing the process.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.8 Software Used

Definition	A description of any noteworthy software used in performing the process.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.9 Steps

Definition	A description of the actions performed to initiate and complete the process.
Standards Used	
Data Type	String
Required	M
Values	
Notes	

2.10 Result

Definition	The outcomes of the process where these are worth being recorded. Can include confirmation of intended changes, loss of functionality or highlighting of unexpected outcomes.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.11 Guidelines

Definition	Any standards or procedures used when performing the process.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.12 Completion Date/Time

Definition	The date and time the process was completed.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	<p>8 digit numeric field in the format yyymmddthhmmssddd Example is: 20030226t171136044</p> <p>All date/time fields will be held in the above format with seconds recorded and also three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software.</p> <p>Ideally this information will be system generated as a by-product of the process and then fed back into the preservation metadata repository. This may not be feasible initially and be built as a later enhancement.</p>

2.13 Comments

Definition	Any other matter of relevance to preservation involved in the process.
Standards Used	
Data Type	String
Required	O
Values	
Notes	

2.3. Entity 3 - File

DEFINITION: Contains technical information about a single file associated with a digital object.

PURPOSE: The File entity contains technical information required for preservation purposes about the characteristics of all files pertaining to digital objects. Elements 3.1 through to 3.11 are common to all file types. Elements numbered 3.12 onwards are specific to the file type specified.

Note that elements 3.12 through 3.12.8 have been mapped to the NISO Data Dictionary - Technical Metadata for Digital Still Images draft standard (see http://www.niso.org/standards/resources/Z39_87_trial_use.pdf).

3.1 Object Identifier

Definition	An internal identifier assigned to digital objects within NLNZ. This is a non-intelligent numeric number allocated sequentially to digital objects.
Standards Used	
Data Type	Positive Integer
Required	M
Values	
Notes	This is the key field that associates file details with the appropriate object (see 1.3 Object Identifier above).

3.2 File Identifier

Definition	An internal identifier – a number assigned by NLNZ to a digital file.
Standards Used	
Data Type	Reference
Required	MA
Values	
Notes	This field is used to distinguish files within a complex object. The File Identifier will contain the same values as the Object Identifier where the files are associated with a complex object. This number will be added to by a running number for all the files in the complex object. For example, where a complex object has an Object Identifier of 1234, each file within that complex object will be allocated a File Identifier based on that Object Identifier, eg. 1234-1, 1234-2 etc.

3.3 File Path

Definition	The directory path within a complex object that points to this file.
Standards Used	
Data Type	Reference
Required	MA
Values	
Notes	This field is Mandatory for complex objects only. The file path is internal to the complex object and excludes both the object directory name and the file name. For example for the file kiwi egg.jpg within the object “Birds of New Zealand” this field could contain /images/kiwi/ This field is required as it is possible for two files within a complex object to have the same file name. Without this path data it would not be possible to differentiate between the files and so associate the file level metadata with the correct physical file.

3.4 Filename & Extension

Definition	The name of the file as it is held in the digital archive.
Standards Used	
Data Type	Reference
Required	M
Values	
Notes	<p>For files associated with a complex object, the file names are unchanged from the source material e.g. 'Annual Report.html'. For files associated with simple objects or object groups, the file name is altered to contain the Object IID, a role code and the file extension e.g. '123456_PM.doc'. The role type for Preservation Master files will always be _PM.</p> <p>The reason for files associated with complex objects retaining their original names is that each complex object will have its own directory, thereby ensuring that path/file name duplication cannot occur. Other files will be held in a shared directory and therefore a name change is necessary to ensure that names are not duplicated.</p>

3.5 Former File Name

Definition	The name (including extension) of the file as it existed in the source material.
Standards Used	
Data Type	Reference
Required	MA
Values	
Notes	<p>This field is Mandatory for simple objects only.</p> <p>Files associated with simple objects do not retain their original names and instead have the file name replaced by a combination of the file identifier and a 'role' type. This is to ensure that filename duplication cannot occur. The purpose of this field is to record the original name in the event of having to recreate the preservation master from the source material.</p>

3.6 File Size

Definition	The size of the file in bytes.
Standards Used	
Data Type	Positive integer
Required	M
Values	
Notes	

3.7 File Date/Time

Definition	Date/Time that this file was created.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	<p>8 digit numeric field in the format <code>yyymmddthhmmssddd</code> Example is: 20030226t171136044</p> <p>All date and time fields will be held in the above format with seconds recorded and also three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software.</p>

3.8 MIME Type

Definition	The Multipurpose Internet Mail Extensions Type for the file expressed as type/subtype.
Standards Used	http://www.iana.org/assignments/media-types/index.html
Data Type	Enumerated Type (restricted to external standard)
Required	M
Values	
Notes	Examples are: video/quicktime, image/jpeg

3.9 File Format

Definition	The format of the file taken from the name of its type or the application used to create it.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Examples are: MPEG, Microsoft Word Although file formats may be able to be verified against a list, the volatility of file types may make this impractical to enforce.

