

Taming the World “Wild” Web with Metadata for Everyone

Joan Weeks

Cooperative and Instructional Programs Division, Library of Congress, Washington, DC,
United States.

E-mail address: jwee@loc.gov



Copyright © 2015 by Joan Weeks. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Abstract:

Librarians create files that are uploaded on their library websites with little or no metadata which makes the content hard for search engines to find and for patrons with visual challenges using screen readers. This is particularly difficult as projects are underway to convert these files to new formats or just add them to digital archives. Many librarians believe that only a high-tech computer specialist could create and preserve web files with metadata. With the amount of work and lack of resources everyone needs to be involved! This presentation will demonstrate how everyone can add metadata quickly and easily to MS Word, PowerPoint, PDFs and basic webpages, as well as discuss how to set up guidelines for the best practices for accessibility.

Keywords: metadata libraries digital archives accessibility

Did you ever wonder why search engines rank many web resources at the bottom of the search results? What chance does a hit that is 53 millionth have of seeing the light of day? Most people fail to retrieve websites that are more than 10 hits down in the list.

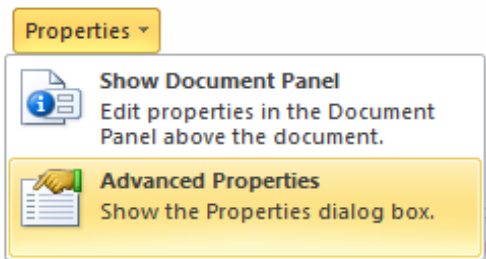
On a daily basis librarians create files that are uploaded on their library websites with little or no metadata which makes the content hard for patrons to find and even more so for individuals with visual challenges using screen readers. The lack of good metadata presents even more hurdles as projects are underway to convert these files to new formats or just add them to digital archives. Recently, in our division at the Library of Congress, in migrating documents to a new website I needed to identify who had created the documents to have them help in updating them. Many times it was difficult to find the creator and organize the material in categories so they could be searched and retrieved.

Research Methodology:

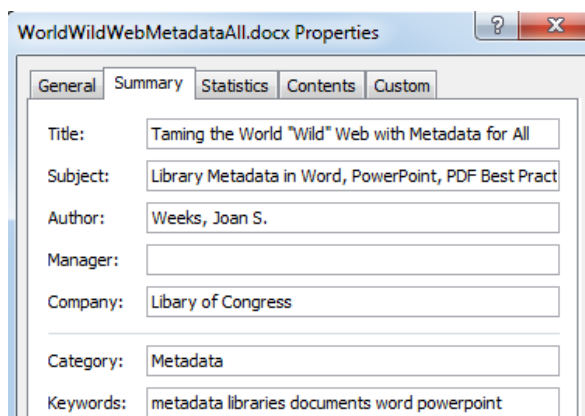
In the Google Advanced search the terms **library metadata** were entered and then MS Word, PowerPoint, PDF were selected for the file types. For MS Word, Yale University came up to the top of the results and their Word Document about Library Metadata Best Practices was exemplar. However, they could have added a subject and keywords in their metadata in addition to the author and title they included. After going through 200 hits to find result with a much lower ranking one from another prestigious university was reviewed and found to have used “name of Librarian” as the title in the metadata and “preferred customer” as the author with no subject or keywords. These terms had nothing to do with the title of their document.

Many librarians believe that only a high-tech computer specialist could create and preserve web files with metadata. Job titles such as metadata librarian or systems librarian only reinforce this perception. With the amount of work and lack of resources everyone needs to be involved! So let's get started.

In MS Word, select **File>>Info** and then select **Properties** and **Advanced Properties**.



In the properties window select the **Summary** tab. Complete the title, subject, author and keywords metadata at a minimum.



