

HTML Meta Tags for Dynamically-Created LC Web Pages
Metadata Schemes: Dublin Core, MODS, Google Scholar Tag Set
July 2011

Effective use of trusted, clearly-defined HTML meta tags can improve search engine optimization and the citation services of Web 2.0 applications. The Library of Congress has, therefore, initiated a phased effort to add a consistent set of HTML meta tags to dynamically-created web pages in LC digital content discovery applications. These pages are produced "on the fly," using templates to build web presentations of LC's digital files and metadata.

The HTML meta tags added to each web page have been selected from a core set of descriptive metadata developed by a subgroup of the LC Metadata for Digital Content (MDC) group. These descriptive HTML tags are hidden to most users in the <head> element of each web page and do not affect what users actually see. In LC discovery applications, these tags support services such as the "cite this page" functionality for RSS feeds and blogs, web-based citation managers, OpenURL, and cross-application search refinements. As consistently applied, trusted metadata for digital content in the form of HTML meta tags are also potentially useful for structured search engine crawling.

Library applications have incorporated these tags on digital content displays include the Prints and Photographs Online Catalog (see URL: <<http://www.loc.gov/pictures/>>), the Library's finding aid (EAD) search application (see URL: <<http://www.loc.gov/findingaids/>>), and *Chronicling America* (see URL: <<http://chroniclingamerica.loc.gov/>>). These applications use one or more of the following sets of HTML meta tags, with values extracted from existing descriptive metadata:

1. Dublin Core metadata

Required in HTML:

```
<link rel="schema.DC" href="http://purl.org/DC/elements/1.1/" />
```

DC tags used in LC HTML meta tags:

```
<meta name="DC.title" content="xxx" />
```

Required

Non-Repeatable

Content source: 245\$afgknp or main title equivalent in source metadata; stylesheet may append additional subfields

Encoding guidelines: Remove non-filing articles (this information affects sorting)

```
<meta name="DC.creator" content="xxx" />
```

Repeatable

Content source: 100\$abcdgq -110\$abcd -111\$acdenq or creator equivalent in source metadata; stylesheet may append additional subfields

Encoding guidelines: Use last name-first name order for personal names.

```
<meta name="DC.publisher" content="xxx" />
```

Repeatable

Content source: 260\$b or publisher in source metadata

```
<meta name="DC.date" content="xxx" />
```

Repeatable

Content sources: controlled date(s) in source metadata (such as MARC 008 or other fields)

Encoding guidelines: Follow EDTF for date transcription:

<http://www.loc.gov/standards/datetime/>

- <meta name="DC.identifier" content="xxx" /> *Repeatable*
Content sources: handles, LCCN Permalinks, persistent URLs
Encoding guidelines: Content should be a URL
- <meta name="DC.type" content="xxx" /> **Required**
Repeatable
Content source: source metadata
Encoding guidelines: controlled list of terms from mods:typeofResource
<http://www.loc.gov/standards/mods/mods-outline.html#typeOfResource>
- <meta name="DC.rights" content="xxx" /> *Repeatable*
Content source: use free text or URL
- <meta name="DC.language" content="xxx" /> *Repeatable*
Content source: 008/35-37 or 041 using two- or three-letter codes from ISO 639 (depending on source metadata)
- <meta name="DC.subject" content="xxx" /> *Repeatable*
Content source: 600\$abcdgqtvxyz, 610\$abcdtvxyz, 611\$acdenqtvxyz, 630\$adflmnopsvxyz, 648\$avxyz, 650\$abvxyz, 651\$abvxyz, 653\$a, 655\$avxyz, 656\$avxyz, 662\$bcdefgh. Stylesheet may append additional subfields (such as title subfields in fields 600-610-611)
Encoding guidelines: Include dashes between \$vxyz and 662\$bcdefgh, if possible
- <meta name="DC.relation" content="xxx" /> *Repeatable*
Content source: Could be used to reference an item that is part of a collection or aggregation
 Mapping needs to be considered on a case-by-case basis, depending on the source of the record (applications may want to use 773\$t or 985)

DC tags not used by LC in HTML meta tags:

- <meta name="DC.description" /> Use instead <meta name="description" />
 <meta name="DC.coverage" /> Consider more refined GIS tags, based on other MDC pilots
 <meta name="DC.format" /> LC believes this is not needed