3.10 File Format Version

Definition	The format of the file identified in 3.9.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Examples are: V2.0, XP Although file format versions may be able to be verified against a list, the proliferation of new versions may make this impractical to enforce.

3.11 Target Indicator

Definition	An indicator to show whether this file is the entry point for accessing a complex object.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	MA
Values	Yes No
Notes	Applies only to complex objects.

3.12 Image

3.12.1 Resolution

3.12.1.1 Image Resolution – Sampling Frequency Unit

Definition	The unit of measurement for X Sampling Frequency and Y Sampling Frequency.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	1= no absolute unit of measurement 2 = inches 3 = centimetres
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.1.2

3.12.1.2 Image Resolution – X Sampling Frequency

Definition	The number of pixels per Sampling Frequency Unit in the image width (horizontal dimension).
Standards Used	
Data Type	Positive integer
Required	MA
Values	
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.1.3 Mandatory when Sampling Frequency Unit = 2 or 3. When Sampling Frequency Unit = 1, the value of this field must be null.

3.12.1.3 Image Resolution – Y Sampling Frequency

Definition	The number of pixels per Sampling Frequency Unit in the image length (vertical dimension).
Standards Used	
Data Type	Positive integer
Required	MA
Values	
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.1.4 Mandatory when Sampling Frequency Unit = 2 or 3. When Sampling Frequency Unit = 1, the value of this field must be null.

3.12.2 Dimensions**3.12.2.1 Image Dimension – Width**

Definition	The width of the digital image, i.e. horizontal or X dimension, in pixels.
Standards Used	
Data Type	Positive integer
Required	MA
Values	
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.1.5

3.12.2.2 Image Dimension - Length

Definition	The length of the digital image, i.e. vertical or Y dimension, in pixels.
Standards Used	
Data Type	Positive integer
Required	MA
Values	
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.1.6

3.12.3 Image - Bits Per Sample

Definition	The number of bits per component for each pixel.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	M
Values	<p>1 = 1 bit (bitonal)</p> <p>4 = 4 bit grayscale</p> <p>8 = 8 bit grayscale or palletised colour</p> <p>8,8,8 = RGB</p> <p>16,16,16 = TIFF, HDR (high dynamic range)</p> <p>8,8,8,8 = CMYK</p>
Notes	<p>Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.2.1</p> <p>Note that this field allows a different number of bits per component for each component corresponding to a pixel. For example, RGB colour data could use a different number of bits per component for each of the three colour panes. Most RGB files will have the same number of Bits Per Sample for each component. Even in this case, the writer must write all three values.</p>

3.12.4 Image – Photometric Interpretation - Colour Space

Definition	Designates the colour space of the decompressed image data.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	<p>0 = WhiteIsZero</p> <p>For bilevel and grayscale images: 0 is imaged as white. $2^{**BitsPerSample-1}$ is imaged as black. This is the normal value for Compression=2.</p> <p>1 = BlackIsZero</p> <p>For bilevel and grayscale images: 0 is imaged as black. $2^{**BitsPerSample-1}$ is imaged as white. If this value is specified for Compression=2, the image should display and print reversed.</p> <p>2 = RGB</p> <p>In the RGB model, a colour is described as a combination of the three primary colours of light (red, green, and blue) in particular concentrations. For each of the three components, 0 represents minimum intensity, and $2^{**BitsPerSample-1}$ represents maximum intensity. Thus an RGB value of (0,0,0) represents black, and (255,255,255) represents white, assuming 8-bit components. For Planar Configuration = 1, the components are stored in the indicated order: first Red, then Green, then Blue. For Planar Configuration = 2, the Strip Offsets for the component planes are stored in the indicated order: first the Red component plane Strip Offsets, then the Green plane Strip Offsets, then the Blue plane Strip Offsets.</p> <p>3 = Palette colour</p> <p>In this model, a colour is described with a single component. The value of the component is used as an index into the red, green, and blue curves in the Colour Map field to retrieve an RGB triplet that defines the colour. When Photometric Interpretation = 3 is used, Colour Map must be present and Samples Per Pixel must be 1.</p> <p>4 = Transparency Mask</p> <p>This means that the image is used to define an irregularly shaped region of another image in the same TIFF file. Samples Per Pixel and Bits Per Sample must be 1. PackBits compression is recommended. The 1-bits define the interior of the region; the 0-bits define the exterior of the region.</p> <p>5 = CMYK</p> <p>6 = YCbCr</p> <p>8 = CIE Lab</p>
Notes	Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 6.1.4.1