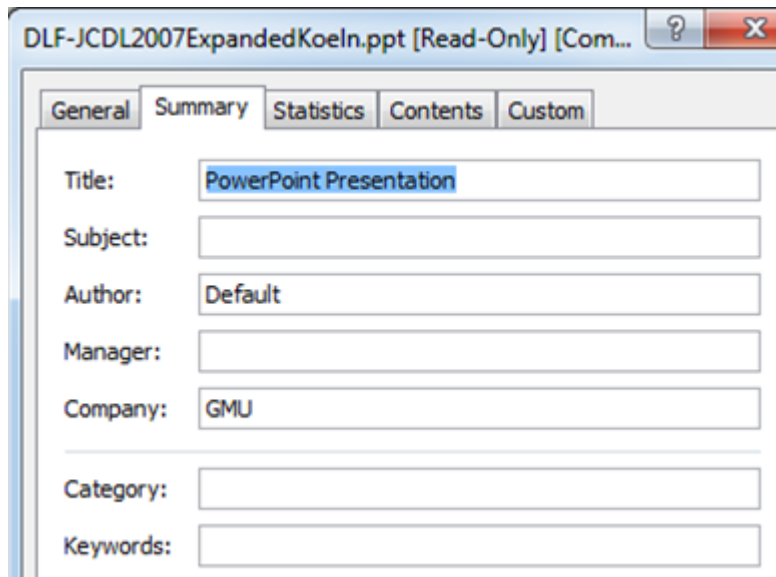
PowerPoint presentations are a popular addition to library websites. Using the same advanced search technique in Google only this time specifying PowerPoint as the file type a similar situation occurred. Hit 191 seemed like a very interesting presentation entitled “Digital Library Content Model” but most people would never see it because it was so far down in the results.

[PPT] [Digital Library Content Model.](#)

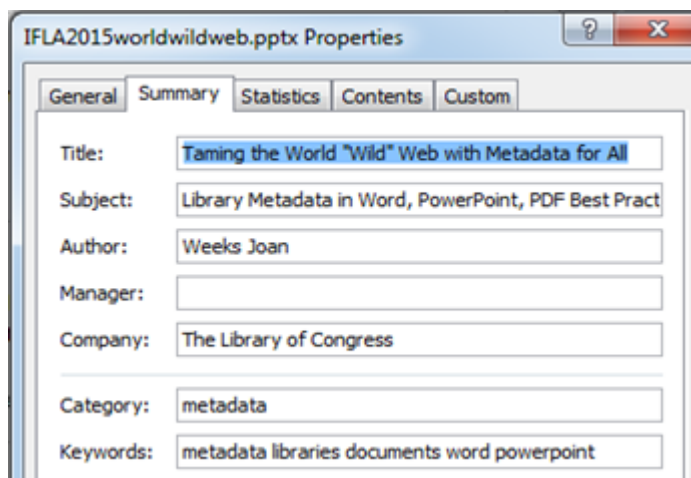
www.dsoergel.com/DLF-JCDL2007ExpandedKoeln.ppt ▼

This requires a good metadata model; Support annotation of information objects. ... of basic notions regarding the content of a digital library (or, more generally, ...

Why the low ranking? The title was PowerPoint Presentation and the author was default.



Putting metadata in PowerPoint follows the same methodology as MS Word. In MS Word, select **File>>Info** and then select **Properties** and **Advanced Properties**. In the properties window select the **Summary** tab. Complete the title, subject, author and keywords metadata elements and the category if many items on various topics are going on your website.



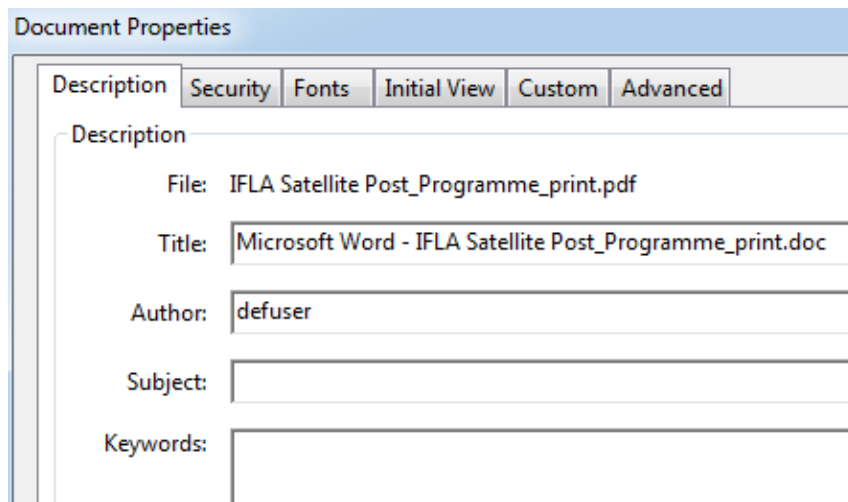
Portable document format (PDF) documents are widely used to preserve the original appearance of the content and to prevent it from being altered. Many libraries put PDFs on their websites for procedures and public announcements, news releases etc. When

