



# **LES MÉTADONNÉES EN LIGNE, UNE SACRÉE AVENTURE**

LOU BALLANGÉ - LE SLIC - 2014 / 2015

Seconde feuille de route. Rédigée pour le 31 novembre 2014 par Lou Ballangé, dans le cadre du cours de Structuration et gestion des métadonnées, donné par M. Antoine Moreau.

## Sommaire

Introduction.....	1
Quelques de métadonnées.....	2
Dublin Core.....	2
RDF.....	2
SCORM/LOM.....	2
ID3.....	3
Wikimédia Commons et Commonist.....	4
Wikimédia Commons.....	4
Commonist.....	6
Big Data.....	7
Le site Internet.....	8
Les métadonnées des dossiers .pdf.....	8
Les schémas.....	9
Les balises <meta>.....	9
Les limites.....	11
Conclusion.....	12

## Introduction

Dans la feuille précédente, nous avons exploré l'utilisation de Zotero et de XnView, deux logiciels qui permettent l'extraction et l'ajout de métadonnées sur des photographies, des documents, des pages web, etc. Dans cette seconde partie de notre travail, nous allons nous pencher sur la conception d'une page Internet et à d'autres styles de métadonnées toutes aussi importantes.

Dans une première partie, nous aborderons certains schémas de métadonnées, les plus connus et usités. Nous allons parler des formats RDF, XML, Dublin Core et ID3. Puis nous allons nous pencher sur le site Wikimedia Commons, en aborder les points les plus importants, et nous intéresser également de Commonist. Enfin, nous allons évoquer le sujet des *Big Data*.

Dans une seconde partie, nous aborderons le sujet du site Internet qui nous a été demandé de créer. Il s'agit d'une unique page d'accueil, une introduction aux métadonnées et deux documents au format .pdf (cette feuille de route et la précédente). Dans la création de la page Internet, le travail va se porter sur les métadonnées, les balises <meta>, présentes dans le <head> du code HTML.

Le but de cette seconde feuille de route est de démontrer l'importance des métadonnées dans un travail d'intégration de contenu et de mise en ligne.

## Quelques de métadonnées

### Dublin Core

Le Dublin Core est à la base de la structuration des métadonnées et des informations mises en ligne à l'heure actuelle. Il s'agit d'un schéma qui permet la description précise de ressources numériques ou physiques. Le Dublin Core fait partie intégrante de l'ISO, intégré dans la norme internationale ISO 15836 depuis 2003.

### RDF

Le format RDF (*Ressource Description Framework*) est un format de métadonnées qui enregistre la structure et le contenu des informations présentes sur le net, contrairement au XML qui n'enregistre que la structure. Le traitement des informations via le RDF permet de mieux les organiser, et le référencement par les moteurs de recherche est de ce fait plus efficace.

Le RDF possède plusieurs extensions et versions. L'une d'entre elle, le FOAF, permet de relier des personnes entre elles et de fournir un véritable réseau social organisé, comme Twitter ou Facebook par exemple.

### SCORM/LOM

Le SCORM (*Scharable Content Object Reference Model*) et le LOM (*Learning Object Metadata*) sont deux schémas de métadonnées disponibles dans la mise en ligne d'informations dans le cadre de d'outils d'apprentissage.

Le SCORM est une norme qui permet la création d'objets pédagogiques organisés et structurés. Ainsi, les standards pédagogiques peuvent se partager, communiquer... Une amélioration apportée au SCORM, l'API Tin Can, permet de faire de n'importe quel support une plateforme d'apprentissage.

Le LOM, permet l'explication de ressources éducatives numériques. Ces métadonnées permettent de classer, partager, retrouver et modifier facilement les informations mises en ligne. Cela apporte une efficacité plus importe dans le traitement et le partage des données.

## **ID3**

Les métadonnées ID3 sont des métadonnées pouvant être ajoutées à des fichiers audio MP3. Elles permettent d'avoir toutes les informations désirées sur une piste audio (titre, durée, date, auteur...). Cependant, les informations étaient limitées, de même que l'espace disponible.

Elles peuvent être modifiées et rajoutées à partir de logiciels en libre accès. Le plus connu est Audacity, qui permet la création, la modification et le travail sur des pistes audio. Ce logiciel, assez simple d'utilisation et facilement téléchargeable, permet l'édition de métadonnées sur les pistes créées/analysées. Le logiciel gratuit VLC permet également de modifier et compléter les métadonnées d'un fichier audio.