3.12.5 Image – Photometric Interpretation – ICC Profile Name

Definition	The name of the International Color Consortium (ICC) profile used.
Standards Used	
Data Type	String
Required	MA
Values	
Notes	Mandatory if a registered profile is used. If the ICC profile used is a well-known and well-documented profile, record the information in this data element. Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 6.1.4.2.1

3.12.6 Image - Colour Map - Reference

Definition	The location of the file containing the colour map.
Standards Used	http://www.color.org/icc_specs2.html
Data Type	Reference
Required	MA
Values	
Notes	Mandatory for palletised colour images, Photometric Interpretation = 3. As noted in the TIFF (Revision 6.0) definition, Colour Map is synonymous with colour lookup table (CLUT). When Photometric Interpretation = 2, there is no Colour Map ; in other words, there is no Colour Map in RGB images (TIFF, p.24). The reference data type accommodates the practice of generating a colour map at the beginning of each session. If the colour map exists in an external file, it must be referenced in this element. Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 8.2.4.1

3.12.7 Image - Orientation

Definition	Orientation of the image saved on disk e.g. normal, normal rotated 180°.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	1 = normal* 3 = normal rotated 180° 6 = normal rotated cw 90° 8 = normal rotated ccw 90° 9 = unknown
Notes	* “Normal” is defined as follows: when opened, the top (0th) row of pixels corresponds to the visual top of the image and the first (0th) column of pixels on left corresponds to the visual left-hand side of the image. Consult TIFF (Revision 6.0) for additional values referring to mirrored images. (Note that TIFF/EP supports only five values, which are proposed above as the finite list of enumerated type values.) This field is to be used to record only the orientation of the image, <i>not</i> the orientation of the source to the device (e.g., camera) used to capture the image. Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 6.2.4

3.12.8 Compression

3.12.8.1 Image - Compression Scheme

Definition	The compression scheme used to store the image data.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	M
Values	1 = Uncompressed 2 = CCITT 1D 3 = CCITT Group 3 4 = CCITT Group 4 5 = LZW 6 = JPEG 32773 = PackBits (simple byte-oriented run-length scheme)
Notes	Values above are drawn from TIFF specification (p. 117) though institutions are encouraged to devise a local enumerated list to allow for the addition of new values as technology changes. This data element allows for the designation of sub-elements in order to record the level of compression applied (see 3.12.8.2 Compression Level). Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 6.1.3.1

3.12.8.2 Image - Compression Level

Definition	Designates the level of compression used in 3.12.8.1 Compression Scheme .
Standards Used	
Data Type	Positive integer
Required	MA
Values	
Notes	This field is mandatory if the Compression Scheme does not = 1. Equates to National Information Standards Organisation (NISO) – Technical Metadata for Still Images - Field 6.1.3.2

3.13 Audio

3.13.1 Audio - Resolution

Definition	The rate of sampling, in samples per second, used to create the audio file. Also known as sample rate or sample frequency.
Standards Used	
Data Type	Positive integer (5 or 6 digits)
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A174 Examples are 32100, 44100, 192000 KHz

3.13.2 Audio - Duration

Definition	The length of the audio recording in hours, minutes and seconds and three digits for representing decimal fractions of a second.
Standards Used	ISO 8601
Data Type	Date/Time (duration)
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A428 Format hhmmssddd, example would be 01:27:38:247

3.13.3 Audio - Bit Rate

Definition	The word length used to encode the audio. Consequently an indication of dynamic range. It is the maximum number of significant bits for the value without compression.
Standards Used	
Data Type	Positive integer
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A175 Examples are: 16, 20, 24

3.13.4 Audio - Compression

Definition	The name of the compression scheme, noise reduction scheme, or other non-linear processing applied to an audio signal. Note that audio compression, or bit rate reduction is a non-reversible, “lossy” process.
Standards Used	
Data Type	String
Required	MA
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A257 Examples are: MPEG 3; Dolby A; IEC pre-emphasis

