HUD Metadata Implementation Guide

Contents

1 Introduction

Metadata can be a powerful tool for managing information and ensuring the public can find what they need on our websites efficiently. This document provides implementation guidance on how to include metadata on web pages for www.hud.gov, hudatwork.hud.gov, and espanol.hud.gov. While most of the metadata elements addressed here are recommended by the Web Content Standards Working Group of the Interagency Committee on Government Information to be placed on the home page and all major entry points, we at HUD will use some additional elements. In addition, we hope to place metadata on most, if not all, web pages comprising our websites.

1.1   What is metadata?

Metadata is information which describes an item, not the item itself. In web terms, Metadata doesn't actually appear on a web page but describes the contents and attributes of the page. (If you want to see the metadata, you have to reveal the HTML codes. For example if you look at http://www.gc.ca/ you'll see the metadata as the collection of lines starting with <meta> in the <head> portion of the web page.)

Metadata is used by search engines and other automated tools to help users find the information they need more efficiently. Think about a library card catalog. A card catalog doesn't contain the books or periodicals, but contains data relating to them: their titles, authors, publishers, published date, etc. So, a card catalog would be a collection of Metadata. Another example is a department store catalog: it lists the items, brand, price, color, capacity, etc. All this information could be thought of as Metadata.

1.2   Why is metadata important?

Metadata is important for three major reasons:

a. We want to create information that will help our readers find what they want based on specific criteria: title, creator, date, subject, audience, etc. For example, what if our readers want all the speeches given by a particular Secretary? Maybe they want everything related to Senior Citizens. Or, maybe they are interested in all speeches given by Secretary Martinez in 2002 related to Senior Citizens. Metadata will enable us to help them find this information.

b. Metadata will help us manage our sites better. Using Metadata, we will be able to quickly find old, outdated content (content with dc.date.valid dates from a year ago, for example). Using the dc.creator element, we'll be able to quickly determine who "owns" a particular page (regardless of where the page resides on the website).

c. Metadata will allow information to be tracked and assembled government-wide. If all federal agencies use the same set of Metadata the same way, we can do a search to find all the home-buying programs across the federal government, or find all the pages across the federal government that were created for a senior-citizen audience group.

As you can see, Metadata can be a very powerful tool in managing and accessing information on the Internet and intranet. This is why creating good metadata is so important.

1.3   What are metadata elements?

You've already seen a couple of the elements we'll be using: dc.creator is the element which captures who created or owns a particular web page. A metadata element is simply the type of information we will capture. For example, the Title of a document is one "element," Subject is another element. For metadata to be truly useful, these elements need to be standardized and used consistently. The elements in this guide are based on the Dublin Core metadata element set which is both a NISO standard (z.39.85-2001) and an ISO standard (ISO 15836-2003).

Going back to our card catalog system—the Card Catalog is a Metadata Collection. Now think of the rules that get applied to the Card Catalog—the Dewey Decimal system, what is a "title," how do you display the published date. The aggregation of these rules and the element definitions comprise a metadata system.

At HUD, we are going to use the "Dublin Core" metadata element set and will be coding our web pages according to their scheme. Dublin Core—named for Dublin, Ohio where it was created—is the international standard that served as the basis for the metadata appearing on government web pages in the United Kingdom, Australia, Canada, as well as many other countries. Should the U.S. federal government require a standard set of metadata, it will undoubtedly be based on the concepts, if not the actual terms, set forth in Dublin Core.

1.4   Mandatory Elements

We will be using 8 Dublin Core elements on our web pages:

1. dc.audience
2. dc.creator
3. dc.date.created
4. dc.date.valid
5. dc.description
6. dc.language
7. dc.subject
8. dc.title

In order for this metadata to be truly useful, we need to understand and use them correctly. The official definitions and additional guidance are found in Section 3, but here's a quick rundown on the elements and what they mean. First, the "dc." part of the element says this element is part of the Dublin Core metadata registry.

1. dc.audience –Who is the intended audience for this web page? This can be repeated for as many audiences as needed and we will use a controlled vocabulary.
2. dc.creator—Who created or "owns" this document? In most cases, it will be the office responsible for the content. For example, the Office of Housing. However, there will be times that we might want to capture the official (the actual person) responsible for the content—such as speeches given by the Secretary. Or, sometimes we need to capture both the official and their office. For this reason, there can be multiple creators.
3. dc.date.created—When was the page created? If unknown, we'll use our best estimate at when the page was created. If you really can't figure this out, use the last time it was reviewed.
4. dc.date.valid—initially, this date should be the date the page was created. Then, when we do our quarterly certifications, this is where you'll put the date you last certified that the content was still current and accurate.
5. dc.description—What is on this web page? In plain language, how would you describe this web page to someone?
6. dc.language—For most of our pages, this will be the code for English. However, on espanol.hud.gov, it will be Spanish, and in those rare cases when we have content in some other language, the codes are provided in Section 3 below. It is possible for a page to be in two or more languages—for example a page that has text in both English and Spanish. For this reason, this element can also have more than one entry.
7. dc.subject—What is the subject of the page? Is it homebuying? renting, or something else? We'll develop a controlled vocabulary from which you can choose. Again, there can be more than one subject for a page, so this element can have multiple entries.
8. dc.title—There's the title that shows on the page, but sometimes we'll want to capture alternative titles. The home page is a good example. We call it "homes and communities." However, another title could be "U.S. Department of Housing and Urban Development (HUD) Home Page." For that reason, title can also be repeated as a metadata element.