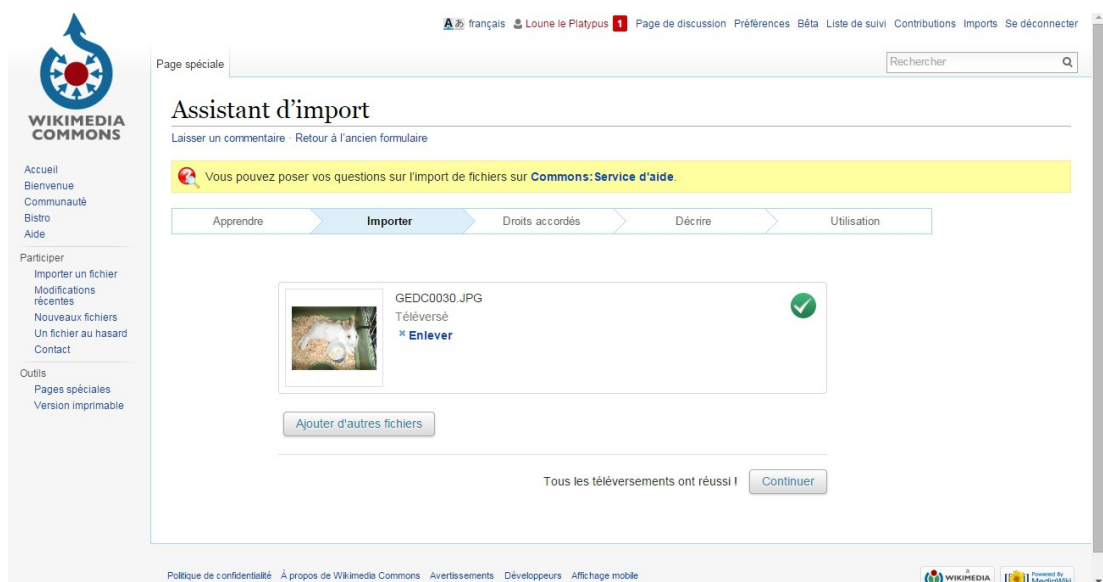
ID3 n'en est pas à son coup d'essai, puisque deux versions ont été rajoutées. La dernière, ID3v2, permet d'enregistrer plus de données, dans un espace plus important (les paroles et les images par exemple peuvent désormais être éditées, ce qui n'était pas le cas dans les versions précédentes). La toute dernière version est la ID3v2.4.

# Wikimédia Commons et Commonist

## Wikimédia Commons

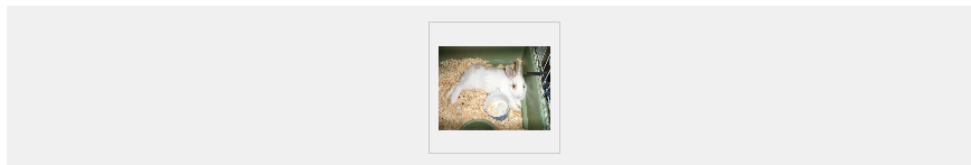
Wikimédia Commons est un site Internet permettant le partage d'images. Afin d'accéder au contenu et de pouvoir poster des documents, il est nécessaire de s'y inscrire (ce qui est gratuit). Dès la première connexion, le site propose un petit didacticiel afin de comprendre le fonctionnement du site. Ce petit didacticiel informe également l'internaute sur le côté libre accès et libre service du site, et il indique enfin les points interdits, comme poster une œuvre inspirée de quelqu'un d'autre sans l'accord de la personne en question.

Dans un premier temps, l'utilisateur importe son/ses documents :



Il est possible d'exporter plusieurs photographies en même temps, de les retirer et d'en ajouter de nouvelles autant de fois que désiré avant de passer à l'étape suivante.

Ensuite, l'utilisateur doit donner la source de l'œuvre postée. Soit il s'agit de son travail, comme ci-dessous, et en ce cas il donne son consentement pour son partage et sa diffusion :



Ce site vous demande de fournir les informations de droit d'auteur pour cette œuvre, pour s'assurer que tout le monde puisse la réutiliser.

**Ce fichier est de mon propre travail.**

Je, , détenteur du droit d'auteur sur cette œuvre, accorde de façon irrévocable à quiconque le droit d'utiliser cette œuvre sous la licence Creative Commons Attribution ShareAlike 4.0 (contenu légal).

