

<copyright>



copyrightMD User Guidelines



<copyright>

Version 0.91

California Digital Library
September 2009

1 INTRODUCTION

This document provides implementation guidelines for the copyrightMD XML (Extensible Markup Language) schema. copyrightMD defines a set of metadata elements that can be used to express copyright information for digital resources. It is compatible with and can be used as an extender to the Metadata Encoding and Transmission Standard (METS).

1.1. Background

In 2004, the California Digital Library (CDL) formed a short-term Rights Management Group (RMG) to advise on issues concerning copyright protections and fair use. From 2005 through 2006, the RMG did an analysis of the functional requirements related to copyright metadata, identified key data elements for expressing copyright metadata, and formalized these elements in the copyrightMD XML schema.

1.2. Functional Objectives

There are three primary functional objectives of the copyrightMD schema:

1. To express factual information relating to copyright that allows users to make an informed copyright assessment of a given work.
2. To connect users to sources of further information about the copyright status, or to a person or institution that can be a resource for obtaining permissions relating to copyright-protected uses of the resource.
3. To explicitly associate item-level copyright information with discrete digital objects, which enriches digital objects with copyright metadata at the point of creation, thereby preventing the creation of "orphaned works" in the future.

Explicit copyright assertions alone give users some idea of the presumed copyright status of the work, but do not take the place of factual copyright metadata. The copyrightMD schema is therefore user-oriented, and not intended to serve as a recordkeeping mechanism for other copyright-related information collected and gathered by institutions for internal management (such as donor agreement information, copyright permission request histories, etc.).

One of the chief characteristics of digital resources, whether digitally reformatted or born digital, is the ease with which they can be re-used either in whole or in part. Reuse rights are determined based on the copyright status of an item, as well as the nature of the intended use. It is incumbent on the users of materials to assess the copyright status of an item and to understand whether use is permitted by law, or if it may be necessary to obtain permission. To make this copyright assessment, users need to have certain information about the item. In the case of digitally reformatted materials in particular, this information may be inherent in the context of the archive or collection in which the physical items reside. However, the digitization of works often removes that context, and connections between the surrogate and the original work may weaken

over time. It is therefore necessary that information that allows the user to make the best assessment of copyright status be bundled with digital objects. Unfortunately, the recording of copyright-related information is often not part of the tradition of cataloging or metadata creation in libraries, museums, and archives.

An important potential digital library service is to provide the user with all available information relating to the copyright status of the work. Current thinking is that this type of contact information will be provided through the user interface. Additionally, systems could also be configured to allow users to scope searches and browses to collections with particular copyright statuses. For example, a search could be limited to known "public domain" works, or to materials where the rights holder and contact information are known.

2 Implementation Notes

copyrightMD is expressed using the XML schema language of the World Wide Web Consortium. By using the XML schema language, copyrightMD defines main elements, child elements (i.e. subelements), and attributes of elements and subelements. Content of elements is included in the lowest level elements so as to avoid "mixed content", or character data interspersed with subelements. For example, the element <rights.holder> contains the subelements <name>, <contact>, and <note>, making it a wrapper element that does not contain any character data. Information about the rights holder can only be added within the subelements <name>, <contact>, and <note>.

Attributes serve to modify elements and may be associated with elements at any level. Attributes are defined with the element(s) they modify as part of the XML schema.

A copyrightMD document contains a schema declaration that indicates the copyrightMD namespace. Within a record or group of records it is optional to use the "copyrightMD" prefix before each element (and before the "copyrightMD" namespace declaration), since the copyrightMD namespace is indicated in the record. It is most useful to use the prefix "copyrightMD:" before each element when combining a copyrightMD record with XML data from another namespace, e.g. a copyrightMD record within a METS document.

The schema declaration for copyrightMD Version 0.9:

```
<copyright copyright.status="" publication.status=""
xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
```

2.1. Use of Content Standards

Any content standard or cataloging guideline may be used with relevant copyrightMD elements (e.g., for formatting names or dates). Institutions are encouraged to apply content standards consistently across records and document their usage.

2.2. Required Elements

The copyrightMD schema reflects the full suite of elements that could be utilized in any given record. A valid copyrightMD record requires only a single element, the top-level <copyright> element. Applications may develop profiles specifying mandatory elements as needed.

2.3. Element Repeatability

Repeatability of all copyrightMD elements is indicated in the element descriptions in section 3.3 below.

The copyrightMD data elements may be applied at any level of description, and should be used at the level that is most appropriate to document the copyright status of a resource. This may be at an aggregate level or at an item level, depending on the degree to which the components of the resource share given copyright information. In some cases, the copyright data elements could be used at the component level for a particular item; for example, in the case of a photograph album comprising images created by different photographers, each page of the album may need to reflect a different copyright status. The mechanics of applying copyrightMD elements will, of course, depend on the system you use for managing digital objects (e.g., METS).

2.4. Order of Elements

Within a single copyrightMD record instance, the schema does not constrain the order in which the elements and subelements are encoded.

The order of encoding elements in a record created using the copyrightMD schema does not assume display order. A stylesheet is used to control the order of information presented in a display to end users, as noted in section 2.8 below.