1.5 Example of a complete metadata record

Here's how the metadata for the front page of www.hud.gov might look:

```
<html>
<head>
<title>Homes and Communities: U.S. Department of Housing and
Urban Development</title>
<meta name="dc.title" content="Homes and Communities">
<meta name="dc.title" content="U.S. Department of Housing and
Urban Development Home Page">
<meta name="dc.creator" content = "Departmental Web Team">
<meta name="dc.creator" content="Office of the Deputy Secretary">
<meta name="dc.date.created" content = "1995-04-15">
<meta name="dc.date.valid" content="2004-09-21">
<meta name="dc.language" content="en-us">
<meta name="dc.description" content="HUD's official website is a
clearinghouse of information and services about housing,
```

```
homebuying, renting, and community development for citizens and
for business partners.">

</head>
```

You'll notice we didn't use the Audience or Subject elements in this record. Audience and Subject will use a "controlled vocabulary"—a collection of responses we all agree to use—which hasn't yet been created.

2 The Details

For those that are interested, here are the details for each of the elements. For ease of reference, the elements are listed in alphabetical order. For each element, you will find the following data:

2.1 Definition

The formal definition of the element (e.g., what do we mean by "creator" or "title" so that if someone else were to look at our definition and our metadata, they'd be able to understand what we mean).

2.2 Repeating

Some elements can have more than one entry. Therefore, they are repeatable. For example, if a resource has more than one "creator" you can repeat the element to show all creators.

2.3 Purpose

Why are we including this element? What is its purpose? How could/will it be used in the future?

2.4 Notes

Here's where you will find additional information that might be useful when you go to use the element on your pages.

2.5 Not to be confused with

This part of the record will explain, where appropriate, what the difference between this and another element.

2.6 Examples

Examples of appropriate or correct entries for a given element. Examples are used in an informal way and are fictitious, as they are only intended to demonstrate the meaning or refinement of the element.

2.7 HTML syntax

This is where you will find the actual HTML code to put into the <head></head> section of your web page.

2.8   Value Domain

Examples and rules for valid entries. If it's text, how many words, if it's a number, what's the format, etc.

2.9   Validation

These are the rules that will be used to determine whether or not the metadata was created correctly or not. For example, the date for dc.date.created must be equal to or earlier than the date for dc.date.valid.

2.10 Mapped to

There are other metadata registries and schemas throughout the world. Where the elements express concepts that are the same in other registries, we've tried to "map them" one for one so that others can see the similarities. The other schemas compared are:

- DoD 5015.2-STD
- e-GMS: used by the United Kingdom
- GILS: Government Information Locator Service
- NARA LCDRG

3 The Elements

3.1   dc.audience

Definition	A class of entity for whom the resource is intended or useful.
Repeating	Yes
Purpose	To allow searches for information based on type of audience. This element may also be used for creating cross-agency searches based on audience.
Notes	The value for this element should be selected from a controlled vocabulary. The Department of Education has one controlled vocabulary (found at http://www.ed.gov/admin/reference/index.jsp); HUD will create one for our use as well. Controlled vocabularies should be harmonized wherever possible.
Not to be confused with	-
Examples	Students News Media Homebuyers Researchers
HTML syntax	<meta name="dc.audience" content="students"> <link rel=schema.dc href="http://purl.org/dc/terms/audience"> Note: for multiple audiences, it is acceptable to use a single <meta> line and separate entries with a semicolon such as: <meta name="dc.audience" content="students; researchers"> <link rel=schema.dc href="http://purl.org/dc/terms/audience">
Value Domain	The text value must not exceed 100 words nor contain restricted characters.