3.13.5 Audio - Encapsulation**3.13.5.1 Audio Encapsulation - Name**

Definition	The name of the delivery format of the file.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A222 Examples are: Real Audio II, BWF

3.13.5.2 Audio - Encapsulation - Version

Definition	The version level of the delivery format of the file.
Standards Used	
Data Type	String
Required	O
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A222 Example: ver 1

3.13.6 Audio - Channels

Definition	A classification of the sound format type identifying the number of channels and how they are related to each other.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	M
Values	MonoSingle channel 2 channel stereo 2 channel mono 4 channel 5 channel surround
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A132 This list is not exhaustive and is likely to grow as more audio media is archived.

3.14 Video**3.14.1 Dimensions****3.14.1.1 Video - Frame Dimension - Width**

Definition	The width of the digital image, i.e. horizontal or X dimension, in pixels.
Standards Used	
Data Type	Positive integer
Required	M
Values	
Notes	

3.14.1.2 Video - Frame Dimension -Length

Definition	The length of the digital image, i.e. vertical or Y dimension, in pixels.
Standards Used	
Data Type	Positive integer
Required	M
Values	
Notes	

3.14.2 Video - Duration

Definition	The length of the video recording in hours, minutes and seconds and three digits representing decimal fractions of a second.
Standards Used	ISO 8601
Data Type	Date/Time (duration)
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A428 Format hhmmssddd, example would be 01:27:38:247

3.14.3 Video – Number of Frames

Definition	The number of frames present in the video recording.
Standards Used	
Data Type	Positive Integer
Required	M
Values	
Notes	

3.14.4 Video - Frame Rate

Definition	The rate at which the video should be shown to achieve the intended effect – expressed in frames per second (fps).
Standards Used	
Data Type	Positive integer
Required	M
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A44 Example: 25

3.14.5 Video - Codec Method**3.14.5.1 Video - Codec Method - Name**

Definition	The name of the codec method applied to the video.
Standards Used	
Data Type	String
Required	MA
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A226 The possible compression schemes could be treated as a restricted list. However the possibility of new schemes coming into existence without notice could cause problems with maintaining this list. This field does not specify the actual codec used to create the file but rather the method used. It is possible for a method to be usable by multiple codecs. The name and version level are sourced from a single field.

3.14.5.2 Video - Codec Method - Version

Definition	The version level of the codec method applied to the video.
Standards Used	
Data Type	String
Required	O
Values	
Notes	Equates to European Broadcasting Union – Preservation Metadata Scheme - Attribute A226 The possible compression schemes could be treated as a restricted list. However the possibility of new schemes coming into existence without notice could cause problems with maintaining this list. This field does not specify the actual codec used to create the file but rather the method used. It is possible for a method to be usable by multiple codecs. The name and version level are sourced from a single field.

3.14.6 Video - Aspect Ratio

Definition	The desired aspect ratio of the image on screen.
Standards Used	
Data Type	String
Required	M
Values	
Notes	Contains 2 values separated by a colon. Some files produced for display on non-square-pixel monitors have a desired aspect ratio that differs from the ratio of horizontal to vertical pixels. Example: 4:3

3.14.7 Video - Scan Mode

Definition	An indicator showing whether the digital item is scanned in a progressive or interlaced mode.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	Progressive Interlaced
Notes	

3.14.8 Video - Sound Indicator

Definition	An indicator of the presence of sound in the video file.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	M
Values	Yes No
Notes	If the value is 'yes', then the video file will also be associated with an instance of the Audio metadata (3.13) in addition to the Video metadata (3.14)

3.15 Text**3.15.1 Text - Character Set**

Definition	The character set used when creating the file.
Standards Used	http://www.iana.org/assignments/character-sets
Data Type	Enumerated Type (restricted to external standard)
Required	M
Values	
Notes	Examples are: ASCII, Unicode, EBCDIC, UTF-8

3.15.2 Text – Markup Language

Definition	The type of mark up language used to mark up the document.
Standards Used	
Data Type	Enumerated Type (restricted to list)
Required	O
Values	SGML XML HTML TEI
Notes	

2.4. Entity 4 - Metadata Modification

DEFINITION: Details of a change to an existing metadata record.

PURPOSE: The Metadata Modification entity records information about the history of changes made (by whom and when) to the preservation metadata subsequent to the initial metadata creation. Recording changes to the preservation metadata record acknowledges that the record is itself an important body of data about the object that requires management over time.