2.5. Unknown Data Values

copyrightMD elements can be coded as "unknown" when that is the appropriate data value for a specific situation. This alerts users to the fact that they will not be able to rely on the provided metadata to aid in copyright determination, and suggests that they should contact your institution (or the appropriate services contact) for further information. Be certain to distinguish between indicating the value "unknown" (meaning the information is truly unknown), and indicating the value "no attempt to code" (meaning that information may or may not be available, but no attempt was made to encode it in the record).

2.6. Character Encoding

Use ISO-8859-1, UTF-8, or UTF-16 standard character sets or encodings, with a preference for UTF-8, which is the normal character set used in XML encoding. Use standardized forms of names for character sets, as documented by the Internet Assigned Numbers Authority (e.g., use "UTF-8" and not "UTF8").

Characters reserved for XML markup delimiters (ampersand, left angle bracket, and right angle bracket) need to be replaced with character entities.

2.7. Headings, Labels, Formatting, and Punctuation

Do not include HTML formatting for headings, labels, line breaks, list formatting, or other display encoding within copyrightMD elements. Punctuation within elements is optional. Institutions are encouraged to apply punctuation consistently across records and document that usage. Remember that purely formatting characters—such as tabs, carriage returns, and excessive spacing—may or may not be honored by output systems.

2.8. Display and Output

XSLT (Extensible Stylesheet Language Transformation) stylesheets may be used to transform the copyrightMD data in some way for output. For example, an institution might use a stylesheet to place information from a copyrightMD record into a display for end users that contains easy-to-understand labels in HTML, or might use a stylesheet to transform information encoded as attribute values in a copyrightMD record into textual form for the purposes of display.

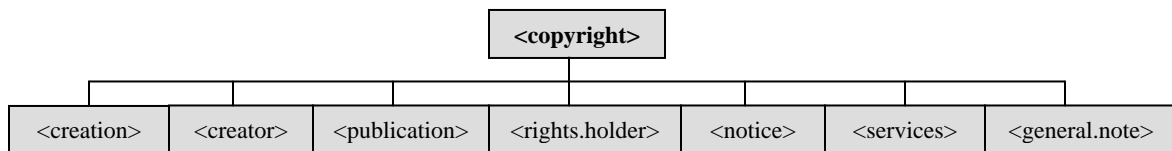
3

Detailed Description of copyrightMD

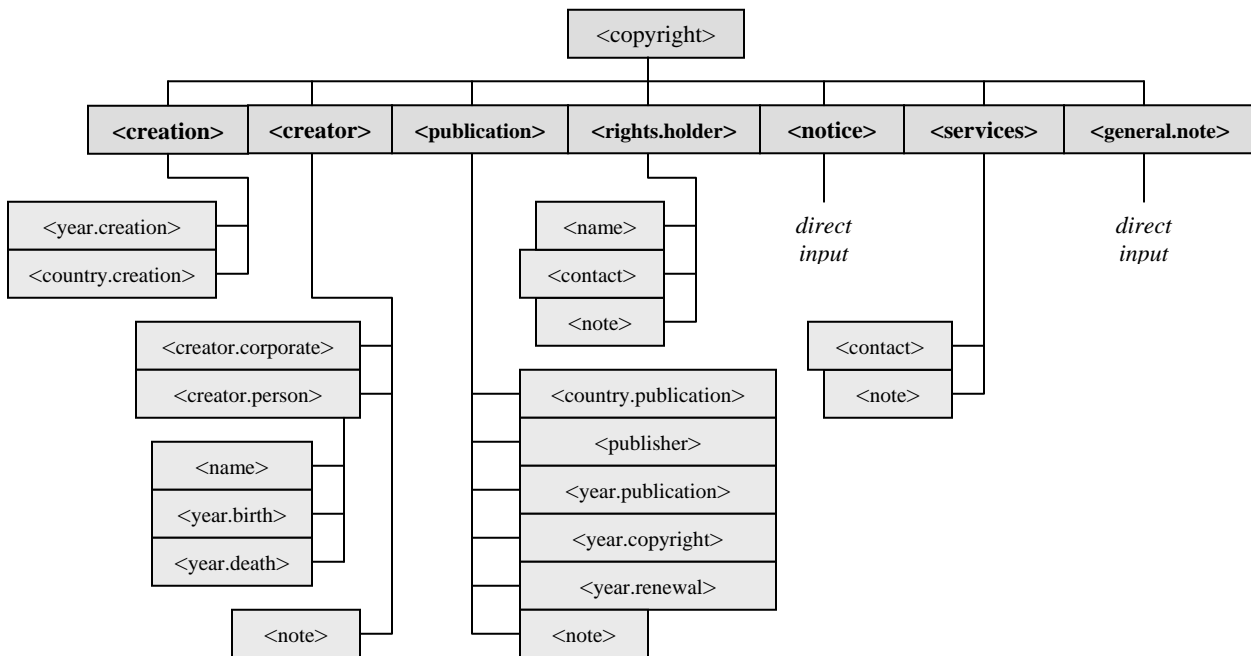
3.1. General Application Guidelines

At a minimum, in order to validate, a copyrightMD record must include, as part of the `<copyright>` element, encoded values for the attributes *copyright.status* and *publication.status*. Use of all of the remaining copyrightMD elements is optional and will be determined by a combination of institutional policies, workflows and staffing, and the types of material being described.