Validation	The value of this element is presumed to be "all audiences" under certain conditions: if this element is absent; or if the value of this element is empty, spaces, or null.
Mapped to	Dublin Core: Audience

3.2   dc.creator

Definition	An entity primarily responsible for making the content of the resource.
Repeating	Yes
Purpose	Enables the user to find resources that were written or otherwise prepared by a particular individual or organization. Also can be used to find the individual or organization that "owns" the content for maintenance purposes.
Notes	Using the job title rather than a person's name enhances the ability to locate information, although personal names may be needed for legal purposes and/or audit trails. The "creator" element is further enhanced when the full organizational hierarchy and full contact information are provided. Since acronyms may not be well known, it is best to use the full official title or cross-reference an appropriate glossary or explanatory note. Cross-agency portals should use the "creator" element to list the primary sponsoring agency or agencies who manage the website.
Not to be confused with	-
Examples	Office	U.S. Department of Justice, Federal Bureau of Investigation (FBI), Records Management Division
	Office	U.S. National Archives and Records Administration, Office of Records Services – Washington, DC, Modern Records Program
	Person	John Carlin, Archivist of the United States
HTML syntax	<meta name="creator" content="John, Carlin, Archivist of the United States"> <meta name="creator" content="National Archives and Records Administration, Office of the Archivist"> Note: for multiple creators, separate entries with a semicolon.
Value Domain	For personal author names, the text value is not required to be "normalized" (i.e., structured according to lexical rules

	distinguishing family name, honorific, etc.).
Validation	The set of metadata is incomplete under certain conditions: if this element is absent or if the value of this element is empty, spaces, or null.
Mapped to	DoD 5015.2-STD – Contributor, Creator
	Dublin Core – Contributor, Creator
	e-GMS – Contributor, Creator
	GILS – Author (Corporate name)
	NARA LCDRG – Contributor

3.3   dc.date.created

Definition	Date of creation of the resource.
Repeating	No
Purpose	To show the date the information resource was "created." Among the many uses of this element are to determine how long the information resource has been available, the interval between when the resource was created and when it was last reviewed, or when the content should expire.
Notes	The organization or individual listed in the "creator" element will usually determine the date created. This date may not necessarily reflect the date the information resource was actually created as there are often resources that are created but embargoed until a certain date and time. Information that was created and made publicly available in another format (e.g., a speech given by a prominent official) at an earlier date may carry the date the information was first made available (as an oral presentation or transcript) and not the date the web version was made available. Suggestions for information resources which were created before the release of this implementation guide include: • the date created (if known) • the date the information resource was last re-designed • the date the information resource was last reviewed Pay particular attention to the value domain field below. If a date does not include the month in a two-digit format (e.g., January=01), the data for this element will be considered invalid.
Not to be confused with	Date Valid—The "date valid" element identifies the last time the information resource was updated or reviewed to ensure correctness and currency.
Examples	2002-12-02
HTML syntax	<meta name="date.created" content="2002-12-02">
Value Domain	Date is represented in "YYYY-MM-DD" format, one of the ISO 8601 formats, consisting of the four digit Gregorian

	calendar year (YYYY), the two digit month (MM) valued from 01 to 12, and the two digit day (DD) valued from 01 to 31.
Validation	The set of metadata is incomplete under certain conditions: • if this element is absent or occurs more than once • if the value of this element is empty, spaces, or null • if the date is not presented in the correct yyyy-mm-dd format • or if the "date created" value is a date later than any "date reviewed" value when present.
Mapped to	Dublin Core—Created, dateCreated e-GMS—date.created

3.4   dc.date.valid

Definition	Date of validity of a resource.
Repeating	No
Purpose	To show the date the information resource was last reviewed and certified current and accurate.
Notes	To comply with HUD's policies and standards, each web page should be reviewed at least once every 3 months to ensure it is current and accurate.
	This date can (and probably should) be the same as the date displayed on the page that shows when the page was last reviewed.
	Pay particular attention to the value domain field below. If a date does not include the month in a two-digit format (e.g., January=01), the data for this element will be considered invalid.
Not to be confused with	dc.date.created—The date created element is used to identify when the web page was originally made available. Date Valid and Date Created could, technically, be the same date. However, the intention of the "date reviewed" element is to record when the last review, after creation, occurred.
Examples	2003-01-01
HTML syntax	<meta name="date.reviewed" content="2003-01-01">
Value Domain	Date is represented in "YYYY-MM-DD" format, one of the ISO 8601 formats, consisting of the four digit Gregorian calendar year (YYYY), the two digit month (MM) valued from 01 to 12, and the two digit day (DD) valued from 01 to 31.
Validation	The set of metadata is incomplete if: • the date is not presented in the correct yyyy-mm-dd format • the "date reviewed" value is a date earlier than any "date createded" value when present.
Mapped to	Dublin Core: Date Valid

