

Herbert J. White II
Digital Imaging Operations Manger
Family and Church History Department
The Church of Jesus Christ of Latter-Day Saints

INTRODUCTION

This document will present a set of Categories, which will contain Metadata Elements and Sub-Elements that will assist in the description of a stored digital object.

Anne J. Gilliland-Swetland in the document "Introduction to Metadata, Pathways to Digital Information, setting the Stage" States:

"All of these perspectives on Metadata become important in the development of networked digital information systems, but they lead to a very broad conception of metadata. To understand this conception better, it is helpful to break it down into distinct categories – administrative, descriptive, preservation, use and technical metadata – that reflect key aspects of metadata functionality."

(<http://www.getty.edu/gri/standard/intrometadata>)

This document will outline for the user, information regarding the categories for metadata, as well as the element and sub-elements within each of the categories. This will also bring together diverse document imaging metadata elements, which have been described across the document imaging industry, under one umbrella of understanding.

It needs to be understood that Metadata is very fluid. There is not attempt to state that these are the only elements for a Document Imaging System. It is our belief that elements will be added and taken from this standard, as needs change in the document imaging industry. There is also no attempt to state that all of the elements listed must be used to describe a digital object. One must only use what is available and necessary to obtain access to the object.

OBJECTIVE

The object is to define the critical metadata elements within a document imaging system, enabling users to begin saving common meta elements. These elements can be directly tied to the digital object for better understanding of those objects.

Again, Anne J. Gilliland-Swetland in the document "Introduction to Metadata, Pathways to Digital Information, setting the Stage" States:

"Metadata creation and management have become a very complex mix of manual and automatic processes and layers created by many different functions and individuals at different points in the life of an information object As they move through each phase, the object acquires layers of metadata that can be associated with the object in several ways.

This metadata can be contained within the same envelope as the information object Metadata can also be attached to the information object through bi-directional pointers or hyperlinks, while the relationships between metadata and information objects, and between different aspects of metadata can be documented by registering them with a metadata registry."

(<http://www.getty.edu/gri/standard/intrometadata>)

The categories and elements identified in this document will be submitted to the Association for Information and Image Management (AIIM). AIIM is the Standard Developer for the American National Standards Institute (ANSI) who is responsible for the development of Standards in the United States.

WHAT IS METADATA?

In discussions with Archivists and Librarians it has been identified that metadata elements are not something new. These elements have been used in Catalogs, Registers, Abstracts, and Indexes for years on end. It has only been in recent history that "Metadata" has taken on a new meaning in the computer environment. The "Dublin Core" has stated:

"The term Metadata has been used only in the past 15 years, and has become particularly common with the popularity of the World Wide Web. But the underlying concepts have been in use for as long as collections of information have been organized. Library catalogs represent a well established variety of metadata that has served for decades as collection management and resource discovery tools."

(<http://www.oclc.org/research/projects/core/education/index.htm>)

Metadata has been described by many as "DATA ABOUT DATA".

Metadata is specifically identified data about other data or objects. It is the internet-age term for those elements, as stated previously, that librarians and archivists have traditionally put into Catalogs, Registers, Abstracts, and Indexes. In today's world these elements are most commonly referred to as descriptive information about Web based resources.

Metadata records consist of a set of elements with attributes, necessary to describe the resource, other data or objects, in question. For example within the Family and Church History Library Catalog, systems commonly use sets of metadata records with elements that describe a microfilm, book or resource file. Other library elements such as author, title, date of creation or publication, subject coverage, and the call number specifying location of the item on the shelf are referred to as Metadata.

Although the concept of metadata predates the Internet and the Web, worldwide interest in metadata standards and practices has exploded with the increase in digital imaging. The wide scale adoption of descriptive standards and practices for electronic resources will improve retrieval of resources from common internet sites. As noted by Weible and Lagoze, two leaders in the field of metadata development:

"The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself." (Weibel and Lagoze, 1997) (<http://www.oclc.org/oclc/research/projects/core/documents/wd-guide-current.htm>)

It is the need for "standardized descriptive metadata" that has compelled document-imaging industry and the Family and Church History Department to develop the overall metadata elements for digital images.