Within the `<copyright>` element there are seven (7) available blocks of subelements:



Each of these blocks either contains additional subelements to facilitate encoding of important information relating to copyright, or allows encoding of such information directly in the highest-level element in the block. Available subelements for each block are indicated below, with additional application information supplied in section 3.3 of these guidelines.



Decisions regarding the appropriate blocks to use in any given encoding situation should be informed by the choices made in encoding the required attributes *copyright.status* and *publication.status* within the <copyright> element. The following are some examples of this alignment. These examples illustrate how copyrightMD blocks might be used and are not intended to be either exhaustive or prescriptive. Institutions are encouraged to implement the copyrightMD schema consistently for each scenario encountered, and to document that implementation.

Example 1:

```
<copyright copyright.status="copyrighted" publication.status="published">
```

- At minimum, consider providing additional information using the <publication>, <rights.holder>, and <notice> blocks.

Example 2:

```
<copyright copyright.status="copyrighted" publication.status="unpublished">
```

- At minimum, consider providing additional information using the <creation>, <creator>, and <rights.holder> blocks.

Example 3:

```
<copyright copyright.status="pd" publication.status="published">
```

- At minimum, consider providing additional information using the <publication> and <notice> blocks.

Example 4:

```
<copyright copyright.status="unknown" publication.status="unknown">
```

- At minimum, consider providing as much additional information as is available using the <creation>, <creator>, and <publication> blocks.

3.2. Alphabetical List of Attributes

copyright.status

Availability: <copyright>

Required: Yes

Definition: Indicates copyright status for the resource.

Application: Six (6) enumerated values enforced by the copyrightMD schema:

- *copyrighted* - Under copyright.
- *pd* - Public domain: No further information.
- *pd_usfed* - Public domain: US Federal document.
- *pd_holder* - Public domain: Item dedicated to the public domain by the rights holder.
- *pd_expired* - Public domain: Item in the public domain because of expiration of copyright based on U.S. law.
- *unknown* - Copyright status of the resource is unknown.

iso.code

Availability: <country.publication>, <country.creation>

Required: No

Definition: Specifies a standardized code for the name of the country of creation or publication.

Application: Data values should be taken from the ISO 3166-1 alpha 2-letter code list.¹

publication.status

Availability: <copyright>

Required: Yes

Definition: Indicates the publication status for the resource.

Application: Three (3) enumerated values enforced by the copyrightMD schema:

- *published* - Resource is published.
- *unpublished* - Resource is unpublished.
- *unknown* - Publication status of the resource is unknown.

year.type

Availability: <year.copyright>, <year.renewal>, <year.publication>, <year.creation>, <year.birth>, <year.death>

Required: No

Definition: Provides additional information about the year encoded.

Application: Three (3) enumerated values enforced by the copyrightMD schema:

- *exact* - Year is known.
- *approximate* - Year is unknown, but it is possible for a reasonable estimation to be given based on additional available evidence.
- *unknown* - Year is unknown and it is impossible to give a reasonable estimate. This value should only rarely be used, and institutions are encouraged to supply an approximate date for undated materials whenever feasible.

¹ ISO 3166-1, the 2-letter code list for countries, is available at <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>. Accessed 2006-03-28.

3.3. Alphabetical List of Elements

<contact>

Availability: <rights.holder>, <services>

Required: No

Repeatable: Yes

Definition: Contains contact information regarding the rights holder, the rights holder's designated agent, or a provider of services related to the resource.

Application: Input the data directly within this element.

- <contact> within <rights.holder>: Use to provide relevant contact information, when available and not confidential, for the person or institution to whom a user should address questions regarding usage of or permissions regarding the resource.
- <contact> within <services>: Use to specify the service(s) available and contact information for the person or institution providing the service(s).

Attributes: None

Subelements: None

<copyright>

Availability: Root element.

Required: Yes

Repeatable: No

Definition: Wrapper element for copyright-related information for the resource.

Application: Data values must be encoded either in attributes or subelements.

Attributes: *copyright.status*, *publication.status*

Subelements: <creation>, <creator>, <general.note>, <notice>, <publication>, <rights.holder>, <services>

<country.creation>

Availability: <creation>

Required: No

Repeatable: No

Definition: Contains the name of the country in which the resource was created.

Application: Supply the name of the country in which the resource was created.

Institutions are encouraged to use a standard thesaurus for the form of name supplied.²

Input the data directly within this element. Use the available attribute, if desired, to encode a machine-readable standard code for the name of the country. This element can also be used as an empty element, with data provided only as an attribute value.

Attributes: *iso.code*

Subelements: None

<country.publication>

Availability: <publication>

Required: No

Repeatable: No

Definition: Contains the name of the country in which the resource was published.

Application: Supply the name of the country in which the resource was published.

Institutions are encouraged to use a standard thesaurus for the form of name supplied.³

Input the data directly within this element. Use the available attribute, if desired, to encode a machine-readable standard code for the name of the country. This element can also be used as an empty element, with data provided only as an attribute value.

Attributes: *iso.code*

Subelements: None

² One source might be the ISO 3166-1, the standardized list of 2-letter codes for countries, which is available at <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>. Accessed 2006-03-28.