2. MODS metadata used in LC HTML meta tags to provide more granular metadata for resource discovery:

Required in HTML:

<head profile="http://www.loc.gov/mods/simple#">

MODS tags recommended by LC for use in HTML meta tags:

<meta name="mods.title" content="xxx" />

Required if present
Non-Repeatable

Content source: 245\$afgknp or main title equivalent in source metadata; stylesheet may append additional subfields

Encoding guidelines: Remove non-filing articles (this information affects sorting)

<meta name="mods.uniformTitle" content="xxx" />

Required if present
Non-Repeatable

Content source: 130\$adflmnoprs, 240\$adflmnoprs, or uniform title equivalent in source metadata; stylesheet may append additional subfields

<meta name="mods.personalName" content="xxx" />

Repeatable

Content source: 100\$abcdgq or personal creator equivalent in source metadata; use when \$t is not present

Encoding guidelines: Use last name-first name order for personal names.

<meta name="mods.personalNameTitle" content="xxx" />

Repeatable

Content source: 100\$abcdgqt or personal name-title creator equivalent in source metadata; include \$t; stylesheet may append additional subfields

Encoding guidelines: Use last name-first name order for personal names.

<meta name="mods.corporateName" content="xxx" />

Repeatable

Content source: 110\$abcd in source metadata or corporate creator equivalent in source metadata; use when \$t is not present

<meta name="mods.corporateNameTitle" content="xxx" />

Repeatable

Content source: 110\$abcdt in source metadata or corporate name-title creator equivalent in source metadata; include \$t; stylesheet may append additional subfields

<meta name="mods.conferenceName" content="xxx" />

Repeatable

Content source: 111\$acdeng in source metadata or conference creator equivalent in source metadata; use when \$t is not present

<meta name="mods.conferenceNameTitle" content="xxx" />

Repeatable

Content source: 111\$acdengt in source metadata or conference name-title creator equivalent in source metadata; include \$t; stylesheet may append additional subfields

<meta name="mods.placeOfOrigin" content="xxx" />

Repeatable

Content source: 260\$a or equivalent for place of publication/issuance in source metadata

<meta name="mods.publisher" content="xxx" />

Repeatable

Content source: 260\$b or equivalent in source metadata

- <meta name="mods.dateIssued" content="xxx" /> *Repeatable*
Content sources: controlled date(s) in source metadata (MARC 008/07-14)
Encoding guidelines: If date range (If 008/06=c,d,i,k,m, or u) insert "/" between dates. Follow EDTF for date transcription for date ranges: <http://www.loc.gov/standards/datetime/>. If an open date use first date/open (e.g. 1992/open)
- <meta name="mods.dateCreated" content="xxx" /> *Repeatable*
Content sources: controlled date(s) in source metadata (MARC 008/07-14) or 245\$f for manuscripts
Encoding guidelines: If from 245\$f, convert "-" between 2 dates to "/" (if "-" is followed by space, insert "open" after "/"; if no "-" output value as string)
 Follow EDTF for date transcription for date ranges: <http://www.loc.gov/standards/datetime/>
 Examples: 2004/2006 or 2004/open
- <meta name="mods.url" content="xxx" /> *Repeatable*
Content sources: handles, Permalinks, persistent URLs
Encoding guidelines: Content should be a URL
- <meta name="mods.typeOfResource" content="xxx" /> **Required if present**
Repeatable
Content source: source metadata against a controlled list of terms from mods:typeOfResource <http://www.loc.gov/standards/mods/mods-outline.html#typeOfResource>
 if a collection, repeat typeOfResource with value "collection"
 if a manuscript, repeat typeOfResource with value "manuscript"
See mapping information below.
- <meta name="mods.accessCondition" content="xxx" /> *Repeatable*
Content source: Free text or URL
- <meta name="mods.subjectPersonalName" content="xxx" /> *Repeatable*
Content source: 600\$abcdgqvxyz or personal name subject equivalent in source metadata; use when \$t not present; stylesheet may append additional subfields
Encoding guidelines: Use last name-first name order for personal names and, if possible, include dashes between \$vxyz
- <meta name="mods.subjectPersonalNameTitle" content="xxx" /> *Repeatable*
Content source: 600\$abcdgqtvxyz or personal name-title subject equivalent in source metadata; include \$t; stylesheet may append additional subfields
Encoding guidelines: Use last name-first name order for personal names and, if possible, include dashes between \$vxyz
- <meta name="mods.subjectCorporateName" content="xxx" /> *Repeatable*
Content source: 610\$abcdvxyz in source metadata; use when \$t not present; stylesheet may append additional subfields
Encoding guidelines: Include dashes between \$vxyz, if possible

