

Kentucky Digital Library Metadata Guidelines

Update June 2013

1 Introduction

The Kentucky Digital Library (KDL) is built to enhance scholarship, research and lifelong learning through the establishment of access to shared digital archival collections in the state of Kentucky. It also provides guidance and instruction for Kentucky libraries, archives, historical societies and museums on applying appropriate technologies used in the production of digital library resources.

- from the mission statement

The Kentucky Digital Library is a statewide resource whose infrastructure is managed by the University of Kentucky's Digital Library Services. KDL also serves as regional hub for the Digital Public Library of America. These guidelines are intended to support the aim of providing technological guidance to the diversity of participating cultural heritage institutions whose unique collections comprise the KDL. In turn, well-formed and consistent metadata will support effective long-term preservation, enhance discovery and improve interoperability. Ultimately, good metadata helps users by making sure Kentucky's digitized cultural heritage collections are accessible and searchable, now and in the future.

Specifically, these guidelines are designed to:

cover basic metadata principles and best practices

increase transparency by describing KDL's workflows and processes regarding metadata

provide specifications and practical examples for descriptive metadata fields

facilitate effective reuse of existing descriptions already created by contributors, such as cataloging records and finding aids

1.1 Metadata Overview

Robust models for describing digital objects categorize metadata as either descriptive, administrative and structural. While useful, these categories can overlap in several areas if they are not further defined. KDL generates much of the administrative and structural metadata describing objects and collections through automated processes. On the other hand, KDL believes contributors know their own collections best, and KDL depends on them to provide descriptive metadata that will best facilitate discovery. At the same time, contributed metadata should adhere to best practices where at all possible in order to make the most of KDL's existing infrastructure and any systems that rely on KDL's data, such as the Digital Public Library of America (DPLA).

These guidelines are designed in the effort to balance the diversity of descriptions associated with contributed collections and the technical needs of systems that rely on KDL data, such as the DPLA. In practice, KDL largely consists of material contributed by special collections and archives and has made every effort to incorporate aspects of standards appropriate to those collections, especially Description and Arrangement Content Standards (DACS) and Encoded Archival Description (EAD). KDL recognizes three levels of "required-ness" across descriptive metadata fields -- required, required if available, recommended and collection level.

Required fields must be submitted in order for content to be ingested:

title
format
resource type
language
rights
Kentucky-specific subject

Required if available

creator

date

source

Recommended

KDL describes a number of additional fields that map to common description and cataloging standards. We highly recommend that contributors submit all fields appropriate to existing descriptive metadata that may have previously been collected in finding aid, accession record or catalog record by contributing institution.

Collection level

KDL provides a number of recommended fields specific to collection level records in order to capture information contained in finding aids that may not be appropriate for item level records.

1.2 Clarifications

Items and Collections

There are two types of records that may be entered: those that describe individual items and those that describe collections of items. Fields that only pertain to collections appear in the recommended fields section.

Source Material

In the case of digitized collections as opposed to born-digital materials, the metadata fields as they are construed here privilege the source material. While the fields descriptions in the following section should make that explicit, occasionally ambiguity may arise, in which case contributors should opt for descriptions of the original.

Character Encoding

KDL encodes metadata records using UTF-8 character encoding, which supports a range of foreign characters. We encourage contributors to adhere as closely to information as it appears on the original record, for example, including accented characters in the title of records or the names of contributors.

Dates

To enhance discoverability, KDL requires the use of ISO 8601 for the entry of the primary data field. This takes the form of YYYY-MM-DD. We encourage the use of the same format in every other field in which a data might appear, including titles, creators

Authority Lists and Controlled Vocabulary

Wherever possible we support authority lists and controlled vocabulary over free-text submission. A list of resources mentioned in this guidelines appears in the Appendix B.

2 Data Fields

2.1 How to Use this Manual

Descriptive metadata fields submitted by contributors appear below with the following fields:

Title: human-readable title of the field

Repeatable: whether the field can be used multiple times or - whether it is unique to an item

Scheme: how data should be inputted, whether entering free - unstructured text or entering values according to an authority list

Qualifier: whether qualifiers are available to distinguish field - elements

Definition: a succinct definition of the field's intended use, - largely taken from Dublin Core guidelines but further specified for use by KDL

Usage Guidelines: notes on how to use the field and issues that may come up in practice

Examples: examples taken from the KDL of successful fields

Resources: links to external resources relevant to the field

Note that crosswalks for each field appear in Appendix A.