(N'importe qui peut utiliser, partager ou reprendre cette œuvre, du moment qu'il m'en accorde le crédit et partage toute œuvre qui en est dérivée sous cette licence.)

[Utiliser une licence différente](#)

Ce fichier n'est pas de mon propre travail.

Suivant

Soit ce n'est pas son travail, et dans ce cas il est obligé de donner la source de son document avec le plus de précision possible, afin de conserver le droit de le poster. Dans le cas contraire, il s'agirait de plagiat.

Ensuite, des informations, des métadonnées, sont entrées dans le but d'identifier le document :

\* = Ce champ est obligatoire.

	<p>*Titre ? <input type="text" value="Shura - Bordeaux 2008"/></p> <p>Description ? <input type="text" value="français"/> <input type="text" value="Photographie de mon lapin Shura à son arrivée dans son nouveau foyer. Bordeaux 2008."/> <a href="#">ajouter une description dans une autre langue</a></p> <p>*Date de création ? <input type="text" value="2008-06-10 00:06:52"/></p> <p>Catégories ? <input type="text"/> <input type="text"/> <a href="#">Ajouter</a></p> <p><a href="#">Ajoutez emplacement et plus d'informations...</a></p>
--	--

Suivant

Parmi ces métadonnées, l'utilisateur peut entrer le titre du document (information obligatoire), une description, la date de création (elle est déjà



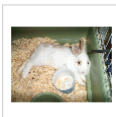
entrée de base, mais elle peut être modifiée, elle fait parti e des métadonnées enregistrées dans la photographie) ainsi qu'une catégorie dans laquelle classer le document sur le site, afin qu'il soit plus rapidement trouvable.

Enfin, l'utilisateur a accès à deux liens, afin de pouvoir partager son document sur des wikis ou, plus largement, sur le net :

Apprendre Importer Droits accordés Décrire **Utilisation**

**Merci d'avoir téléversé vos œuvres !**

Vous pouvez maintenant utiliser ces fichiers sur des wikis, ou les lier sur le web.

 Afin d'utiliser le fichier sur un wiki, vous pouvez copier ce texte dans une page :  
[[File:Shura - Bordeaux 2008.JPG|thumb|Photographie de mon lapin Shura à son arrivée dans son nouveau foyer]]

Pour faire un lien vers lui en HTML, copiez cette URL :  
[https://commons.wikimedia.org/wiki/File:Shura\\_-\\_Bordeaux\\_2008.JPG](https://commons.wikimedia.org/wiki/File:Shura_-_Bordeaux_2008.JPG)

Shura - Bordeaux 2008.JPG

[Aller à la page d'accueil du wiki](#) [Téléverser plus de fichiers](#)

## Commonist

Commonist est un logiciel libre en Java à installer sur un ordinateur, et qui permet de charger un grand nombre de documents dans n'importe quel médiawiki. Une fois les documents envoyés, il apparaîtront accompagnés de métadonnées préalablement complétées avant l'envoi.

Ces métadonnées renseignent les internautes des wikis sur l'identité de l'auteur du document. Mais ce n'est pas tout. Il est possible d'entrer l'adresse du wiki dans lequel les documents seront postés ainsi que tout un tas d'autres informations. Une petite description pourra accompagner chaque document afin d'apporter des précisions aux internautes qui pourraient le souhaiter. Il est également possible d'indiquer si le posteur n'est pas le créateur du document, il doit donc remplir des champs afin de donner le plus de renseignement possible sur la véritable source (c'est une question d'éthique et de bonne foi). Enfin, entre autre, il est possible d'informer les internautes sur la licence sous laquelle le document est.

Contrairement aux métadonnées qui sont entrées sur Wikimedia Commons en même temps que l'enregistrement des images, dans Commonist il est possible de les remplir avant. De plus, les champs sont bien complets, si tout est

rempli il ne manquera aucune information relative au document. C'est un bon logiciel, libre d'accès, qui permet de nouveau d'accroître son importance sur le net en postant des documents bien renseignés, propres et fiables.

## Big Data

*Big Data*, un terme que nous entendons souvent sans trop savoir à quoi il se raccroche ni même ce qu'il représente. Pourtant, lorsque l'on se penche sur le sujet des métadonnées, il est impossible de passer à côté de ces mots et il faut en comprendre le sens et l'importance.