- <meta name="mods.subjectCorporateNameTitle" content="xxx" /> *Repeatable*
Content source: 610\$abcdtvxyz in source metadata; include \$t; stylesheet may append additional subfields
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectConferenceName" content="xxx" /> *Repeatable*
Content source: 611\$acdenqvxyz in source metadata; use when \$t not present; stylesheet may append additional subfields
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectConferenceNameTitle" content="xxx" /> *Repeatable*
Content source: 611\$acdenqtvxyz; include \$t; stylesheet may append additional subfields
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectUniformTitle" content="xxx" /> *Repeatable*
Content source: 630\$adfklnoprsvxyz
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectTemporal" content="xxx" /> *Repeatable*
Content source: 648\$avxyz
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectTopic" content="xxx" /> *Repeatable*
Content source: 650\$abvxyz, 653\$a in source metadata
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectGeographic" content="xxx" /> *Repeatable*
Content source: 651\$abvxyz in source metadata
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectOccupation" content="xxx" /> *Repeatable*
Content source: 656\$akvxyz in source metadata
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.subjectHierarchicalGeographic" content="xxx" /> *Repeatable*
Content source: 662\$abcdefgh in source metadata
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.genre" content="xxx" /> *Repeatable*
Content source: 655\$avxyz in source metadata
Encoding guidelines: Include dashes between \$vxyz, if possible
- <meta name="mods.hostRelatedItem" content="xxx" /> *Repeatable*
Content source: 773\$at in source metadata; stylesheet may append additional subfields
- <meta name="mods.precedingRelatedItem" content="xxx" /> *Repeatable*
Content source: 780\$at in source metadata; stylesheet may append additional subfields

<meta name="mods.succeedingRelatedItem" content="xxx" /> *Repeatable*
Content source: 785\$at in source metadata; stylesheet may append additional subfields

<meta name="mods.identifier" content="xxx" /> *Repeatable*
<meta name="mods.lccn" content="xxx" /> *Not Repeatable*
<meta name="mods.issn" content="xxx" /> *Repeatable*
<meta name="mods.isbn" content="xxx" /> *Repeatable*
Content sources: LCCN Permalink; 020 (ISSN), 022 (ISBN)
Encoding guidelines: use <identifier> if undistinguished type

<meta name="mods.language" content="xxx" /> *Repeatable*
Content source: two- or three-letter codes from ISO 639 (depending on source metadata)

Conversion of MODS typeOfResource and DC Type values from MARC 21 Ldr/06

<i>if</i> Leader/07 (code c)	repeat typeOfResource with value "collection"
<i>if</i> Leader/06 (codes d,f,p,t)	repeat typeOfResource with value "manuscript"
a,t	text
e,f	cartographic
c, d	notated music
i	sound recording—non-musical
j	sound recording—musical
k	still image
g	moving image
o	kit
r	three-dimensional object
m	software, multimedia
p	mixed material

3. Google Scholar tag set (as described in: <http://scholar.google.com/intl/en/scholar/inclusion.html>)
LC may use for relevant LC content (such as articles from serials). Rules to be determined.

Possible tags include:

- citation_language
- citation_doi
- citation_abstract_html_url
- citation_title
- citation_authors
- citation_issn
- citation_date
- citation_journal_title
- citation_volume
- citation_firstpage
- citation_lastpage
- citation_publisher
- citation_issue
- citation_pdf_url
- citation_pmid
- citation_keywords (multiple instances OK)
- citation_conference
- citation_dissertation_name
- citation_dissertation_institution
- citation_patent_number
- citation_patent_country
- citation_technical_report_number
- citation_technical_report_institution