13

3.5 dc.description

Definition	An account of the content of the resource.
Repeating	No
Purpose	This is the text describing the web page. The text should help the user decide if it fits their needs.
Notes	Description should use complete words and phrases that describe the subject or contents of the information resource.
	The description is often used as the text returned by search engines to give the user a sense of what is available on the information resource.
	Description may include but is not limited to: an abstract, table of contents, or a free-text account of the content.
Not to be confused with	Do not confuse with dc.subject, which should use keywords preferably selected from a controlled vocabulary.
Examples	Definitions of common metadata as applied to government information resources, agreed for the U.S. Federal Government under the E-Government Act of 2002.
HTML syntax	<meta name="description" content="Definitions of common metadata as applied to government information resources, agreed for the U.S. Federal Government under the E-Government Act of 2002.">
Value Domain	The text value must not exceed 100 words or contain restricted characters.
Validation	The set of metadata is incomplete under certain conditions: if this element is absent; or if the value of this element is empty, spaces, or null.
Mapped to	e-GMS—Description
	GILS—Abstract
	Dublin Core—Description

3.6   dc.language

Definition	A language of the intellectual content of the resource.
Repeating	Yes
Purpose	Enables users to limit their searches to pages in a particular language.
Notes	This element is repeatable to indicate when more than one language is present. The code set for the metadata is assumed to be Latin-1. Setting this element should indicate the ability of the speaker of such a language to extract useful content, rather than simply the appearance of a word or phrase from a given language.
Not to be confused with	-
Example	For a web page written in English: Language: eng For a resource written in Spanish: Language: spa For a resource written in both Spanish and English: Language: eng; spa
HTML syntax	<meta name="language" content="eng"> <meta name="language" content="spa"> <meta name="language" content="eng; spa">
Value Domain	The text value must exactly match one of the natural language identifiers listed in ISO 639-2 which can be found at http://www.loc.gov/standards/iso639-2/englangn.html.
Validation	The value of this element is presumed to be "eng" under certain conditions: if this element is absent; or if the value of this element is empty, spaces, or null.
Mapped to	Dublin Core—Language GILS—Language of resource e-GMS—Language

3.7   dc.subject

Definition	The topic of the content of the resource.
Repeating	Yes
Purpose	Enables users to limit their searches to resources in a particular language.
Notes	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. We will use a controlled vocabulary (which has not been created yet). Department of Education has a controlled vocabulary found at http://www.ed.gov/admin/reference/index.jsp
Not to be confused with	Distinguish from dc.description, which is a plain text description of the web page.
Example	Homebuying Renting Public Housing
HTML syntax	<meta name="subject" content="Public Housing"> <meta name="subject" content="Rental Assistance"> <meta name="subject" content="Public Housing; Rental Assistance">
Value Domain	The text value must match one of the choices from a controlled vocabulary.
Validation	The value of this element is presumed to be invalid under certain conditions: if this element is absent; or if the value of this element is empty, spaces, or null.
Mapped to	Dublin Core—Subject GILS— e-GMS—Subject

3.8 dc.title

Definition	A name given to the resource (web page).
Repeating	Yes
Purpose	Enables the user to find a resource with a particular title or carry out more accurate searches.
Notes	If the information resource does not have a formal title, the creator should establish a meaningful title that is user oriented and brief. Title is commonly used as a key reference result in lists of search results. There can be multiple titles for the same information resource. For example, there might be a title for the informal name of the website (HUD home page), another title element for the formal (U.S. Department of Housing and Urban Development Home Page).
Not to be confused with	-

Examples	Web site title	U.S. Department of Housing and Urban Development Home Page
	Web site title	Homes and Communities
	[document](#)	Application by Thomas McCarthy for Admission to Western Branch Soldier's Home

HTML syntax	<meta name="title" content="U.S. Department of Housing and Urban Development Home Page"> <meta name="title" content="Homes and Communities">
Value Domain	The text value must not exceed 100 words nor contain restricted characters.
Validation	The set of metadata is incomplete under certain conditions: if this element is absent; or if the value of this element is empty, spaces, or null.
Mapped to	Dublin Core—Title e-GMS—Title GILS—Document Title

Appendix: Links and References

Standards and other organizations

- **DCMI** (Dublin Core Metadata Initiative) – http://www.dublincore.org

- **DoD 5015.2-STD** -- http://www.dtic.mil/whs/directives/corres/html/50152std.htm

- **e-GMS** – e-Government Metadata Standard (United Kingdom)
  http://e-government.cabinetoffice.gov.uk/assetRoot/04/00/78/86/04007886.pdf

- **ISO 639** (codes for languages) -- http://www.w3.org/WAI/ER/IG/ert/iso639.htm

- **ISO 11179** (metadata standards) -- http://metadata-stds.org/11179/index.html

- **ISO 8601** (recording date and time) -- http://www.iso.ch/iso/en/prods-services/popstds/datesandtime.html

- **GILS** (Government Information Locator Service) –
  http://www.access.gpo.gov/su_docs/gils/