At this point in time, extensive metadata is our best way of minimizing the risk of a digital object becoming inaccessible. Properly used metadata can:

- *Identify the name of the work, who created it, who reformatted it, and other descriptive information*
- *Provide unique identification and links to organizations, files, or databases which have more extensive descriptive metadata about the work (this is particularly important in the likely event that the digital file and its external metadata become separated)*
- *Explain the technical environment needed to view the work, including applications and version numbers needed, decompression schemes, other files that need to be linked to it etc...*

There are various types of metadata, which appear unimportant today, but may prove critical for properly viewing these files in the future. (For example, saved information about a particular scanner's color profile will be critical for future color management systems to account for display device differences and to properly display colors on a particular device.) A good rule of thumb is to save any metadata that is cheap/easy to capture, or that someone has indicated might eventually be important.

*(Howard Besser, 1999 Digital Longevity, Chapter for School for Scanning Book
(<http://www.gseis.ucla.edu/~howard/papers/sfs-longevity.html>))*

What is Metadata? In the document-imaging context our definition would be:

Useful as well as usable informational data which provides the Administrative, Descriptive, Preservation, Technical and Use descriptions regarding the digital or analog object, within a document imaging environment."

WHAT ARE THE CORE ELEMENTS IN THIS SCHEME?

This Metadata Standard is intended to be a set of simple yet effective elements for describing a wide range of informational resources. The meta elements are listed within Five major categories. These Categories are:

- Administrative
- Discovery and Description
- Preservation
- Technical
- Use

Within these five categories are listed the meta elements and their sub-elements. These elements have been obtained by reviewing various metadata recommended practices on the World Wide Web. From these recommended practices we have taken those elements which apply to the document imaging industry. The web sites used to acquire this data are:

- Pandora:
 - <http://www.nla.gov.au/pandora/ldmv2.html>
- Dublin Core:
 - <http://purl.org/DC/Documents/rec-dces-19990702.htm>
- RLG:
 - <http://www.rlg.org/preserv/presmeta.html>
- Cedar Preservation Elements
 - <http://users.ox.ac.uk/cedars/papersAIW02.html>
- Meta Data Coalition:
 - <http://www.mdcinfo.com>
- Visual Resources Association Data Standards Committee (VRA Core Categories V3.0):
 - <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.html>
- ICA
 - [http://www.ica.org/cds/isad\(g\).pdf](http://www.ica.org/cds/isad(g).pdf)
- DIG35
 - <http://www.digitalimaging.org>

ELEMENT CONTENT AND CONTROLLED VOCABULARIES

The elements and sub-elements have been selected from those Standards Practices listed above. This provides for a "Controlled Vocabulary" which will unify the Practices through common definition. This will enable the sharing of common Meta elements between organizations. Through the adoption of these elements as the "Document Imaging Metadata Elements," Organizations can share images and metadata associated with them.

COMMON DESCRIPTION TABLES FOR THE METADATA ELEMENTS

For each of the Elements and Sub-Elements there is a common table, which provide the following information:

- Element Name'
- Definition of the Element
- An area to list the various computer systems where the data can be found and if the element is applicable.
- The Sub-Elements with Definitions
- XML Code for the element
- The crosswalk between other Recommended Practice for Metadata
- The version number that the present metadata element is under
- Who is the registration authority for the metadata
- The reason for the Data to exist
- The data type
- The language of the Data
- The Origin of the Data
- The obligation level
- Any comments that would be necessary.

ELEMENTS AND SUB-ELEMENTS

The following list of Metadata Categories, Elements and Sub-Elements are the recommended elements that should be considered for use in the Document Imaging Environment.