*Big Data*, littéralement, signifie « grosse donnée ». Cela a été traduit au fil du temps par « données de masse » ou « données massives ». Il s'agit d'un groupe réunissant une masse importante de données. Cependant, il est compliqué de travailler avec autant de volume en passant par des logiciels classiques de gestion de base de données ou autre. Il a donc fallu trouver d'autres moyens de les exploiter. L'analyse de ces *Big Data* passe par une règle nommée la règles des 3V<sup>1</sup> :

- un important **V**olume de données à traiter,
- une grande **V**ariété d'informations,
- un niveau de **V**élocité à atteindre.

Dans l'article « *Big Data* : l'analyse des données en masse », en ligne sur le *Journal du Net*, il est indiqué qu'il y a plusieurs technologies possibles dans l'analyse des *Big Data*. Tout d'abord, les *Big Data* dites NoSQL. Les données SQL sont organisées et gérées via des bases de données. Pour les NoSQL, à cause du volume important de données à travailler, des technologies autres que les bases de données sont utilisées. Il s'agit d'espace de stockage plus importants, afin de faciliter l'hébergement et la gestion.

Ensuite, nous avons ce qui est appelé « le traitement massivement parallèle », c'est-à-dire des serveurs, des structures spécialement conçus pour le traitement et l'envoi de données en masse.

Enfin, les données peuvent être stockées dans des espaces mémoires, ce qui permet un traitement plus rapide.

Les *Big Data* sont utilisées par toutes les grandes entreprises de notre

---

<sup>1</sup> Les informations sur les Big Data viennent de l'article « *Big Data* : l'analyse des données en masse », en ligne sur le *Journal du Net* (<<http://www.journaldunet.com/solutions/analytics/big-data/>>).

société afin de recevoir de nouvelles informations, d'établir un profil de client toujours plus perfectionné ou encore dans le but d'accroître leur notoriété et leur importance. Les données proviennent de tous nos appareils connectés (téléphones, ordinateurs, tablettes...), et servent à nous cerner, nous comprendre, afin de créer des produits et des services toujours plus adaptés et qui nous toucheront. Cependant, cela ne concerne pas uniquement les domaines commerciaux. Ces données sont utiles également dans le monde de la politique, afin d'analyser les tendances des votants, en médecine, dans les sciences, dans les domaines sociaux...

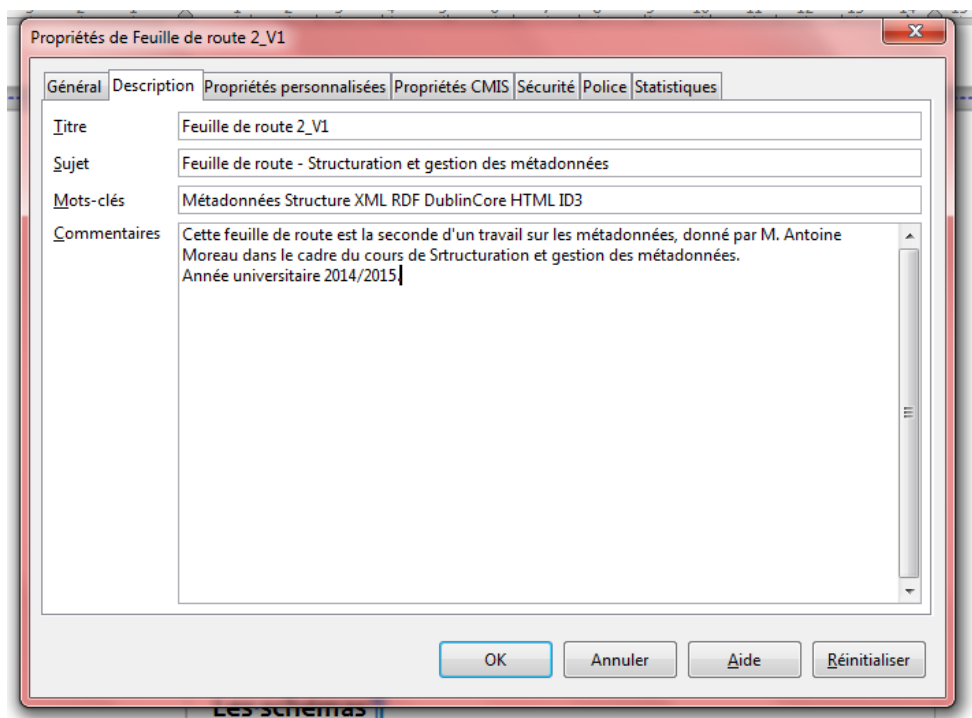
En somme, les *Big Data* sont les yeux et les oreilles des grands de ce monde. Dans une société basée énormément sur le principe de veilles documentaires et commerciales, ces données sont d'une importance capitale pour toute entreprise, association et même tout gouvernement. Il s'agit de transformer le public, les consommateurs, en données exploitables.

