

## Dompter le monde “sauvage” du web avec des métadonnées pour tous

*French translation of the original paper: “Taming the World “Wild” Web with Metadata for Everyone”.*

*Translated by: Nathalie Yakovleff, Gan Assurances, Paris La Défense, France.*

*Le texte de ce document a été traduit de l’anglais, des différences avec le texte original peuvent exister. Cette traduction est uniquement fournie à titre d’information.*

### **Weeks Joan**

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

E-mail address: [jwee@loc.gov](mailto:jwee@loc.gov)



This is a French translation of “Taming the World “Wild” Web with Metadata for Everyone” copyright © 2015 by **Nathalie Yakovleff**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

---

### **Résumé :**

*Les bibliothécaires créent des fichiers à télécharger sur leurs sites web comprenant peu ou aucune métadonnée, ce qui rend leur contenu difficile à trouver par les moteurs de recherche et par les usagers déficients visuels qui utilisent des lecteurs d’écran. Ceci est aussi particulièrement compliqué pour des projets en cours visant à convertir des fichiers dans de nouveaux formats ou simplement les ajouter à des archives numériques. De nombreux bibliothécaires croient que seul un spécialiste high-tech de l’informatique peut créer et conserver des fichiers web avec des métadonnées. Compte-tenu de la quantité de travail et du manque de ressources, nous devons tous nous impliquer ! Cette présentation va démontrer comment chacun peut ajouter rapidement et facilement des métadonnées à des documents MS Word, PowerPoint, Pdf et à de simples pages web et expliquera comment mettre en place des instructions de bonnes pratiques pour l’accessibilité.*

**Mots-clés :** métadonnées bibliothèques archives numérique accessibilité

---

Vous êtes-vous jamais demandé pourquoi les moteurs de recherche classent beaucoup de ressources web en fin de liste de résultat ? Quelle chance a une réponse qui se classe 53millionième de voir la lumière du jour ? La plupart des gens négligent les sites web qui dépassent la 10<sup>ème</sup> place de la liste.

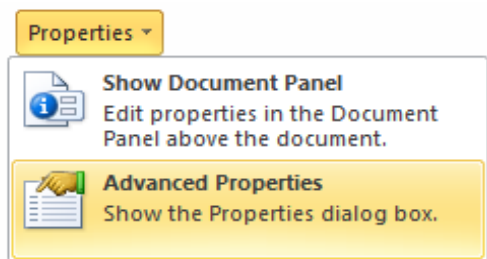
Les bibliothécaires créent quotidiennement des fichiers qui sont téléchargés sur leurs sites web et qui contiennent peu ou pas de métadonnées, ce qui rend leurs contenus difficiles à trouver et encore plus encore pour les personnes déficientes visuelles qui utilisent des lecteurs d'écran. Le manque de bonnes métadonnées constitue un véritable obstacle quand des projets sont entrepris pour convertir ces fichiers dans de nouveaux formats ou pour simplement les ajouter à des archives numériques. Récemment, en migrant des documents sur un nouveau site web pour notre département de la bibliothèque du Congrès, j'ai eu besoin d'identifier qui les avait créés afin de pouvoir les mettre à jour. Il fut plusieurs fois difficile de trouver le créateur et d'organiser le corpus en catégories permettant de les rechercher et de les retrouver.