2.2 Required Fields

Account Information

In addition to descriptive metadata fields, KDL requires repositories to provide basic account information in order to facilitate contributors editing metadata and uploading content.

The required information for repository accounts are:

Abstract: A succinct description of the contributing repository to appear online alongside contributed materials

Department: The name of the institutional department in custody of the collections published online

Institutional Contact Information to appear alongside online collections

Street Address: Physical location of the contributed collections

Email: An institutional email address for the repository for directing reference requests and other collection-specific queries

URL: A stable web address linking to the contributing institution

Depositor Contact Information for internal use should KDL need to contact the person responsible for contributing material

Depositor name

Depositor email address

Title

Repeatable: Yes

Scheme: Free text

Qualifier(s): Alternate

Definition: An identifying and/or descriptive name given to the resource

Usage Guidelines:

Input the title exactly as it appears as it appears, including punctuation and foreign characters.

The title should be the name by which the resource is formally known, typically given by the creator.

If no title exists, please provide a succint description of the resource, if possible constructed according to DACS. Only use "Untitled" in cases where that is the designated title.

If more than one title exists, additional title fields should be designated by the qualifier "alternate".

Examples:

Formal / "Why Blood Flows in Harlan," The Literary Digest, December 5, 1931

Supplied / Group portrait of Elmendorf students, with teacher Ella Ford, standing in front of school located off Paris Pike.

Resources:

Describing Archives: A Content Standard (http://files.archivists.org/pubs/DACS2E-2013.pdf)

Format

Repeatable: Yes

Scheme: AAT

Definition: The file format or physical medium of the present resource

Usage Guidelines:

When using the metadata editor, Format choices will appear as options in a selectable drop down menu.

KDL uses a custom subset of Getty AAT. See Resources below.

In the case of digitized resources, Format refers to the analog source.

Examples: Books, Diaries, [mixed materials]

Resources:

KDL Format Authority List

(http://www.uky.edu/~eweig/kdl_metadata_2013/kdl_metadata_2013_dcmiformat.xsd)

Getty AAT (http://www.getty.edu/research/tools/vocabularies/aat/index.html)

Resource Type

Repeatable: Yes

Scheme: DCMI Type Vocabulary

Definition: The functions, genres or aggregation levels for the content of the resource

Usage Guidelines:

When using the metadata editor, Resource Type choices will appear as options in a selectable drop down menu.

The Resource Type corresponds to those described by the Dublin Core Type Vocabulary.

Examples: Collection, Text, StillImage

Resources:

DCMI Type Vocabulary (http://dublincore.org/documents/dcmi-type-vocabulary/#H7)

Kentucky-Specific Subject

Repeatable: Yes

Scheme: KDL Kentucky Subject Headings

Definition: Topic headings related to Kentucky history that describe the intellectual

content of the resource

Usage Guidelines:

Select Kentucky specific topic(s) as indicated on KDL's Kentucky subject headings list.

If using the metadata editor, these options appear in a drop down list.

Examples: African American history, Appalachia, Coal Mining

Resources:

KDL Kentucky Subject Headings (http://www.uky.edu/~eweig/kdl_metadata_2013/kdl_metadata_2013_dcmisubject.xsd)

Language

Repeatable: Yes

Scheme: ISO 639

Definition: The language of the intellectual content of the present resource

Usage Guidelines:

When using the metadata editor, Language choices will appear as options in a selectable drop down menu.

The Language should be indicated by the two-letter code where available or the three-letter code, if the two-letter is unavailable.

Examples: en, es

Resources: ISO 639 (http://www.loc.gov/standards/iso639-2/php/code_list.php)

Rights

Repeatable: Yes

Scheme: Free text

Definition: An identifier that links to a rights management statement or to a service providing information about rights management

Usage Guidelines:

Contributors can include a statement and/or a URL appropriate to that institution's policies.