## Le site Internet

Dans le cadre de ce travail sur les métadonnées, il nous a été demandé de mettre en place une page Internet qui puisse accueillir une image ainsi qu'un dossier .pdf. Nous allons donc voir quelques métadonnées concernant les dossiers .pdf et les pages web.

### Les métadonnées des dossiers .pdf

Pour les dossiers en .pdf, avant de les transformer dans cette extension, il est possible de leur ajouter un nombre important de métadonnées. Cela se passe via le traitement de texte, qui permet l'édition de ce genre d'informations :



## Les schémas

Les schémas de métadonnées dans le cadre de la création d'un site Internet permettent l'ajout d'informations complémentaires qui se trouveront dans la partie non visible du site. Elles ne s'afficheront donc pas sur la page, mais elles seront bien présentes. Si les balises sont soigneusement remplies, cela va augmenter le référencement du site. Ces schémas permettent la description précise du site ainsi qu'une standardisation de ces informations.

Nous allons voir deux schémas de métadonnées applicables au code HTML :

- DC : Dublin Core<sup>2</sup>,
- OG : Open Graph<sup>3</sup>.

## Les balises <meta>

Dans la création d'un site Internet, les métadonnées ne seront pas visibles de l'utilisateur. Il ne va pas les voir s'afficher sur son écran. Dans l'absolu, ces informations n'intéressent pas l'internaute lambda. Elles vont servir aux moteurs de recherche et aux navigateurs. Elles vont être utiles dans le référencement ainsi que dans l'identification du site, de son auteur, et son organisation interne. Un site dit « propre » a des balises clairement classées, bien remplies et ordonnées.

Les métadonnées vont être entrées dans la balise <head>. Il s'agit de toutes les informations invisibles. Les balises de métadonnées les plus utilisées, celles qu'il ne faut en aucun cas oublier, sont la balise <title>, pour le titre du site, et la balise <description> dans une moindre mesure. C'est elle qui va apporter par exemple le petit résumé affichés par les moteurs de recherche.

Dans la conception du site, il faut en premier lieu lier les schémas de métadonnées dans la balise <head>, comme ci-dessous. Les deux schémas abordés fournissent les même informations, elles ne sont juste pas notées de la même manière et toutes accessibles par tous les navigateurs ou logiciels de traitement des métadonnées.

---

2 Toutes les informations relatives aux métadonnées DC sont disponibles sur le site ci-joint: <[http://openweb.eu.org/articles/dublin\\_core/](http://openweb.eu.org/articles/dublin_core/)>.

3 Toutes les informations relatives aux métadonnées OG sont disponibles sur le site ci-joint: <<http://ogp.me/>>.

```

<!-- LES SCHEMAS. Chargement des schémas qui vont être utilisés
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/">
<link rel="schema.og" href="http://ogp.me/ns#">

```

Ensuite, il suffit d'entrer les balises <meta> avec les bonnes informations. Ces informations vont être la carte d'identité du site, elles vont être utiles pour son référencement ainsi que pour l'image du créateur qui va avoir son nom accessible plus facilement via les moteurs de recherche.

```

<!-- Le schéma Open Graph -->
<meta property="og:description" content="Ce site a été
mis en place dans le cadre du cours de Structuration et gestion
des métadonnées, de M. Antoine Moreau. Il présente une
introduction à ce que sont les métadonnées, un exemple de métadonnée
via le site XnView et deux dossiers .pdf plus complets sur le sujet.">
<meta property="og:type" content="website">
<meta property="og:locale" content="fr_FR">
<meta property="og:site_name" content="Apprends-moi les métadonnées">