³ One source might be the ISO 3166-1, the standardized list of 2-letter codes for countries, which is available at <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html>. Accessed 2006-03-28.

<creation>

Availability: <copyright>

Required: No

Repeatable: No

Definition: Wrapper element for information about the creation of the resource.

Unpublished resources are evaluated for copyright status based on the year of creation and the country in which creation occurred.

Application: Use primarily for unpublished resources, but can also be used for published resources if creation information is known and significant to the copyright analysis. Data values must be encoded in subelements.

Attributes: None

Subelements: <country.creation>, <year.creation>

<creator>

Availability: <copyright>

Required: No

Repeatable: No

Definition: Wrapper element for information about the creator(s) of the resource.

Application: Use primarily for unpublished resources, but can also be used for published resources if creation information is known and significant to the copyright analysis. Data values must be encoded in subelements.

Attributes: None

Subelements: <creator.corporate>, <creator.person>, <note>

<creator.corporate>

Availability: <creator>

Required: No

Repeatable: Yes

Definition: Contains the name of a corporate entity responsible for creating the resource.

Application: Supply the name of a corporate entity responsible for creating the resource. Institutions are encouraged to use and document a standardized approach to formatting names in this element. Input the data directly within this element. See guidelines under <creator.person> for information on families as creators.

Attributes: None

Subelements: None

<creator.person>

Availability: <creator>

Required: No

Repeatable: Yes

Definition: Wrapper element for information regarding an individual responsible for creating the resource. For unpublished resources created by an individual (or individuals), information regarding her/his death date is critical for evaluating copyright status. Families, though viable creators of resources, are not legally recognized as such in terms of copyright analysis. If a family name is the only name available as a creator of a resource, treat as an unknown creator in terms of copyright metadata and provide relevant information about the family in the descriptive metadata for the resource.

Application: Data values must be encoded in subelements.

Attributes: None

Subelements: <name>, <year.birth>, <year.death>

<general.note>

Availability: <copyright>

Required: No

Repeatable: Yes

Definition: Contains a general note regarding copyright information for the resource.

Application: Use for any relevant general information relating to copyright for the resource, or to the information provided in the record. Use this element only for information that cannot be accommodated in a more specific copyrightMD element. Input the data directly within this element. This is the appropriate place to encode a general boilerplate statement regarding copyright that the institution wishes to appear with the metadata about the resource.⁴

Attributes: None

Subelements: None

⁴ The CDL Rights Management Group recommends the following boilerplate statements:

- *When copyright.status is unknown:* Copyright status unknown. Some materials in these collections may be protected by the U.S. Copyright Law (Title 17, U.S.C.). In addition, the reproduction, and/or commercial use, of some materials may be restricted by gift or purchase agreements, donor restrictions, privacy and publicity rights, licensing agreements, and/or trademark rights. Distribution or reproduction of materials protected by copyright beyond that allowed by fair use requires the written permission of the copyright owners. To the extent that restrictions other than copyright apply, permission for distribution or reproduction from the applicable rights holder is also required. Responsibility for obtaining permissions, and for any use rests exclusively with the user.
- *When copyright.status is public domain:* Material in the public domain. No restrictions on use.
- *When copyright.status is copyrighted:* Transmission or reproduction of materials protected by copyright beyond that allowed by fair use requires the written permission of the copyright owners. Works not in the public domain cannot be commercially exploited without permission of the copyright owner. Responsibility for any use rests exclusively with the user.

<name>

Availability: <creator.person>, <rights.holder>,

Required: No

Repeatable:

- <name> within <creator.person>: No (note that <creator.person> *is repeatable* within <creator>)
- <name> within <rights.holder>: Yes

Definition: Contains the name of either an individual responsible for creating the resource, or an individual or corporate entity identified as copyright holder for the resource.

Application: Institutions are encouraged to use and document a standardized approach to formatting names in this element. Input the data directly within this element.

- <creator.person><name>: Supply the name of an individual responsible for creating the resource.
- <rights.holder><name>: Supply the name of either an individual or a corporate entity identified as copyright holder for the resource.

Attributes: None

Subelements: None

<note>

Availability: <creator>, <publication>, <rights.holder>, <services>

Required: No

Repeatable: Yes

Definition: Contains a general note relating to one of the following aspects of the resource: its creator(s), its publication, its rights holder(s), or services available pertaining to it.

Application: Input the data directly within this element. Use this element only for information that cannot be accommodated in a more specific copyrightMD element within the parent element where it is used.

Attributes: None

Subelements: None

<notice>

Availability: <copyright>

Required: No

Repeatable: No

Definition: Contains the copyright notice as it appears on the resource.

Application: Transcribe the copyright notice exactly as it appears on the resource. Input the data directly within this element.

Attributes: None

Subelements: None

<publication>

Availability: <copyright>

Required: No

Repeatable: No

Definition: Wrapper element for publication information regarding the resource.

Application: Use only for published works. Data values must be encoded in subelements. When <publication> is used, also use the following encoding: <copyright publication.status="published">.