using the library metadata terms in the advanced search in Google and specifying PDF for file type, an IFLA hit on the 19th page of results out of 1,040,000 appeared.

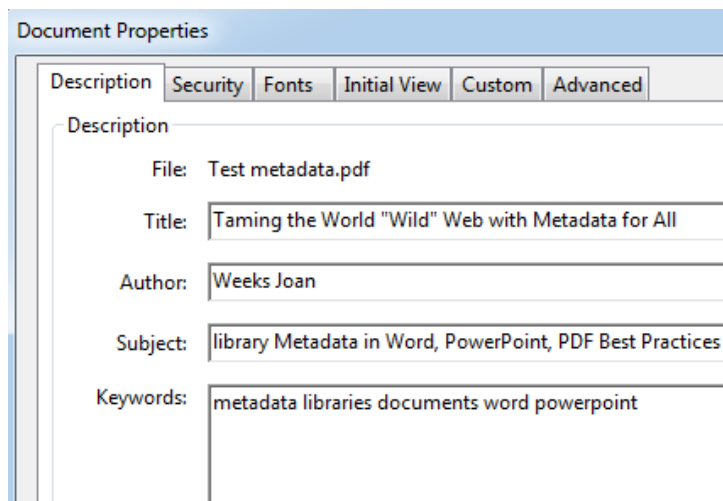
[\[PDF\] Beyond libraries – subject metadata in the digital environment and ... www.nlib.ee/.../IFLA%20Satellite%20Post_Programme_print.pdf](#) ▼

IFLA Satellite Post-Conference. Beyond **libraries** – subject **metadata** in the digital environment and semantic web. 17-18 August 2012, Tallinn. Programme.

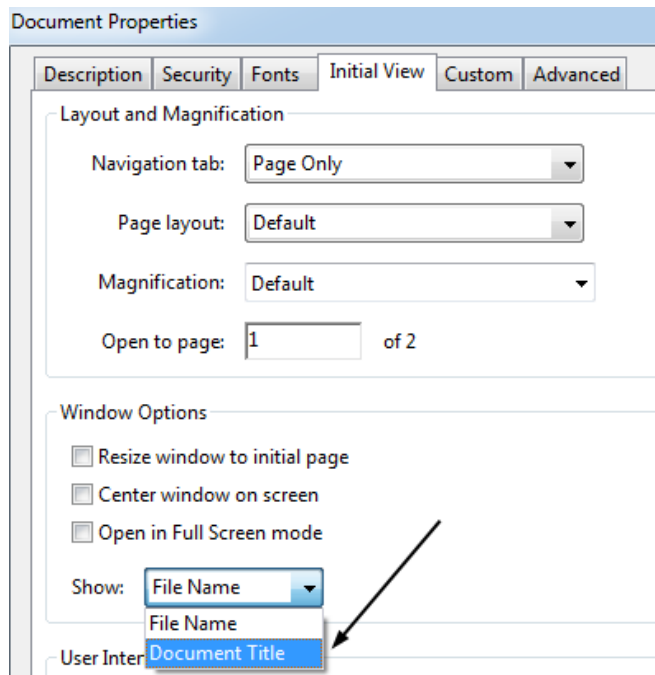
In reviewing the metadata, we see the file name used as the title and no author, subject or keywords.