If no institution-specific rights statement is included, KDL's default rights statement will be used.

Examples:

Default / This digital resource may be freely searched and displayed. Permission must be received for subsequent distribution in print or electronically. Physical rights are retained by the owning repository. Copyright is retained in accordance with U. S. copyright laws.

Please go to http://kdl.kyvl.org for more information.

2.3 Required If Available

Creator

Repeatable: Yes

Scheme: LC Name Authority Headings (recommended), Free text

Definition: The person, agency or organization primarily responsible for creating the

intellectual content of the resource

Usage Guidelines:

In the case of an individual, begin with the creator's last name followed by the first, separated by a comma.

In case of ambiguity, include the name exactly as it appears in the resource.

Include birth and death dates where applicable.

In the case of corporate entities, use the name exactly as it appears in the resource.

Wherever possible, use the form of names as they appear in the Library of Congress Name Authority Headings.

Examples:

Individual Creator / Brezing, Belle, 1860-1940

Corporate Creator / University of Kentucky. College of Agriculture. Agricultural Experiment Station

Resources: Library of Congress Name Authority Headings (http://authorities.loc.gov/)

Date

Repeatable: Yes

Scheme: ISO 8601 (YYYY-MM-DD)

Definition: The date when the present resource was created, or in the case of digitized objects, when the analog source was created

Usage Guidelines:

When using the metadata editor, this field will be entered using the drop down calendar.

When only a year is available, please format the date [year]-01-01.

When an item is described by a date span, please enter the earliest date.

Items require a single date in order to facilitate effective, faceted search across digital collections, which may give rise to ambiguity. Other date fields that appear under Recommended Fields can be used to further describe items.

Examples: 1848-01-01

Resources:

ISO 8601 (http://www.w3.org/TR/NOTE-datetime)

Source

Repeatable: No

Scheme: A formal identifier (recommended), Free text

Definition: An unambiguous reference to the resource from which the present resource is derived

Usage Guidelines:

In practice, the source is often the title for the collection of which the present resource is a member.

The source may also be a formal identifier (e.g., URI, ISBN, ISSN or accession number).

The source is not applicable to most born-digital objects, that is, objects that are not derivatives of other resources.

Examples:

Collection title / Herndon J. Evans Collection

Formal / ISBN: 0613685725, Accession Number: 2009MS189

2.4 Recommended

LCSH Subject

Repeatable: Yes

Scheme: LCSH, TGM I, TGM II, Free text

Definition: The subjects, topics or keywords describing the intellectual content of the

resource

Usage Guidelines:

In practice, descriptive metadata fields can be used as subjects. The values entered for Creator, Contributor, Spatial Coverage and/or Temporal Coverage can also be entered as values for the Subject field.

Where appropriate, include subjects from the following authority lists: LCSH, TGM I and TGM II. Ideally, these subject strings can be reused from existing records such as finding aids or MARC records.

Examples:

Creator as subject /

Brezing, Belle, 1860-1940

University of Kentucky. College of Agriculture. Agricultural Experiment Station

LCSH / Botany--Kentucky--Lexington

Entomology--Kentucky--Lexington

Resources:

Library of Congress Subject Authority Headings (LCSH) (http://authorities.loc.gov/)

Thesaurus for Graphic Materials I (TGM I) (http://www.loc.gov/rr/print/tgm1/)

Thesaurus for Graphic Materials II (TGM II) (http://www.loc.gov/rr/print/tgm2/)

Publisher

Repeatable: Yes

Scheme: LCSH, TGM I, TGM II, KDL Kentucky Topics, Free text

Definition: The subjects, topics or keywords describing the intellectual content of the

resource

Usage Guidelines:

In practice, descriptive metadata fields can be used as subjects. The values entered for Creator, Contributor, Spatial Coverage and/or Temporal Coverage can also be entered as values for the Subject field.

Also include Kentucky specific topic(s) as indicated on KDL's Kentucky subject headings list.

Where appropriate, include subjects from the following authority lists: LCSH, TGM I and TGM II. Ideally, these subject strings can be reused from exsiting records such as finding aids or MARC records.