4.1 Object Identifier

Definition	An internal identifier assigned to digital objects within NLNZ. This is a non-intelligent numeric number allocated sequentially to digital objects.
Standards Used	
Data Type	Positive Integer
Required	M
Values	
Notes	This is the key field that associates metadata modification details with the appropriate object (see 1.3 Object Identifier above).

4.2 Metadata Record Modifier

Definition	The name of the person who made this modification to the preservation metadata record.
Standards Used	
Data Type	String
Required	M
Values	Free form text
Notes	It is anticipated that this information will be obtained from the logon Id of the individual initiating the change. This logon Id will be expanded to the person's name before being stored in this data field.

4.3 Date/Time

Definition	The date/time that this change to the preservation metadata information was completed.
Standards Used	ISO 8601
Data Type	Date/Time
Required	M
Values	
Notes	8 digit numeric field in the format yyymmddthhmmssddd Example is: 20030226t171136044 All date/time fields will be held in the above format with seconds recorded and also three digits for representing decimal fractions of a second – where this level of detail is available. Holding information at such a detailed level does not force this level of granularity to be reported for dates or time as this is under the control of the user interface software.

4.4 Field Modified

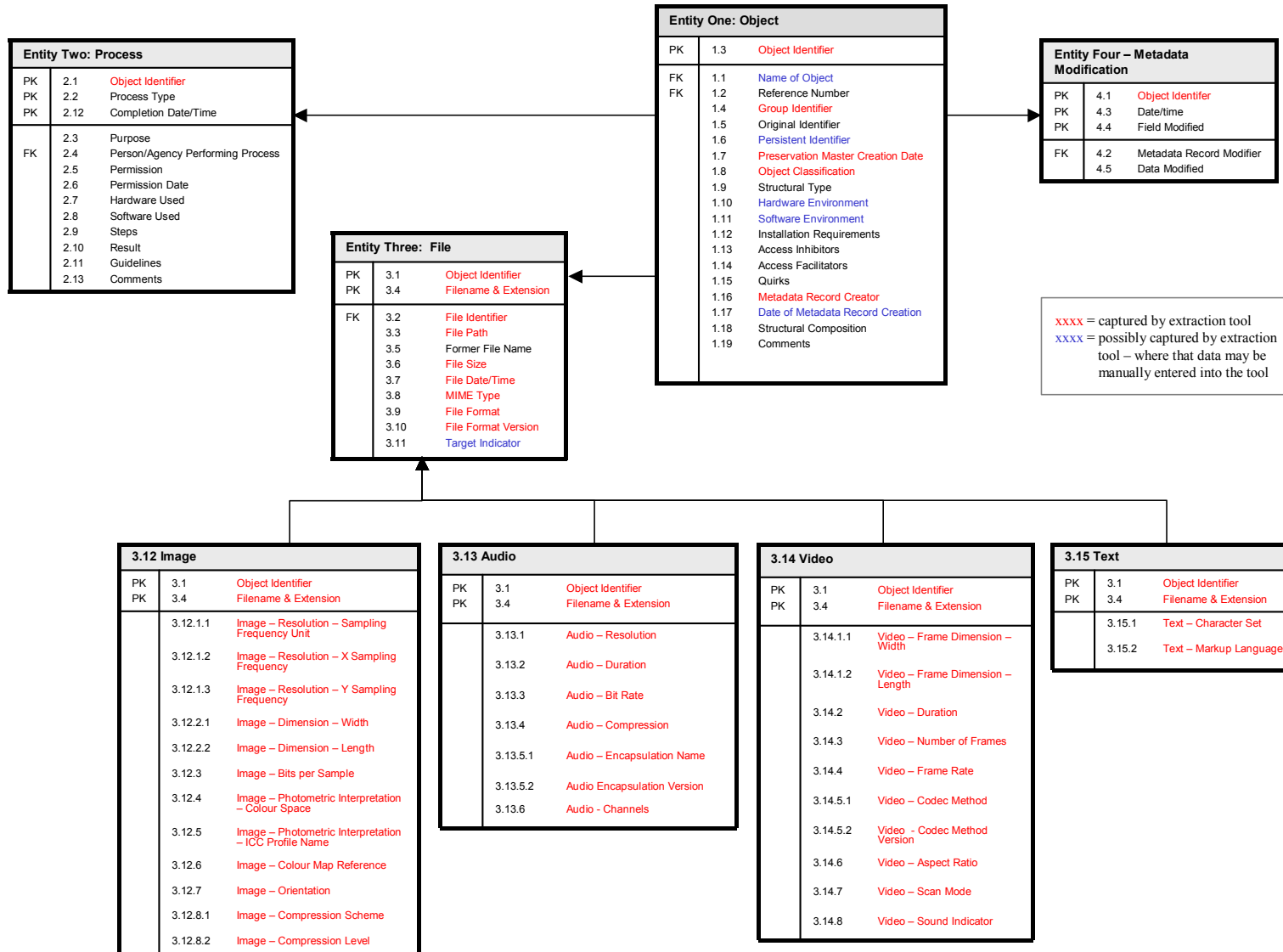
Definition	Identification of the field the modification was made to.
Standards Used	
Data Type	Enumerated type (restricted to list)
Required	
Values	Contains the field identifier (excluding key fields) 1.2 through to 1.19 2.2 through to 2.13 3.2 through to 3.15.2
Notes	Each field changed will result in a set of Metadata Modification information being recorded. The metadata modification fields themselves cannot be subsequently modified.

