

<copyright>



# copyrightMD User Guidelines



<copyright>

## Version 0.9

California Digital Library  
Rights Management Group  
April 2006

Guidelines located at: <http://www.cdlib.org/inside/projects/rights/schema/>

# 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.

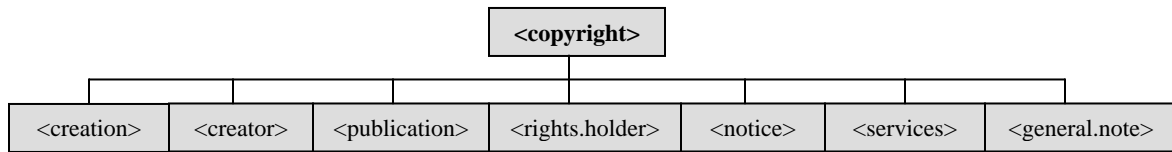
# 3

## Detailed Description of copyrightMD

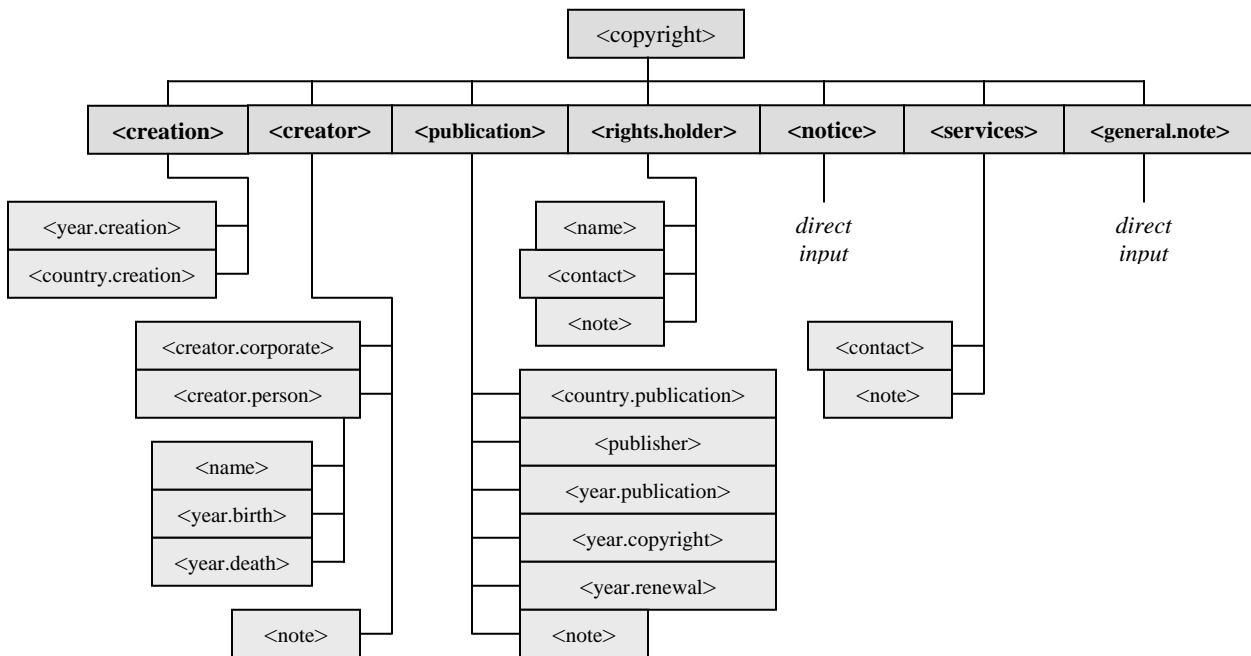
### 3.1. General Application Guidelines

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.<sup>1</sup>

## *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.

---

<sup>1</sup> 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.<sup>2</sup>

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.<sup>3</sup>

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

---

<sup>2</sup> 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.

<sup>3</sup> 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.<sup>4</sup>

**Attributes:** None

**Subelements:** None

---

<sup>4</sup> 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.<sup>5</sup> 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

---

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