4.1.1	ADMINISTRATIVE METADATA	Authentication
	RIGHTS AND SECURITY	Encryption
	INFORMATION	Validation
	RESTRICTIONS	Watermarking
	Terms, License, Use	
	Access Restrictions	GATHERING INFORMATION
	Rights Restrictions	Frequency
	COPYRIGHT	Regime
	Owner	Accruals
	Warnings	
	Conditions	METADATA INFORMATION
	Access Controls	Creator
	RIGHTS	Description
	Negotiation Status	Date created and updated
	Legal Status	
	Retention Period	NOTES AND COMMENTS
	Resource Rights	Free text Notes
	Rules	
	SECURITY	
	Access Inhibitors	4.1.2 DISCOVERY OR DESCRIPTION

CORE METADATA	
CONTRIBUTOR	Census
Individual	Death Records
Library	Birth Records
Archive	Marriage Records
Other	Commissary Court Records
	Etc...
COVERAGE	TYPE
Time	Text
Location	Image
Content - Note	Data
CREATOR / AUTHOR	Database
Individual	Software
Library	Physical Object
Archive	Sound
Other	Video
DATE	Multimedia
Created	Notes
Available	RIGHTS STATEMENT
Issued	Rights Statement (optional)
Modified	NOTES AND COMMENTS
Restoration	Free text Notes
Beginning	
Completion	
DESCRIPTION	4.1.3 PRESERVATION INFORMATION
Open Text Field - Notes	IDENTIFIERS
Defined Category MARC	Persistent Identifiers
FORMAT (See also Preservation)	Version or Position
IDENTIFIER (PERSISTENT)	Use Information for Patron
Project Name	Title
Project Number	Level of Granularity
Current Document ID	Culture or Period
Former Document ID	Genre
Current Object ID	Location
Former Object ID	HISTORY
Unique Number Assignment	Preservation History of Object
Link with external Catalog	Preservation History of Change
ENTITY IDENTIFIER TITLE	Process
Working Title	Preservation History Copies
Official Title	Agencies Responsibilities for
LANGUAGE	Preservation
Language	FORMAT
PUBLISHER	Digital Storage Formats
Individual	Human Readable (Y/N)
Library	HARDWARE
Archive	Name
Other	Details
RELATION	Modality
Relationship between works	SOFTWARE
Reference to another work	Name
Identification from which the	Critical Software
digital object was created	Modality
SUBJECT	QUIRKS
Key-word describing resource	Quirks

4.1.4 TECHNICAL METADATA

CAPTURE

- Bit Depth (Tonal Resolution)
- Color Space / Management
- Color Bar, Grayscale Bar
- Scanner Profile
- Compression - Uncompression
- File Type and Version
- Procedures
- Dimensions
- Orientation
- Capture Details
- Resolution
- Targets
- Reduction
- Film Size
- File Equivalents
- File Description Format
- Version
- Image Size
- Number of Files
- Operator Information
- Filming Location
- History

CHARACTER SETS

- Text Character Sets
- Character encoding

FINDING AND SEARCHING AIDS

- Open Text

HARDWARE

- Name
- Details
- Modality

SOFTWARE

- Name
- Critical Software
- Modality

SYSTEM

- Dependencies
- Requirements

NOTES AND COMMENTS

- Free text Notes

4.1.4.1 EVALUATION AND QUALITY CONTROL

EVALUATION / QUALITY

- Evaluator / Auditor Name / No.
- Operator / Scanner ID No.
- Location of Evaluator
- Audit Session Start Time

Audit Session Stop Time

Audit Session Date

Image File Status (Status of the Image)

EVALUATOR FOLDER ID (Where Image was pulled from)

Evaluator Audit Lot ID

Evaluator Image File ID

Evaluator Image Pass / Fail Status

Evaluator Image Pass / Fail Code

AUDIT FOLDER ID (Where image was pulled from)

Auditor Audit Lot ID

Auditor image File ID

Auditor Image Pass / Fail Status

Auditor Image Pass / Fail Code

4.1.5

USE

IMAGE (OBJECT) VIEW CATEGORY

VIEW CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)

PRINT CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)

DOWNLOAD CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)

METADATA / INDEX VIEW CATEGORY

VIEW CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)

PRINT CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)

DOWNLOAD CATEGORY

- Not Available

- Library / Archive Only

- Intranet (Other Selective Libraries)

- Internet (Available to Public)