## Méthodologie de recherche

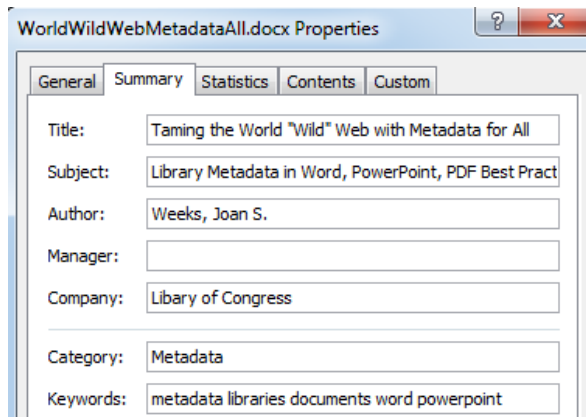
Dans la Recherche avancée de Google les termes « **library metadata** » (= bibliothèque métadonnées) ont été saisis et les types de fichiers MS Word, PowerPoint, Pdf sélectionnés. Pour MS Word, l'Université de Yale est arrivée en tête des résultats et leur document Word sur Les bonnes pratiques en matière de métadonnées en bibliothèque était exemplaire. Cependant, ils auraient pu ajouter un sujet et des mots-clés dans leurs métadonnées en plus de l'auteur et du titre qui étaient eux indiqués. Après avoir parcouru 200 résultats et trouvé plus loin des résultats d'une autre prestigieuse université et j'en ai vu comprenant « nom du bibliothécaire » comme titre dans les métadonnées et « client préféré » en auteur, sans sujet ni mot-clé. Ces termes n'avaient rien à voir avec le titre de leur document.

De nombreux bibliothécaires croient que seul un informaticien spécialiste high-tech est capable de créer et de conserver des fichiers web avec des métadonnées. Des intitulés de métier comme Bibliothécaire métadonnées ou Bibliothécaire système accentuent cette perception. Compte tenu de la quantité de travail et du manque de ressources chacun doit être impliqué ! Commençons-donc !

En MS Word, dans **Fichier>>Info**, cliquer sur **Propriétés** puis **Propriétés avancées**



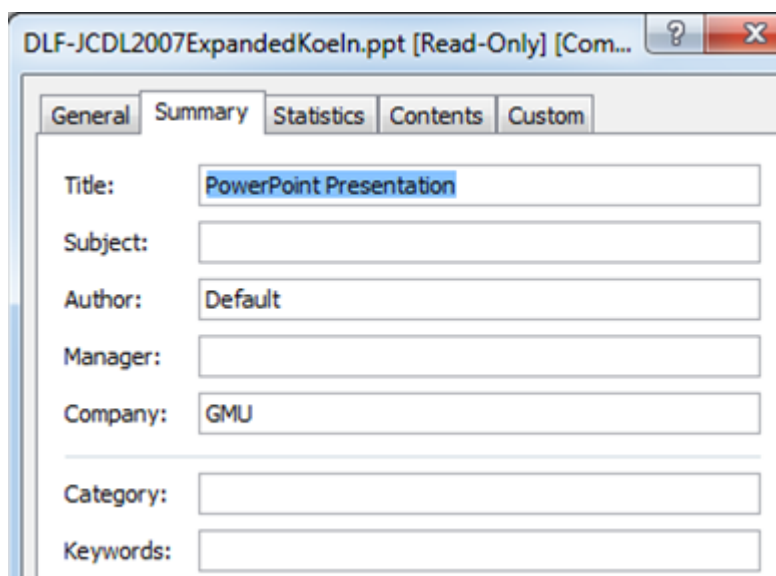
Dans la fenêtre de propriétés, sélectionner l'onglet Résumé. Compléter au minimum les métadonnées titre, sujet, auteur et mots-clés.



Les présentations PowerPoint sont très répandues dans les sites web des bibliothèques. En utilisant la même technique dans la Recherche avancée de Google, et en précisant cette fois PowerPoint en type de fichier, la situation précédente se reproduit. Le résultat 191 semble être une présentation très intéressante intitulée « Modèle de contenu pour une bibliothèque numérique » mais la plupart des gens ne la verra jamais parce qu'elle est trop loin dans la liste de résultats.