4.5 Data Modified

Definition	A record of the preservation metadata content prior to being modified.
Standards Used	
Data Type	String
Required	M
Values	
Notes	If a free text field is modified, the entire text, prior to modification will be recorded. The potential for holding only selected (changed) text needs investigation.

Appendices

Appendix 1 – Schema Map



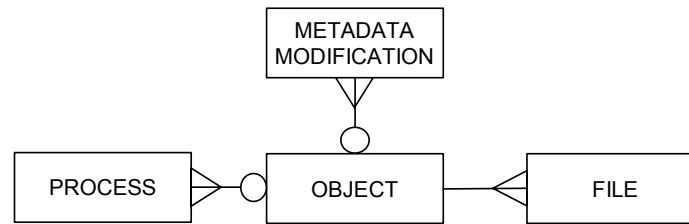
Appendix 2 – Preservation Metadata Model

ENTITY 1 - OBJECT

- 1.1 Name of object
- 1.2 Reference number
- 1.3 Object identifier
- 1.4 Group Identifier
- 1.5 Original Identifier
- 1.6 Persistent identifier
- 1.7 Preservation Master Creation Date
- 1.8 Object Classification
- 1.9 Structural type
- 1.10 Hardware environment
- 1.11 Software environment
- 1.12 Installation requirements
- 1.13 Access inhibitors
- 1.14 Access facilitators
- 1.15 Quirks
- 1.16 Metadata record creator
- 1.17 Date of metadata record creation
- 1.18 Structural composition
- 1.19 Comments

ENTITY 2 - PROCESS

- 2.1 Object identifier
- 2.2 Process type
- 2.3 Purpose
- 2.4 Person/agency performing process
- 2.5 Permission
- 2.6 Permission date
- 2.7 Hardware used
- 2.8 Software used
- 2.9 Steps
- 2.10 Result
- 2.11 Guidelines
- 2.12 Completion date and time
- 2.13 Comments



ENTITY 3 - FILE

- 3.1 Object identifier
- 3.2 File identifier
- 3.3 File path
- 3.4 Filename and extension
- 3.5 Former filename
- 3.6 File size
- 3.7 File date and time
- 3.8 MIME type
- 3.9 File format
- 3.10 File format version
- 3.11 Target indicator

3.12 IMAGE:

- 3.12.1 Resolution
- 3.12.2 Dimensions
- 3.12.3 Bits per sample
- 3.12.4 Colour space
- 3.12.5 ICC profile name
- 3.12.6 Colour map reference
- 3.12.7 Orientation
- 3.12.8 Compression

3.13 AUDIO:

- 3.13.1 Resolution
- 3.13.2 Duration
- 3.13.3 Bit rate
- 3.13.4 Compression
- 3.13.5 Encapsulation
- 3.13.6 Channels

ENTITY 3 – FILE (cont.)

3.14 VIDEO:

- 3.14.1 Frame dimensions
- 3.14.2 Duration
- 3.14.3 Number of frames
- 3.14.4 Frame rate
- 3.14.5 Codec method
- 3.14.6 Aspect ratio
- 3.14.7 Scan mode
- 3.14.8 Sound indicator

3.15 TEXT:

- 3.15.1 Character set
- 3.15.2 Mark up language

3.16 DATASETS:

- Use common elements only
- 3.1 – 3.11

3.17 SYSTEM FILES:

- Use common elements only
- 3.1 – 3.11

ENTITY 4 - METADATA MODIFICATION

- 4.1 Object identifier
- 4.2 Metadata record modifier
- 4.3 Date and time
- 4.4 Field modified
- 4.5 Data modified

Appendix 3 - Logical / Design Data Comparison

This table outlines the changes made to the preservation metadata schema logical model in order to create this data dictionary version.