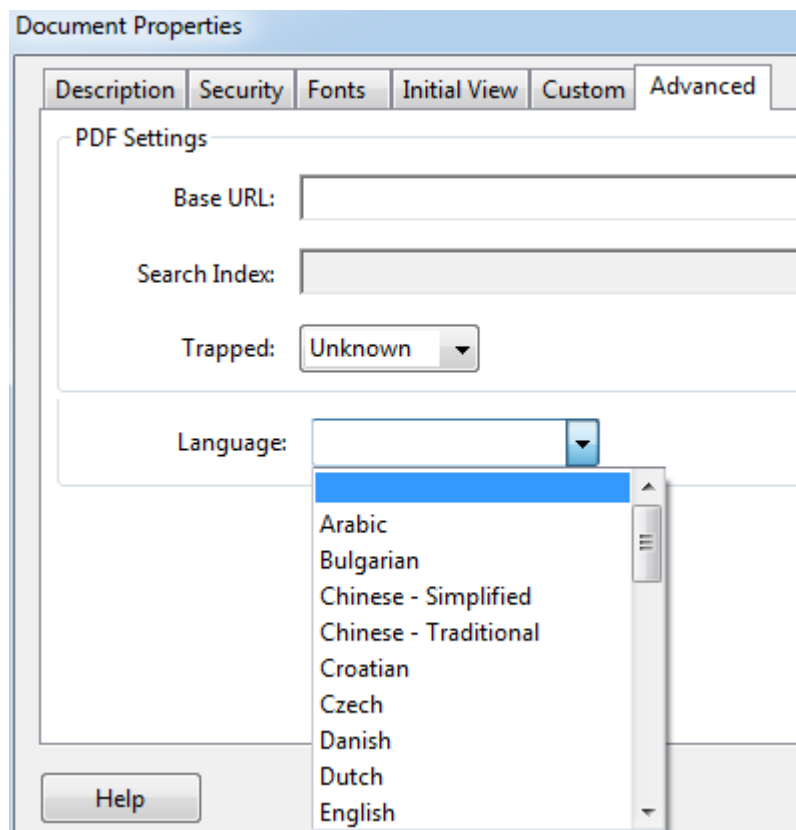
To put metadata in a PDF document a program like Adobe Acrobat Pro is required. With your PDF document displayed select **File** and **Properties** then the **Description** tab.



Enter the title, author, subject and keywords metadata. Next click on the Initial View tab and change the file name to document title. This document title will appear on the tab of the browser as well as raise the ranking in the search results.



Click on the **Advanced** tab and specify your language.



All of these MS Word, PowerPoint and PDF documents are linked up on your library webpage and it is very important to insure that the metadata is in place. With Notepad you can easily insert metadata into your library's webpages. Open the webpage in Notepad.

1. Type `<!doctype html>`
2. Type `<html lang="en">` or your language.
3. Type `<head>`
4. Type `<title>put your title</title>`
5. Type `<meta charset="utf-8">`
6. Type `<meta name="description" content="Your description">`
7. Type `<meta name="keywords" content="Your keywords, other keywords">`
8. Type `</head>`
9. Type `<body> complete webpage</body></html>`

As you build your metadata repertoire you may want to move on to Dublin Core Metadata Elements: <http://dublincore.org/documents/dces> as these are a widely used standard.

```

1
2
3 <!DOCTYPE html>
4 <html lang="en" class="no-js">
5 <head>
6     <meta charset="utf-8">
7     <title>
8     Library of Congress Home
9     | Library of Congress</title>
10
11     <meta name="description"
12         content="The Library of Congress. The
13         Library of Congress is the nation's oldest
14         federal cultural institution, and it serves as
15         the research arm of Congress. It is also the
16         largest library in the world, with more than
17         120 million items. The collections include
18         books, sound recordings, motion pictures,
19         photographs, maps, and manuscripts." />
20
21     <meta name="dc.identifier"
22         content="http://www.loc.gov/" />
  
```

Best Practices

Many libraries prepare a Best Practices manual to keep metadata consistent across the institution, to enhance retrieval and provide a uniform user experience. Some practices that the Library of Congress uses are at: <http://www.loc.gov/standards/mdc/docs/html-metatags.pdf>. Included are the Dublin Core elements that are required and also list of the DC tags that LC does not use such as DC.description and the use instead form `<meta name="description">`

Every librarian can tame the World “Wild” web with good metadata in every document that is placed on their library’s website.

Acknowledgments

LC Metadata for Digital Content (MDC) group