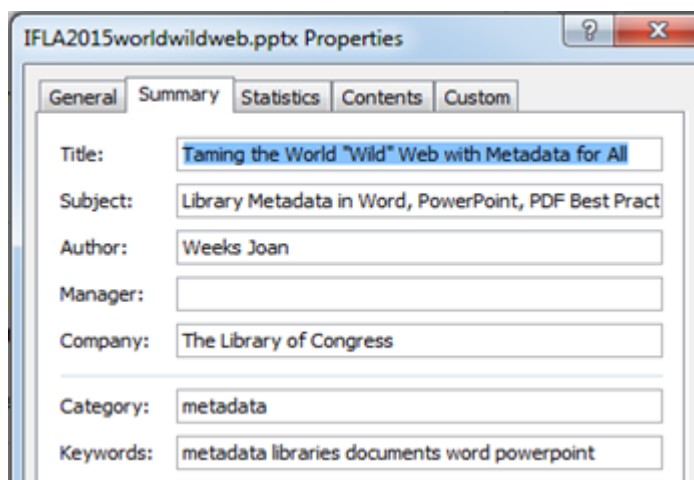
[IPPT Digital Library Content Model.](http://www.dsoergel.com/DLF-JCDL2007ExpandedKoeln.ppt)  
[www.dsoergel.com/DLF-JCDL2007ExpandedKoeln.ppt](http://www.dsoergel.com/DLF-JCDL2007ExpandedKoeln.ppt) ▼  
 This requires a good metadata model; Support annotation of information objects. ... of basic notions regarding the content of a digital library (or, more generally, ...

Pourquoi ce mauvais classement ? Le titre était Présentation PowerPoint et le nom de l'auteur absent.



Pour ajouter des métadonnées dans PowerPoint, la méthodologie est la même que pour MS Word. Dans PowerPoint sélectionner **Fichier>>Info**, cliquer ensuite sur **Propriétés** puis **Propriétés avancées**. Dans la fenêtre de propriétés cliquer sur l'onglet **Résumé**. Compléter le

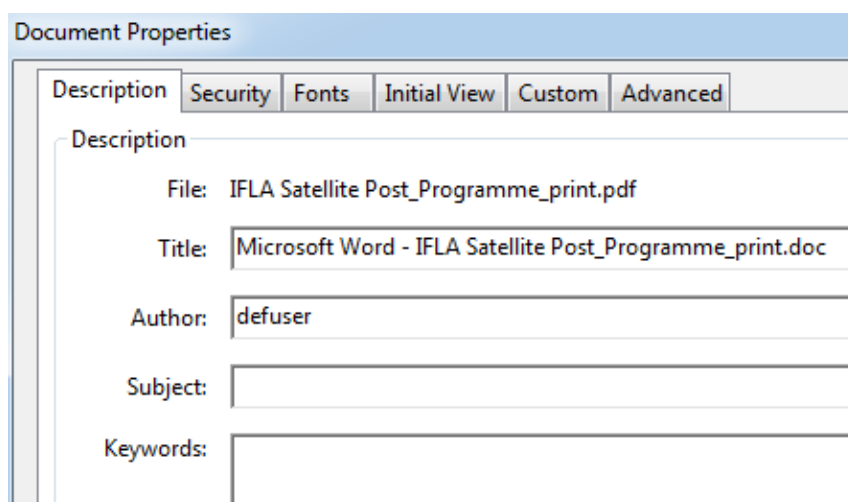
titre, le sujet, l'auteur et ajouter les métadonnées et la catégorie si plusieurs items sur des sujets différents sont traités sur votre site web.