Logical Model ID	Logical Model Name	Design Schema ID	Data Dictionary Name	Logical Changes	Design Changes
1. Object					
1.1	Name of Object	1.1	Name of Object		
1.2	Reference Number	1.2	Reference Number		
1.3	Object Identifier	1.3	Object Identifier		
1.4	Group Identifier	1.4	Group Identifier		
1.5	Original Identifier	1.5	Original Identifier		
1.6	Persistent Identifier	1.6	Persistent Identifier		
1.7	Preservation Master Creation Date	1.7	Preservation Master Creation Date		
1.8	Object Classification	1.8	Object Classification		
1.9	Structural Type	1.9	Structural Type		
1.10	Hardware Environment	1.10	Hardware Environment		
1.11	Software Environment	1.11	Software Environment		
1.12	Installation Requirements	1.12	Installation Requirements		
1.13	Access Inhibitors	1.13	Access Inhibitors		
1.14	Access Facilitators	1.14	Access Facilitators		
1.15	Quirks	1.15	Quirks	The definition includes only problems that were inherent in the original	
1.16	Metadata Record Creator	1.16	Metadata Record Creator		
1.17	Date of Metadata Record Creation	1.17	Date of Metadata Record Creation		
1.18	Structural Composition	1.18	Structural Composition		
1.19	Comments	1.19	Comments		
2. Process					
2.1	Object Identifier	2.1	Object Identifier		
2.2	Process Type	2.2	Process Type		Has moved from being a free text field to a structured classification of why the process was performed.

Logical Model ID	Logical Model Name	Design Schema ID	Data Dictionary Name	Logical Changes	Design Changes
2.3	Purpose	2.3	Purpose		Changed to become an optional field that can be used to record additional details if required.
2.4	Person/Agency Performing Process	2.4	Person/Agency Performing Process		
2.5	Permission	2.5	Permission		
2.6	Permission Date	2.6	Permission Date		
2.7	Hardware Used	2.7	Hardware Used		
2.8	Software Used	2.8	Software Used		
2.9	Steps	2.9	Steps		
2.10	Result	2.10	Result		
2.11	Guidelines	2.11	Guidelines		
2.12	Completion Date/Time	2.12	Completion Date/Time		
2.13	Comments	2.13	Comments		
3. File					
3.1	Object Identifier	3.1	Object Identifier		
3.2	File Identifier	3.2	File Identifier		
3.3	File Path	3.3	File Path	Needed in both logical and physical models due to the possibility of a complex object containing multiple files with the same name. This field allows the metadata to be associated with the correct file.	
3.4	Filename and Extension	3.4	Filename and Extension		
3.5	Former Filename	3.5	Former Filename		
3.6	File Size	3.6	File Size		
3.7	File Date/Time	3.7	File Date/Time		
3.8	MIME Type	3.8	MIME Type		The MIME type and file format have been split into separate fields as these represent different file attributes
3.9	File Format	3.9	File Format		
3.10	File Format Version	3.10	File Format Version		

Logical Model ID	Logical Model Name	Design Schema ID	Data Dictionary Name	Logical Changes	Design Changes
3.11	Target Indicator	3.11	Target Indicator	Needs to change to a Yes/No indicator rather than the file name. The file name is already known as part of this file metadata set.	
Image					
3.12	Resolution	3.12	Resolution		Split into three fields for structured data recording and to conform to NISO standards
		3.12.1.1	Image Resolution – Sampling Frequency Unit		
		3.12.1.2	Image resolution – X Sampling Frequency		
		3.12.1.3	Image resolution – Y Sampling Frequency		
3.12.2	Dimensions	3.12.2	Dimensions		Split into two fields for structured data recording and to conform to NISO standards
		3.12.2.1	Image Dimension - Width		
		3.12.2.2	Image Dimension - Length		
3.12.3	Bits per Sample	3.12.3	Image – Bits per Sample		Name change only to conform to NISO terminology
3.12.4	Colour Space	3.12.4	Image – Photometric Interpretation – Colour Space		Name change only to conform to NISO terminology
3.12.5	ICC Profile Name	3.12.5	Image – Photometric Interpretation – ICC Profile Name		Name change only to conform to NISO terminology
3.12.6	Colour Map Reference	3.12.6	Image – Colour Map - Reference		Name change only to conform to NISO terminology
3.12.7	Orientation	3.12.7	Image Orientation		
3.12.8	Compression	3.12.8	Image Compression		Split into two fields for structured data recording and to conform to NISO standards
		3.12.8.1	Image – Compression Scheme		
		3.12.8.2	Image – Compression Level		
Audio					