Attributes: None

Subelements: <country.publication>, <note>, <publisher>, <year.copyright>, <year.publication>, <year.renewal>

<publisher>

Availability: <publication>

Required: No

Repeatable: No

Definition: Contains the name of the publisher of the resource.

Application: Supply the name of publisher of the resource. Institutions are encouraged to use and document a standardized approach to formatting names in this element. Input the data directly within this element.

Attributes: None

Subelements: None

<rights.holder>

Availability: <copyright>

Required: No

Repeatable: No

Definition: Wrapper element for information about the rights holder for the resource and contact information for the rights holder or rights holder's designee.

Application: Use to direct users to the appropriate point for questions about rights and permissions related to the resource. Data values must be encoded in subelements.

Attributes: None

Subelements: <contact>, <name>, <note>

<services>

Availability: <copyright>

Required: No

Repeatable: Yes

Definition: Wrapper element for information relating to services that might be offered relating to the resource, such as providing copies.

Application: Use to direct users to the appropriate point for obtaining services relating to the resource. Data values must be encoded in subelements.

Attributes: None

Subelements: <contact>, <note>

<year.birth>

Availability: <creator.person>

Required: No

Repeatable: No

Definition: Contains the year of birth of an individual responsible for the creation of the resource.

Application: Supply the year of birth of the creator. Use the standardized form of YYYY, and do not include the month or day of birth. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the birth year encoded, especially if the year is an approximation. This element can also be used as an empty element, with data provided only as an attribute value, when the birth year is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

<year.copyright>

Availability: <publication>

Required: No

Repeatable: No

Definition: Contains the year the resource was copyrighted.

Application: Supply the year the resource was copyrighted, typically based on a copyright notice on the resource itself. The resource does not have to have been registered with the copyright office. Institutions should not approximate copyright year if it does not appear on the work, or in some reliable alternate source for this information. Use the standardized form of YYYY, and do not include month or day information. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the copyright year encoded. This element can also be used as an empty element, with data provided only as an attribute value, when the copyright year is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

<year.creation>

Availability: <creation>

Required: No

Repeatable: No

Definition: Contains the year the resource was created. This is crucial information for making a copyright determination for an unpublished resource created by a corporate entity, or when the death date of an individual responsible for the creation of the resource is unknown.

Application: Supply the year the resource was created. Use the standardized form of YYYY, and do not include month or day information. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the year of creation encoded. This element can also be used as an empty element, with data provided only as an attribute value, when the year of creation is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

<year.death>

Availability: <creator.person>

Required: No

Repeatable: No

Definition: Contains the year of death of an individual responsible for the creation of the resource. This is crucial information for making a copyright determination for an unpublished resource created by an individual.

Application: Supply the year of death of the creator. Use the standardized form of YYYY, and do not include the month or day of death. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the death year encoded, especially if the year is an approximation. This element can also be used as an empty element, with data provided only as an attribute value, when the death year is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

<year.publication>

Availability: <publication>

Required: No

Repeatable: No

Definition: Contains the year the resource was published.

Application: Supply the year the resource was published, typically based on information available on the resource itself. Use the standardized form of YYYY, and do not include month or day information. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the publication year encoded. This element can also be used as an empty element, with data provided only as an attribute value, when the publication year is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

<year.renewal>

Availability: <publication>

Required: No

Repeatable: No

Definition: Contains the year the copyright for the resource was renewed.

Application: Supply the year the copyright for the resource was renewed. Few items are likely to contain a copyright renewal notice. This will only pertain to works that were registered with the copyright office and subsequently renewed.⁵ Such works will have a copyright date before 1964 (i.e., up to and including 1963). Institutions should not approximate the year of copyright renewal if it does not appear on the work, or in some reliable alternate source for this information. Use the standardized form of YYYY, and do not include month or day information. Input the data directly within this element. Use the available attribute, if desired, to supply additional information about the copyright year encoded. This element can also be used as an empty element, with data provided only as an attribute value, when the year of copyright renewal is unknown or no attempt has been made by the institution to encode this information.

Attributes: *year.type*

Subelements: None

⁵ See <http://onlinebooks.library.upenn.edu/renewals.html> for detailed information about searching for copyright renewal information. Accessed 2006-04-04.

4 Examples

Unpublished photograph, copyrighted. Known corporate creator

```
<?xml version="1.0" encoding="UTF-8"?>
<copyright copyright.status="copyrighted"
  publication.status="unpublished"
  xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
  http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
  <creation>
    <year.creation year.type="approximate">1969?</year.creation>
    <country.creation iso.code="us">United States</country.creation>
  </creation>
  <creator>
    <creator.corporate>Rene Laursen Photography</creator.corporate>
  </creator>
  <rights.holder>
    <name>Rene Laursen Photography</name>
    <contact>702 North Grand, Santa Ana, CA 92701. 714-547-4604</contact>
  </rights.holder>
  <services>
    <contact>Berkeley Public Library</contact>
    <note>Consult institution: 2090 Kittredge St., Berkeley, CA 94704. Phone: (510) 981-6100,
    TDD: (510) 981-6100, Web: http://www.berkeleypubliclibrary.org</note>
  </services>
  <general.note>Transmission or reproduction of materials protected by copyright beyond that
  allowed by fair use requires the written permission of the copyright owners. In addition,
  the reproduction of some materials may be restricted by terms of gift of purchase
  agreements, donor restrictions, privacy and publicity rights, licensing and trademarks.
  Works not in the public domain cannot be commercially exploited without permission of the
  copyright owner. Responsibility for any use rests exclusively with the user.</general.note>
</copyright>
```

