

RFC 5013 : The Dublin Core Metadata Element Set

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 1 septembre 2007. Dernière mise à jour le 21 septembre 2007

Date de publication du RFC : Août 2007

<https://www.bortzmeyer.org/5013.html>

Mise à jour, voici la liste des éléments qui forment le fameux "*Dublin Core*", un ensemble d'éléments pour créer des métadonnées, notamment sur les ressources Web.

"*Dublin Core*", qui doit son nom à la ville de Dublin, où s'est tenue la réunion fondatrice, est un jeu de quinze éléments, qui permettent d'attribuer des métadonnées à une **ressource**, typiquement un fichier accessible via le Web.

Ces quinze éléments comprennent, entre autres :

- "*title*", le titre de la ressource,
- "*identifier*", un identificateur unique pour la ressource,
- "*date*", la date de publication de la ressource,
- "*language*", la langue de la ressource,
- etc.

On notera que "*Dublin Core*" ne spécifie pas de **sémantique** rigoureuse pour ces éléments. C'est volontaire, afin de donner le maximum de liberté à ceux qui définissent ces métadonnées. Le RFC se contente de donner des conseils, comme d'utiliser un « système d'identificateurs formels » pour l'élément "*identifier*".

De même, la syntaxe n'est pas spécifiée, uniquement conseillée. Ainsi, le RFC suggère d'utiliser les étiquettes de langue du RFC 4646¹ pour "*language*" ou bien ISO 8601 pour les dates.

De même, la façon dont se représentent les éléments Dublin Core dans la ressource n'est pas précisée dans ce RFC. Elle dépend en effet du format de cette dernière. On peut trouver une liste de techniques en

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4646.txt>

<<http://dublincore.org/resources/expressions/>> et un exemple pour HTML en <<http://dublincore.org/documents/2008/08/04/dc-html/>>.

"Dublin Core" avait été à l'origine normalisé dans le RFC 2413, auquel notre RFC succède. Il n'y a pas de changement radical entre les deux RFC. Il existe aussi une norme ISO sur Dublin Core, la 15836 mais, comme elle n'est pas publique, le RFC est une meilleure source.

Très peu de pages Web publient du "Dublin Core" aujourd'hui. Sur ce blog, par exemple, vous n'en trouverez pas dans les pages HTML mais il y en a dans le flux de syndication Atom. En voici un extrait :

```
<feed xml:lang="fr" xmlns="http://www.w3.org/2005/Atom"
      xmlns:html="http://www.w3.org/1999/xhtml"
      xmlns:dublincore="http://purl.org/dc/elements/1.1/">
<!-- See for discussions about the relationship between Atom and Dublin Core
      and why Atom does not use Dublin Core:
      http://internetalchemy.org/2004/03/theNucleusOfAtom
      http://www.imc.org/atom-syntax/mail-archive/msg03170.html
      http://www.imc.org/atom-syntax/mail-archive/msg04474.html
      http://bitworking.org/news/Not_Invented_Here
-->
...
<dublincore:title>Blog de Stéphane Bortzmeyer</dublincore:title>
<dublincore:language>fr</dublincore:language>
<dublincore:identifiant>tag:bortzmeyer.org,2006-02:Blog</dublincore:identifiant>
<dublincore:date>2007-07-05T16:35:00Z</dublincore:date>
...
```

Autre exemple, le logiciel libre de dessin Inkscape permet d'associer des métadonnées Dublin Core à un document. Comme son format natif de stockage est SVG, Inkscape, logiquement, les exprime en XML dans l'espace de nommage Dublin Core (notons aussi l'utilisation des schémas de "Creative Commons") :

```
<svg
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:cc="http://web.resource.org/cc/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  ...
  <metadata>
    <rdf:RDF>
      <dc:format>image/svg+xml</dc:format>
      <dc:type
        rdf:resource="http://purl.org/dc/dcmitype/StillImage" />
      <dc:title>Forces Architecture</dc:title>
      <dc:date>2007-09-21</dc:date>
      <dc:creator>
        <cc:Agent>
          <dc:title>Stéphane Bortzmeyer</dc:title>
        </cc:Agent>
      </dc:creator>
      <dc:rights>
        <cc:Agent>
          <dc:title>GFDL</dc:title>
        </cc:Agent>
      </dc:rights>
      <dc:language>fr</dc:language>
      <cc:license
        rdf:resource="gfdl" />
      </cc:Work>
    </rdf:RDF>
  </metadata>
```

Voici quelques autres exemples d'utilisation aujourd'hui :

- L'Open Archive Initiative, protocole d'échanges de métadonnées entre des réservoirs d'articles scientifique en libre accès utilise entre autres le Dublin Core.
- TEF <<http://www.abes.fr/abes/documents/tef/>> (Thèses Electroniques Françaises) pour la description des thèses.
- L'excellent projet [revues.org](http://www.revues.org) <<http://www.revues.org>> d'édition électronique en sciences sociales.
- L'École des chartes <<http://elec.enc.sorbonne.fr>>.

C'est mince par rapport aux ambitions anciennes des systèmes de métadonnées.

De même, je ne connais pas de logiciel public (par exemple de moteur de recherche) qui puisse utiliser le "*Dublin Core*". Les seuls cas existants sont dans des mondes plus fermés comme celui d'In-[extenso](http://www.in-extenso.org) <<http://www.in-extenso.org>>, un moteur de recherche spécialisé en sciences humaines et sociales.

En dépit de gros efforts de sensibilisation, les métadonnées restent les grandes absentes du Web. Pour quelles raisons? Outre les raisons spécifiques à Dublin Core, comme le fait que beaucoup d'éléments Dublin Core font double emploi avec des éléments existants (<title> ou l'attribut lang en HTML, les mêmes plus <published> ou <author> en Atom), je vois deux raisons communes à toutes les métadonnées.

L'une est que les moteurs de recherche ont appris à se méfier des informations mises par l'auteur de la page Web. Soit il est peu soigneux, soit il triche en mettant des mots-clés susceptibles de lui attirer du trafic.

L'autre est un problème d'œuf et de poule classique. Les moteurs de recherche n'utilisent pas les métadonnées, donc les auteurs n'en mettent pas, donc les moteurs ne les utilisent pas...

Merci à Gautier Poupeau pour ses nombreuses informations.