Logical Model ID	Logical Model Name	Design Schema ID	Data Dictionary Name	Logical Changes	Design Changes
3.13.1	Resolution	3.13.1	Audio Resolution		
3.13.2	Duration	3.13.2	Audio Duration		
3.13.3	Bit Rate	3.13.3	Audio Bit Rate		
3.13.4	Compression	3.13.4	Audio Compression		
3.13.5	Encapsulation	3.13.5	Audio Encapsulation		Split into two fields for structured data recording
		3.13.5.1	Audio Encapsulation - Name		
		3.13.5.1	Audio Encapsulation - Version		
3.13.6	Channels	3.13.6	Audio Track Number and Type		
Video					
3.14.1	Frame Dimensions	3.14.1	Dimensions		Split into two fields for structured data recording
		3.14.1.1	Video - Frame Dimension - Width		
		3.14.1.2	Video - Frame Dimension - Length		
3.14.2	Duration	3.14.2	Video - Duration		
3.14.3	Number of Frames	3.14.3	Video – Number of Frames		
3.14.4	Frame Rate	3.14.4	Video - Frame Rate		
3.14.5	Codec Method	3.14.5	Video Codec Method		Split into two fields for structured data recording
		3.14.5.1	Video Codec Method - Name		
		3.14.5.2	Video Codec Method Version		
3.14.6	Aspect Ratio	3.14.6	Video - Aspect Ratio		New field. Data available in header and documented in other schemas
3.14.7	Scan Mode	3.14.7	Video - Scan Mode		New field. Data available in header and documented in other schemas
3.14.8	Sound Indicator	3.14.8	Video - Sound Indicator		Changed to a yes/no indicator
Text					
3.15.1	Character Set	3.15.1	Text – Character Set		
3.15.2	Mark up Language	3.15.2	Mark up Language		New field
Entity 4.	Meta Data Modification				

Logical Model ID	Logical Model Name	Design Schema ID	Data Dictionary Name	Logical Changes	Design Changes
4.1	Object Identifier	4.1	Object Identifier		
4.2	Metadata record modifier	4.2	Metadata Record Modifier		
4.3	Date/Time	4.3	Date/Time		
4.4	Field Modified	4.4	Field Modified		
4.5	Data Modified	4.5	Data Modified		Definition changed to include only data prior to modification

Appendix 4 – Associated Documents

The standards and frameworks contained in the following documents support the National Library's Metadata Standards Framework – Preservation Metadata by providing commonly accepted and pre-defined encoding standards for data entry. The purpose of using them is to avoid duplication of effort and to provide for interoperability with other systems.

This implementation schema may be read in conjunction with the earlier versions of the Preservation Metadata Schemas available at <http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta>

Data Dictionary - Technical Metadata for Digital Still Images

http://www.niso.org/standards/resources/Z39_87_trial_use.pdf (Last accessed 3 June 2003)

NISO Z39.87-2002/AIIM 20-2002 is a draft standard for trial use that specifies the technical metadata requirements for digital still images.

DCMI Type Vocabulary

<http://www.dublincore.org/documents/2002/07/13/dcmi-type-vocabulary/> (Last accessed 3 June 2003)

The DCMI Type Vocabulary provides a general, cross-domain list of approved terms that may be used as values to identify the genre of a resource.

ISO 8601: 2000 : 1988 (E) Data elements and interchange formats – Information interchange – Representation of dates and times

<http://www.iso.ch/> (Last accessed 3 June 2003)

ISO 8601: 2000 specifies numeric representations of date and time. This standard notation helps to avoid confusion in international communication caused by the many different national notations and increases the portability of computer user interfaces. The W3C date time format may also be used for better integration with the use of Dublin Core.

List of MIME Types

<ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/media-types> (Last accessed 3 June 2003)

[RFC2045, RFC2046] specifies that Content Types, Content Subtypes, Character Sets, Access Types, and conversion values for MIME mail will be assigned and listed by the Internet Assigned Numbers Authority (IANA).