Unpublished photograph, copyrighted. Known personal creator

```
<?xml version="1.0" encoding="UTF-8"?>
<copyright copyright.status="copyrighted" publication.status="unpublished"
  xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
  http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
  <creation>
    <year.creation year.type="approximate">1969</year.creation>
    <country.creation iso.code="us">United States</country.creation>
  </creation>
  <creator>
    <creator.person>
      <name>Van Young, Sayre</name>
      <year.birth year.type="unknown">Unknown</year.birth>
    </creator.person>
  </creator>
</copyright>
```

```

        <year.death year.type="unknown">Unknown</year.death>
    </creator.person>
</creator>
<rights.holder>
    <name>Van Young, Sayre</name>
</rights.holder>
<services>
    <contact>Berkeley Public Library</contact>
    <note>Consult institution: 2090 Kittredge St., Berkeley, CA 94704. Phone: (510) 981-6100,
        TDD: (510) 981-6100, Web: http://www.berkeleypubliclibrary.org</note>
</services>
<general.note>Transmission or reproduction of materials protected by copyright beyond that
    allowed by fair use requires the written permission of the copyright owners. In addition,
    the reproduction of some materials may be restricted by terms of gift or purchase
    agreements, donor restrictions, privacy and publicity rights, licensing and trademarks.
    Works not in the public domain cannot be commercially exploited without permission of the
    copyright owner. Responsibility for any use rests exclusively with the user.</general.note>
</copyright>

```

Unpublished photograph, copyright unknown. Unknown creator

```

<?xml version="1.0" encoding="UTF-8"?>
<copyright copyright.status="unknown"
    publication.status="unpublished"
    xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
    http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
    <creation>
        <year.creation year.type="exact">1920</year.creation>
        <country.creation iso.code="us">United States</country.creation>
    </creation>
    <creator>
        <creator.person>
            <name>Unknown</name>
            <year.birth year.type="unknown">Unknown</year.birth>
            <year.death year.type="unknown">Unknown</year.death>
        </creator.person>
    </creator>
    <rights.holder>
        <name>Unknown</name>
    </rights.holder>
    <services>
        <contact>Mission Viejo Library</contact>
        <note>Consult institution: 100 Civic Center Mission Viejo, CA 92691. Phone: (949) 830-7100,
            Web: http://cmvl.org</note>
    </services>
    <general.note>Copyright status unknown. Some materials in these collections may be protected by
        the U.S. Copyright Law (Title 17, U.S.C.). In addition, the reproduction of some materials
        may be restricted by terms of gift or purchase agreements, donor restrictions, privacy and
        publicity rights, licensing and trademarks. Transmission or reproduction of materials
        protected by copyright beyond that allowed by fair use requires the written permission of
        the copyright owners. Works not in the public domain cannot be commercially exploited
        without permission of the copyright owner. Responsibility for any use rests exclusively with
        the user.</general.note>

```

</copyright>

Published postcard, copyrighted. Known corporate creator

```
<?xml version="1.0" encoding="UTF-8"?>
<copyright copyright.status="copyrighted"
  publication.status="published"
  xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
  http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
  <publication>
    <country.publication iso.code="us">United States</country.publication>
    <year.copyright year.type="exact">1984</year.copyright>
    <year.publication year.type="exact">1984</year.publication>
    <publisher>Mission Viejo Company</publisher>
  </publication>
  <creator>
    <creator.corporate>Mission Viejo Company</creator.corporate>
  </creator>
  <rights.holder>
    <name>Mission Viejo Company</name>
  </rights.holder>
  <services>
    <contact>Mission Viejo Library</contact>
    <note>Consult institution: 100 Civic Center Mission Viejo, CA 92691. Phone: (949) 830-7100,
    Web: http://cmvl.org/</note>
  </services>
  <notice>&#xA9; 1984 Mission Viejo Company</notice>
  <general.note>Transmission or reproduction of materials protected by copyright beyond that
  allowed by fair use requires the written permission of the copyright owners. In addition,
  the reproduction of some materials may be restricted by terms of gift of purchase
  agreements, donor restrictions, privacy and publicity rights, licensing and trademarks.
  Works not in the public domain cannot be commercially exploited without permission of the
  copyright owner. Responsibility for any use rests exclusively with the user.</general.note>
</copyright>
```

Published plat map, public domain. Known corporate creator (US Federal Government)

```
<?xml version="1.0" encoding="UTF-8"?>
<copyright copyright.status="pd_usfed"
  publication.status="published"
  xmlns="http://www.cdlib.org/inside/diglib/copyrightMD"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.cdlib.org/inside/diglib/copyrightMD
  http://www.cdlib.org/inside/projects/rights/schema/copyrightMD.xsd">
  <publication>
    <country.publication iso.code="us">US</country.publication>
    <year.copyright year.type="unknown">Unknown</year.copyright>
    <year.renewal year.type="unknown">Unknown</year.renewal>
    <year.publication year.type="exact">1860</year.publication>
    <publisher>Surveyor General of the United States</publisher>
  </publication>
  <creator>
    <creator.corporate>Surveyor General of the United States</creator.corporate>
  </creator>
</copyright>
```

</creator>
<services>
 <contact>Oakland Public Library</contact>
 <note>Consult institution: 125 14th St, Oakland, CA. Phone: (510) 238-3134, Web:
 <http://www.oaklandlibrary.org></note>
</services>
<general.note>Material is in the public domain. No restrictions on use.</general.note>
</copyright>

5 Frequently Asked Questions

What are the functional objectives of copyrightMD?