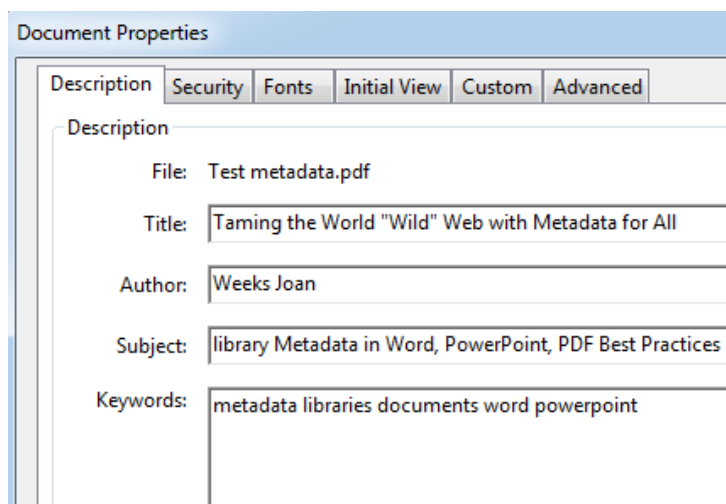
Les documents au format Pdf (Portable Document Format) sont largement utilisés pour préserver l'apparence originale des contenus et empêcher les modifications. De nombreuses bibliothèques utilisent le pdf sur leurs sites web pour la publication de procédures, d'avis publics, des communiqués etc. Quand on saisit les termes « library metadata » (= bibliothèque métadonnées) dans la Recherche avancée de Google et que l'on spécifie Pdf en type de format, un résultat de l'IFLA arrive en 19<sup>e</sup> position sur les 1 040 000 qui s'affichent.

[\[PDF\] Beyond libraries – subject metadata in the digital environment and ...  
www.nlib.ee/.../IFLA%20Satellite%20Post\\_Programme\\_print.pdf](http://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.

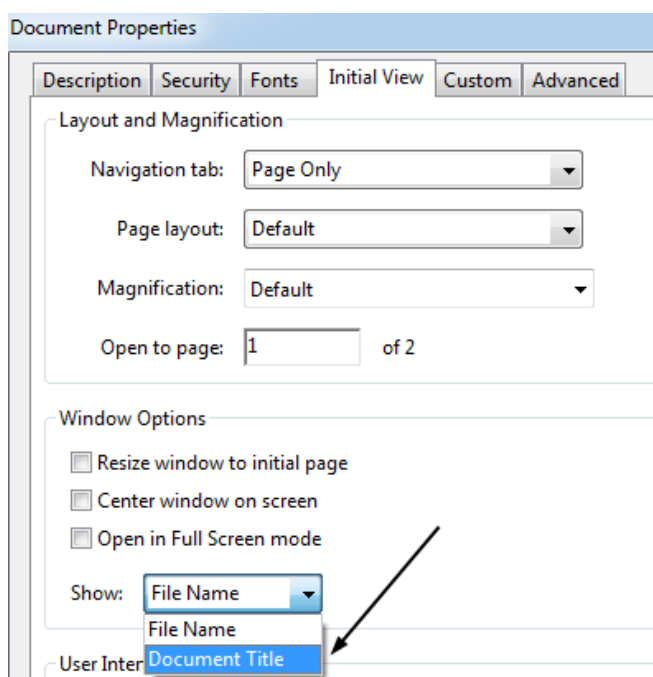
Quand on regarde les métadonnées, nous voyons que le nom du fichier est indiqué en titre et qu'il n'y a ni auteur, ni sujet, ni mot-clé.