Examples:

Creator as subject /

Brezing, Belle, 1860-1940

University of Kentucky. College of Agriculture. Agricultural Experiment Station

LCSH / Botany--Kentucky--Lexington

Entomology--Kentucky--Lexington

Resources:

KDL Kentucky Subject Headings

(http://www.uky.edu/~eweig/kdl_metadata_2013/kdl_metadata_2013_dcmisubject.xsd)

Library of Congress Subject Authority Headings (LCSH) (http://authorities.loc.gov/)

Thesaurus for Graphic Materials I (TGM I) (http://www.loc.gov/rr/print/tgm1/)

Thesaurus for Graphic Materials II (TGM II) (http://www.loc.gov/rr/print/tgm2/)

Contributor

Repeatable: Yes

Scheme: LC Name Authority Headings (recommended), Free text

Definition: The person, agency or organization that contributed significantly or played an important role in creating the intellectual content of the resource other than the creator

Usage Guidelines:

In the case of an individual, begin with the creator's last name followed by the first, separated by a comma.

In case of ambiguity, include the name exactly as it appears in the resource.

Include birth and death dates where applicable.

In the case of corporate entities, use the name exactly as it appears in the resource.

Wherever possible, use the form of names as they appear in the Library of Congress Name Authority Headings.

Examples:

Individual Contributor / Brezing, Belle, 1860-1940

Corporate Contributor / University of Kentucky. College of Agriculture. Agricultural Experiment Station

Resources: Library of Congress Name Authority Headings (http://authorities.loc.gov/)

Collection Label

Repeatable: No

Scheme: Free text

Definition: The human-readable label for the collection of which the resource is a part.

Usage Guidelines:

This field is provided as a convenience to contributing repositories to allow users to

identify an item as part of a collection especially when a formal identifier such as an accession number or ISBN is unavailable.

Use the title of the collection as it appears within internal catalog of the contributing repository.

Examples: Abraham Haptonstall Legal Documents, 1788-1806, Abe Thompson photograph album

Container 1

Repeatable: No

Scheme: Controlled vocabulary, Integer identifier

Definition: The highest level container of a collection's physical arrangement and its

sequential number

Usage Guidelines:

This field allows item level records to indicate a physical location within the contributing institution's holdings.

When using the metadata editor, options appear in a drop down menu.

The optional fields Container 2 and Container 3 describe nested containers, further specifying the physical location of source records.

This field can facilitate the construction of an EAD record from descriptive metadata.

Examples: Box 1, Carton 3, Drawer 6

Resources:

Encoded Archival Description (http://www.loc.gov/ead/tglib/element_index.html)

Container 2

Repeatable: No

Scheme: Controlled vocabulary, Integer identifier

Definition: The second highest level container of a collection's physical arrangement and its sequential number

Usage Guidelines: Please see Usage Guidelines for Container 1.

Examples: Folder 2, Case 3

Container 3

Repeatable: No

Scheme: Controlled vocabulary, Integer identifier

Definition: The third highest level container of a collection's physical arrangement and its

sequential number

Usage Guidelines: Please see Usage Guidelines for Container 1.

Examples: Item 5

Description

Repeatable: Yes

Scheme: Free text

Definition: An account of the physical or intellectual properties of the present resource

Usage Guidelines:

This field allows contributors to include a succinct, relevant summary or abstract that will aid search and discovery of the present resource.

The resources below link to content standards that describe appropriate structures for descriptive and summative notes.

Examples:

Resources:

Encoded Archival Description (http://www.loc.gov/ead/tglib/element_index.html)

Describing Archives: A Content Standard (http://files.archivists.org/pubs/DACS2E-2013.pdf)

Relation

Repeatable: Yes

Scheme: URI (recommended), ARK identifier, Accession number, ISBN

Qualifier(s): Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, Has Format

Definition: A formal string identifier for another resource such as collection or item to which the present resource is related

Usage Guidelines:

This field is intended to provide a way of indicating a machine readable reference to another resource.

Requires the use of a Dublin Core Relation qualifier to specify the relationship between the resources.