There are two primary objectives of copyrightMD. One is to express factual information that allows users to make an informed copyright assessment of a given work. Additionally copyrightMD provides users with sources of further information about the copyright status, or to a person or institution that can be a resource for permissions. copyrightMD is user-oriented and not intended to serve as a recordkeeping mechanism for other copyright-related information collected and gathered by institutions for internal management (such as donor agreement information, copyright permission request histories, etc.).

Why should institutions use copyrightMD?

One of the chief characteristics of digital resources--whether digitally reformatted or "born digital"--is the ease with which they can be re-used either in whole or in part. Reuse rights are determined based on the copyright status of an item as well as the nature of the intended use. Libraries, archives and cultural memory institutions increasingly acquire, collect, create, provide access and preserve digital objects and collections. Recent Association of Research Libraries reports that over 30% of member libraries collections are digital.

How does copyrightMD help users of digital resources?

It is incumbent on the users of materials to assess the copyright status of an item and to understand whether use is permitted by law or if it may be necessary to acquire permission. To make this copyright assessment, users need to have certain information about the item. In the case of digitally reformatted materials in particular, this information may be inherent in the context of the archive or collection in which the physical items reside. But the digitization of works often removes that context, and connections to the original works may weaken over time. copyrightMD provides users the information necessary, to make an assessment of copyright status. Historically, however, the recording of copyright-related information is often not part of the tradition of cataloging or metadata creation in libraries, museums, and archives and thus this information has not been readily available to users. To complicate matters further, since 1978, copyright law no longer requires a notice of copyright to protect a work making it more difficult to identify creators and track down copyright owners.

Libraries have as part of their fundamental mission to provide users with a means to access and use information resources. Providing the user with all available information relating to the copyright status of the work increases the user's ability to access and appropriately use the work and as such is a part of any digital library's service and mission. Current thinking is that this type of contact information will be provided to users through the user interface. Additionally, systems could also be configured so that users could scope searches and browses to collections with

particular copyright statuses. For example, a search could be limited to "public domain" works only, or only to materials where the rights holder and contact information are known.

A second functional objective of the schema is to explicitly associate item-level copyright information with discrete digital objects. By preserving copyright metadata with individual digital objects as they are created, we will be preventing the creation of "orphaned works" in the future.

How does copyrightMD provide information to assist users with making a copyright assessment?

The goal is to provide information that supports a copyright assessment. Assertions give users some idea of the presumed copyright status of the work, but do not take the place of factual copyright metadata. In fact, assertions could potentially be automatically generated by display systems, given parsed copyright metadata.

Copyright status is not a substitute for the copyright metadata, for several of reasons. First, copyright status isn't sufficient to tell the user what to do for works that are under copyright. The copyright data elements add information that could inform a fair use assessment, as well as name and contact information should the user seek to obtain permission. Next, copyright status is a moving target (albeit not moving very quickly), so an item with a status of "copyrighted" will eventually have a status of "public domain." Third, copyright metadata should be seen as part of the archival package. Among other things, it helps future users understand how the copyright assessment was determined.

What tools are available for encoding using copyrightMD?

Any XML authoring tool can be used to create copyrightMD records. Other tools, such as standard databases could also be configured to produce copyrightMD outputs.

The CDL hopes to work with the digital library community to promote the creation of open-source tools and solutions for generating metadata that adheres to copyrightMD. The CDL also hopes to provide data transformation services whereby metadata records in one encoding format could be converted into schema-compliant encoding, using solutions such as [eXtensible Stylesheet Language Transformations \(XSLT\)](#) stylesheets.

Copyright-related information can change. Won't this metadata need a lot of maintenance?

It's true that copyrights can be transferred, so the copyright holder can change, and an unpublished work may be published at a later date. Metadata creators are understandably concerned about having to update this information for stored works. Because one cannot guarantee that this maintenance will take place, we recommend that copyright metadata be clearly date-stamped so that future users can know that the data in the record represents the copyright data at a particular moment in time. Even though some particulars may have changed, this gives the user a solid starting point for investigating current copyright status.

What is the relationship of the copyrightMD to other schemas?

copyrightMD is designed to express copyright metadata within an XML environment, and therefore serves as an extension schema for the [Metadata and Encoding Transmission Standard \(METS\)](#).

The CDL Rights Management Group did an analysis of the functional requirements related to copyright metadata and as part of the analysis did a review of existing metadata schemas. The results of the analysis and review indicated that no existing metadata schema included all necessary data elements and, in fact, most merely contained broad buckets for the collecting and recording of copyright metadata with no functionality for machine actionability.