```

*Illustration 1: Le schéma Open Graph*

```

<!-- Le Schéma Dublin Core -->
<meta name="dc.title" content="Apprends-moi les métadonnées">
<meta name="dc.description" content="Ce site a été mis en place
dans le cadre du cours de Structuration et gestion
des métadonnées, de M. Antoine Moreau. Il présente une
introduction à ce que sont les métadonnées, un exemple de métadonnée
via le site XnView et deux dossiers .pdf plus complets sur le sujet.">
<meta name="dc.creator" content="Ballangé Lou">
<meta name="dc.publisher" content="Ballangé Lou">
<meta name="dc.subject" content="Métadonnées">
<meta name="dc.language" content="Français">
<meta name="dc.date" content="2015-11-26">
<meta name="dc.rights" content="La mise en page, les impressions
d'écran et les textes sont la propriété de Lou Ballangé dès lors
qu'aucune source extérieure n'est précisée. Ces données peuvent
être utilisées à n'importe quelle fin à condition de citer l'auteur.">

```

*Illustration 2: Le schéma Dublin Core*

## Les limites

Malgré leur utilité clairement établie, les métadonnées ont quelques limites qu'il faut connaître et appréhender. Elles sont certes efficaces en ce qui concerne le référencement et l'e-réputation, cependant elles ne sont pas toujours facilement accessibles. Le site Zotero par exemple, présenté dans la première feuille de route, ne prend pas toujours en compte toutes les métadonnées disponibles sur un site Internet. Il est formaté selon certains schémas, ce qui bloque son efficacité.

De plus, certaines métadonnées sont supprimées lors de la mise en ligne. Prenons l'exemple de Facebook, le site supprime toutes les métadonnées IPTC, hormis l'auteur et le *copyright*. Il s'agit d'une nouvelle limite des métadonnées, qui sont cette fois inutilisées et voire même supprimées volontairement<sup>4</sup>.

Comme dernier exemple des limites des métadonnées, je souhaiterais vous parler de nouveau du site Wikimedia Commons. Nous avons vu plus haut comment enregistrer des images et entrer des informations sur le site. Cependant, malgré la petite bande-dessinée qui informe l'internaute sur les droits et les interdictions en vigueur, le principe se base énormément sur la bonne foi des inscrits. En effet, il est long et difficile de vérifier si une photographie est bien l'œuvre de la personne qui la revendique. De plus, il est possible d'entrer n'importe quelle information en tant que métadonnées, et donc volontairement entrer une description complètement à l'opposé de la photographie qu'elle accompagne, dans le seul but d'avoir des vues.

Malgré leur importance et leurs qualités, les métadonnées sont donc à soigneusement vérifier. De plus, il ne faut pas se fier à tous les sites Internet, car certains traitent en partie ces informations ou les occultent complètement.

---

<sup>4</sup> Informations trouvées dans l'article « Le traitement des métadonnées par Facebook », de Denis Szalkowski, sur son site personnel : <http://www.dsfc.net/internet/reseaux-sociaux-internet/traitement-metadonnees-facebook/>.



## Conclusion

Les métadonnées en ligne, une sacrée aventure donc. Avec ce complément à la première feuille de route, nous avons maintenant un panel assez large des qualités mais également des limites des métadonnées.

Dans un processus de mise en ligne, le principal est de se faire voir mais également d'avoir une bonne *e*-réputation (une bonne image sur Internet). Grâce au rajout de métadonnées via des logiciels comme XnView, Audacity ou les traitements de texte, il est possible de signer n'importe quelle création et de lui donner une véritable portée.

De même, lors de la création d'un site Internet, les schémas disponibles permettent d'apporter énormément d'informations complémentaires dans le but de se faire connaître et de produire un rendu propre et impeccablement codé. Cela peut être long, mais dans un travail réellement sérieux, c'est quelque chose d'important à effectuer.

Cependant, malgré tout cela, le plus gros point faible des métadonnées est le bonne foi des utilisateurs. Il est facile de remplacer certaines informations, d'en entrer des fausses, afin de s'accorder le crédit d'une œuvre par exemple. Il faut toujours se méfier, en tant qu'humain, de ce que nous voyons et pensons juste ou faux. Car le moteur de recherche ne le fera pas pour nous.

Enfin, le sujet des *Big Data*, de plus en plus important de nos jours, ne doit pas être à négliger. Il est capital de se maintenir au courant de ce qu'il se passe, des moyens qu'ont les commerciaux et les politiques afin de mieux nous faire acheter ou nous faire adhérer à une idée ou une autre. La veille doit être faite du côté des consommateurs également, afin que nous ne devenions pas que des enveloppes remplies de métadonnées.