Examples: isPartOf: ark:/xt7ht727b08t, isVersionOf: ISBN 0613685725

Resources:

Dublin Core Terms: Relation (http://purl.org/dc/terms/relation)

Collection Level Relation

Repeatable: No

Scheme: URI (recommended), ARK identifier, Accession number, Free text

Definition: A formal string identifier for another resource such as collection or item to which the present resource is related

Usage Guidelines:

This field is intended to provide a way of indicating a machine readable reference to

another resource, specifically that it is part of a collection.

When using the metadata editor, options appear in a drop down menu based on the collection level records submitted for the contributing repository.

Examples: ark:/xt7ht727b08t, Accession Number: 1984ua004

Resources:

Dublin Core Terms: isPartOf (http://purl.org/dc/terms/isPartOf)

Geographic Coverage

Repeatable: Yes

Scheme: Controlled vocabulary

Definition: Kentucky specific place names relevant to the intellectual content of the

present resource

Usage Guidelines:

Choose relevant geographic locations from the subset list of the LCSH subject headings referring to Kentucky's political boundaries.

When using the metadata editor, options appear in a drop down menu.

Examples: World -- North and Central America -- United States -- Kentucky -- Lexington

Resources: KDL Kentucky Place Names

(http://www.uky.edu/~eweig/kdl metadata 2013/kdl metadata 2013 dcmicoverage.xsd)

Retention

Repeatable: No

INO

Scheme: Controlled vocabulary

Definition: The retention schedule of the present resource

Usage Guidelines:

KDL's preservation infrastructure involves different maintenance routines depending on a record's retention schedule.

If using the metadata editor, supported retention schedules appear in a drop down list.

Resources: KDL Retention Schedule List

(http://www.uky.edu/~eweig/kdl_metadata_2013/kdl_metadata_2013_retention.xsd)

Terminal Date

Repeatable: No

Scheme: ISO 8601 (YYYY-MM-DD)

Definition: The calendar date after which a digital record is intended to be deaccessioned

Usage Guidelines:

When using the metadata editor, this field will be entered using the drop down calendar.

Resources: ISO 8601 ([ISO%208601](http://www.w3.org/TR/NOTE-datetime)

Spatial Coverage

Repeatable: Yes

Scheme: Geographical coordinates

Definition: Geospatial coordinates relevant to the intellectual content of the present

resource

Usag		

Examples:

Resources:

File Format

Repeatable: Yes

Scheme: Controlled vocabulary

Definition: The file format of the present digital resource

Usage Guidelines: When using the metadata editor, this field will be entered using the

drop down list of file formats based on IANA.

Examples:

Resources:

KDL File Formats

(http://www.uky.edu/~eweig/kdl_metadata_2013/kdl_metadata_2013_fileformat.xsd)

Internet Assigned Numbers Authority (IANA) (http://www.iana.org/assignments/mediatypes)

Series Statement

Repeatable: No

Scheme: Free text

Definition: A label designating the intellectual series of an archival collection to which the

present belongs

Usage Guidelines:

Enter the series title exactly as it appears within the contributing institutions internal catalog.

This field enhances discoverability by providing intellectual context to item level records.

Examples: Correspondence, Financial documents, Ephemera

2.3 Collection Level Fields

The following fields refer to finding aid specific descriptive information at the collection level. The fields are largely derived from the best practices described by DACS (http://files.archivists.org/pubs/DACS2E-2013.pdf) and EAD

(http://www.loc.gov/ead/tglib/element_index.html). Where indicated, the following definitions are summarized from the Encoded Archival Description Tag Library, 2002.

Scope and Content Note

A prose statement summarizing the range and topical coverage of the described materials, often mentioning the form and arrangement of the materials and naming significant organizations, individuals, events, places, and subjects represented.

Resources: EAD <scopecontent>

(http://www.loc.gov/ead/tglib/elements/scopecontent.html)

Biography or History

A concise essay or chronology that places the archival materials in context by providing information about their creator(s).

Resources: EAD

| EAD

| County | Count

Arrangement Note

Information on how the described materials have been subdivided into smaller units, e.g., record groups into series, identifying the logical or physical groupings within a hierarchical structure.