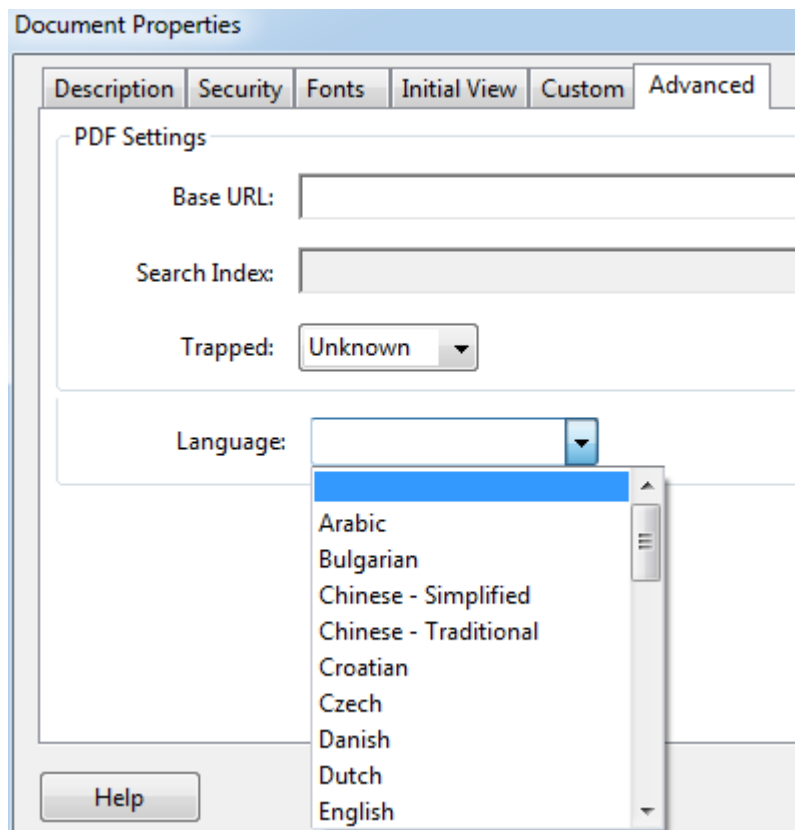
Pour ajouter des métadonnées dans un document Pdf un programme comme Adobe Acrobat Pro est nécessaire. Affichez votre document Pdf et sélectionnez **Fichier** puis **Propriétés** et ensuite cliquez sur l'onglet **Description**.



Saisir le titre, l'auteur, le sujet et les métadonnées de mot-clé. Cliquer ensuite sur l'onglet Vue initiale et dans Affichage sélectionner Titre du document. C'est ce titre qui apparaîtra dans l'onglet du navigateur et qui fera remonter le document dans la liste de résultats.



Cliquez sur l'onglet **Avancé** et spécifiez votre langue.



Tous ces documents MS Word, PowerPoint et Pdf sont reliés à une page web du site de votre bibliothèque, il est donc très important de s'assurer qu'elles comportent elles-mêmes des métadonnées. Avec Notepad vous pouvez facilement en insérer. Ouvrir la page web dans Notepad et compléter le code ci-dessous.

1. Type **<!doctype html>**
2. Type **<html lang="en">** ou votre langue
3. Type **<head>**
4. Type **<title>insérer votre titre</title>**
5. Type **<meta charset="utf-8">**
6. Type **<meta name="description" content="Votre description">**
7. Type **<meta name="keywords" content="Vos mots-clés, d'autres mots-clés">**
8. Type **</head>**
9. Type **<body> saisir la page web</body></html>**

Quand vous construisez votre répertoire de métadonnées vous pouvez vouloir passer à l'utilisation des éléments de métadonnées du Dublin Core : <http://dublincore.org/documents/dces> car la norme est largement utilisée.

```
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/" />
```

## Bonnes pratiques

De nombreuses bibliothèques élaborent un manuel de Bonnes Pratiques afin de maintenir la cohérence des métadonnées dans toute leur institution, d'améliorer la recherche et de mettre à la disposition des utilisateurs une base harmonisée. Quelques pratiques mises en place par la Bibliothèque du Congrès sont consultables ici <http://www.loc.gov/standards/mdc/docs/html-metatags.pdf>. On y trouve les éléments indispensables du Dublin Core et aussi la liste des balises DC que la Bibliothèque n'utilise pas comme DC.Description et l'utilisation de la forme alternative <meta name= « description »>

Chaque bibliothécaire peut dompter le monde « sauvage » du web grâce à de bonnes métadonnées sur chaque document qui figure sur le site web de sa bibliothèque.

## Remerciements

LC Metadata for Digital Content (MDC) group<sup>1</sup>

---

<sup>1</sup> Groupe de travail de la Bibliothèque du Congrès sur les contenus numériques