What are the benefits of using copyrightMD that are not already available in other schemas, such as Dublin Core, MODS, MARCXML, or METSRights?

The metadata schemas mentioned above do not adequately accommodate -- both in terms of extent and granularity -- the range of essential descriptive information necessary to assess the copyright status of a given object. [Dublin Core](#), [Metadata Object Description Schema \(MODS\)](#), and [MARCXML](#), for example, are designed to capture descriptive metadata and have relatively few data elements for capturing copyright-specific data. [METSRights](#) was designed to capture both copyright, contract, and license data, but it only has data elements for rights holder name and contact information. Even if multiple schema such as MODS and METSRights are combined into a composite metadata record for a given digital object, essential copyright information will still be lacking.

Digital license schemas such as [XrML \(eXtensible rights Markup Language\)](#) and [ODRL \(Open Rights Digital Language\)](#) may be able to contain copyright-specific data, but as they are used today their emphasis is on license conditions not on copyright status.

Why does copyrightMD contain elements that overlap with other descriptive metadata schema?

Descriptive metadata serves somewhat different purposes than copyright metadata: the former is primarily geared towards helping users locate and identify items via searching and browsing, while the latter is geared towards assisting users with determining the copyright status of a given item.

Common cataloging practices and content standards for descriptive cataloging are not principally geared towards recording important copyright information. For example, knowing who created a work does not tell you who currently holds the copyright in the work. At the same time, the creator of a work is a key element in copyright law, even if the creator is no longer the copyright holder. Information about the creator is necessary for the expiration of the rights over time. Another difference between descriptive practice and copyright definition is that copyright relies heavily on the creator's death date, since that is the date that determines the length of the copyright. In common descriptive metadata cataloging practices, dates are only included with author's names when those are needed to distinguish between two authors with the same name,

and only the date of birth is needed for that purpose. Occasionally, the only known date is the death date, and therefore that is the date used, with “d.” before it to indicate that it is the date of death. In practice, records with author’s date of birth are rarely updated to include the date of death, so this key piece of information is missing from most cataloging.

It is possible, however, that some data values present in a descriptive metadata record will directly overlap with data values in a copyright metadata record. copyrightMD will seek to provide a mechanism to support referential linking between related data elements, in order to reduce data redundancy in a single record.

Can the copyrightMD XML be used with or imported into other XML schemas?

Yes. In fact, we assume that it will be used in conjunction with other metadata. For example, a copyrightMD record could be combined with other descriptive (e.g., [Dublin Core](#), [Metadata Object Description Schema \(MODS\)](#), or [MARCXML](#)), structural, administrative, and technical metadata records in a [METS](#) wrapper. Our main interest is in the data elements of copyrightMD, and these could presumably be expressed in formats other than XML schema. If you do use the data elements in other formats, please let us know so we can add your implementation to our web site.

What is meant by the term "published" in the copyrightMD schema? How do you know if a work should be considered published vs. unpublished? Why does this matter?

The U.S. copyright law maintains some differences between published works and unpublished works. For a useful summary of the differences, see Peter Hirtle's [Copyright Term and the Public Domain in the United States](#). The law itself defines the act of publication as: "... the distribution of copies or phonorecords of a work to the public by sale or other transfer of ownership, or by rental, lease, or lending." (see [US Code, Title 17, section 101](#)). It is admittedly often not a simple matter to determine whether an item is or is not published, and users of the metadata have the option of coding this element "unknown."

I have hundreds (or thousands) of items already digitized but don't have copyright information for them. What can I do?

All of the copyright data elements can be coded as "unknown," which is the appropriate coding for this situation. This alerts users to the fact that they will not be able to rely on the provided metadata to aid in copyright determination, and suggests that they should contact your institution (or the appropriate services contact) for further information.

I have a collection where most of the objects have the same copyright status. Does copyright information need to be provided for every single object?

The copyright data elements may be used at any level of description, and should be used at the level that is most appropriate to document the copyright status of a resource. This may be at an

aggregate-level or at an item-level, depending on the degree to which the components of the resource share given copyright information. In some cases, the copyright data elements could be used at the component level for a particular item; for example, in the case of a photograph album comprising images created by different photographers, each page of the album may need to reflect a different copyright status.

Does copyrightMD accommodate objects with complex copyright situations, such as multi-media works?

The copyright data elements can be used at the component level for a particular item; for example, in the case of a multi-media work comprising images, music, and textual material, each component may need to reflect a different copyright status. A future version of copyrightMD will seek to address more fully issues with documenting copyright statuses for complex objects. It isn't clear yet if we can just add roles to creators of different components of these complex objects, or if we need to add additional data structures to the schema. If you have interesting materials to provide as test cases, we'd be happy to hear from you.

What copyrightMD elements are required for a valid record?

The copyrightMD schema reflects the full suite of elements that could be utilized in any given record. A valid copyrightMD record requires the top-level <copyright> element with associated @copyright.status and @publication.status attributes; however, no additional elements or attributes are required.

What copyrightMD elements are required by the CDL?

The CDL's *Digital Object Guidelines* provide recommendations for a minimal record (i.e., which elements must be present in any given copyright metadata record). In principle, the CDL's goal is for copyright information to be recorded when it is known by contributing institutions.