Resources: EAD <arrangement>

(http://www.loc.gov/ead/tglib/elements/arrangement.html)

Extent

Information about the quantity of the materials being described or an expression of the physical space they occupy. In practice, this usually involves traditional archival units of measurement such as linear or cubic feet, boxes, cartons or cases. It may also include the number of individual items, such as photographs or film reels.

Resources: EAD <extent> (http://www.loc.gov/ead/tglib/elements/extent.html)

Abstract

A very brief summary of the materials being described, used primarily to encode bits of biographical or historical information about the creator and abridged statements about the scope, content, arrangement, or other descriptive details about the archival unit or one of its components.

Resources: EAD <abstract> (http://www.loc.gov/ead/tglib/elements/abstract.html)

Date Range Start and Date Range End

The date of the earliest record's creation and the latest record's creation. As with other dates, these should take the form of YYYY-MM-DD according to ISO 8601.

Bulk Date

The date or dates describing the predominant date range during which records in a collection were created. This is useful when there is a significant discrepancy between the earliest or latest records and the bulk of record creation.

This date is entered as free text and can be a date (YYYY-MM-DD), a year (YYYY) or span of years (YYYY-YYYY).

Custodial History

Information about the chain of ownership of the materials being described, before they reached the immediate source of acquisition.

Resources: EAD <custodhist> (http://www.loc.gov/ead/tglib/elements/custodhist.html)

Source of Acquisition

The immediate source of the materials being described and the circumstances under which they were received. Includes donations, transfers, purchases, and deposits.

Resources: EAD <acqinfo> (http://www.loc.gov/ead/tglib/elements/acqinfo.html)

Physical Access

Information identifying the place where the described materials are stored, such as the name or number of the building, room, stack, shelf, or other tangible area.

Resources: EAD <physloc> (http://www.loc.gov/ead/tglib/elements/physloc.html)

Technical Access

A description of important physical conditions or characteristics that affect the storage, preservation, or use of the materials described.

Resources: EAD <phystech> (http://www.loc.gov/ead/tglib/elements/phystech.html)

Condition Governing Reproduction and Use

Information about conditions that affect the availability of the materials being described.

Resources: EAD <accessrestrict>

(http://www.loc.gov/ead/tglib/elements/accessrestrict.html)

Accruals

Information about anticipated additions to the materials being described. Can indicate quantity and frequency. Can also be used to indicate that no additions are expected.

Resources: EAD <accruals> (http://www.loc.gov/ead/tglib/elements/accruals.html)

Note

A short statement explaining the text, indicating the basis for an assertion, or citing the source of a quotation or other information. Used both for general comments and as an annotation for the text in a finding aid. Should not be used in cases where a more specific tag would suffice.

Resources: EAD <note> (http://www.loc.gov/ead/tglib/elements/note.html)

Appendix A: Ingest Processes Overview

Once a project is actively being processed, digital objects undergo a number of steps so that when files are later retrieved KDL can ensure they are complete and have not been unintentionally modified. KDL models its system after the Open Archives Initiatives System (OAIS) which, roughly speaking, divides preservation into three stages: submission, archiving and dissemination. At each stage, a packet of information including metadata and data files is generated and passed onto the next for further processing.

1 Submission

Data files, paired with basic descriptive information such as format, are put into an ingest queue, something like a bulletin board listing upcoming projects.

2 SIP

The first stage of processing is creating a Submission Information Packet (SIP). Files are copied into a workspace and associated with appropriate administrative and structural metadata. Files might be further processed, for example, images of printed text might be scanned using Optical Text Recognition (OCR).

3 Identifiers

The data, metadata and any derivative files are then packaged and given unique identifiers. The Archival Information Packet(AIP) identifier will accompany the package of information that will be archived and stored without being accessed. The Dissemination Information Packet(DIP) identifier will accompany the package of information that will be stored on the server accessed by KDL's online interface.

4 Approval

Before being archived or moved to access storage, packages undergo testing to make sure that files and metadata are complete.

5 Storage

Lastly, packages are moved to their appropriate servers and a version of the metadata is added to KDL's index. Archival packages are secure within a preservation repository and dissemination packages are now searchable online. Any temporary files are cleaned up and the process can begin again.

Appendix B: Crosswalks