

Leadership par l'exemple. Comment activer les métadonnées et les annotations web sémantique pour favoriser la découvrabilité?

(L'auteur remercie Diane Mercier pour son travail de révision)

La **découvrabilité** des propositions culturelles en ligne est actuellement un sujet très chaud. Plusieurs acteurs culturels et politiques affirment que la découvrabilité sera garante, ou non, dans le futur, de la diversité des expressions culturelles. Les métadonnées servant à décrire un contenu, une ressource ou un objet culturel numérisé tel une chanson, une photographie, un spectacle et l'endroit où il aura lieu, sont de genres et de couleurs très variés.¹

Pour bien saisir l'enjeu que représente l'activation de la découvrabilité, un premier élément est très important : pour décrire un contenu numérisé avec l'aide de métadonnées, il faut le faire à la fois de façon à ce que les humains puissent comprendre la description de ce contenu, mais aussi que les ordinateurs puissent en faire de même, afin d'offrir ensuite des recommandations et des propositions de contenus pertinents.

Tous les secteurs de création possèdent leur vocabulaire propre. Le défi est maintenant d'affirmer ces vocabulaires et de les traduire au langage que pourront lire et comprendre les machines.

Si j'affirme que *Leonard Cohen est un auteur-compositeur né à Montréal*, il est pris pour acquis que tout le monde sait ce qu'est un auteur-compositeur (ce qui n'est pas toujours vrai), il est pris pour acquis que tout le monde sait que Montréal est une ville au Québec et au Canada. Or, si on ne le spécifie pas, un ordinateur pourrait bien croire que Cohen est une ville et qu'un auteur-compositeur est quelqu'un qui chante.

Ainsi, il faut que les secteurs créatifs s'entendent sur ce qu'est un auteur-compositeur et se le réaffirment entre eux pour être capables de s'entendre sur la façon de le faire comprendre aux machines.

Nous allons décortiquer l'exemple plus haut.

Leonard est un prénom.

Cohen est un nom de famille.

Un auteur écrit des textes.

Un compositeur compose de la musique.

Les textes et les musiques sont des propositions artistiques ou des ressources.

Écrire ou composer sont des métiers, des contributions artistiques.

On naît un certain jour du calendrier.

Montréal est un lieu.

Un lieu est parfois un lieu de naissance.

Il faut expliquer tout cela à une machine.

Dans l'affirmation *Leonard Cohen est un auteur-compositeur né à Montréal*, nos champs de métadonnées sont donc :

Prénom

Nom

Contribution (métier)

Lieu de naissance

Souvent, pour les services qui permettent d'écouter la musique de Leonard Cohen, ces quatre champs d'information, connectés avec sa liste de titres de chansons, sont considérés comme amplement suffisant pour encaisser votre abonnement. Bien souvent, le rôle spécifique et le lieu de naissance de l'artiste sont complètement superflus. Mais tout cela est appelé à changer avec le temps. Selon la logique visant à assurer la plus grande diversité culturelle possible, c'est ce que nous voulons. C'est aussi ce que veulent de plus en plus les services de contenus en ligne. Nous les croyons, car ils gagneront davantage d'argent avec des métadonnées plus riches, bien que cela soit un autre sujet d'article.

Dans le cadre du Plan d'action pour la musique et de la mesure 102 du Plan culturel numérique du Québec², le collectif SOPROQ travaille actuellement au projet *Collection* afin de consulter le milieu québécois de la musique enregistrée sur l'adoption d'une **base commune de champs de données jugée nécessaire** pour permettre d'activer une première phase de découvrabilité. Il est donc nécessaire que : «les secteurs créatifs s'entendent sur ce qu'est un auteur-compositeur et se le réaffirment entre eux», tel qu'énoncé plus haut. Il en va de même avec plusieurs autres concepts et termes industriels.

Les champs décrits dans notre exemple pourraient donc figurer sous peu dans un socle commun de données. Lorsque le milieu sera parvenu à définir ces champs obligatoires, il sera possible de les traduire en langage machine.

Il existe plusieurs types de langage machine. De très nombreux modèles de langage machine sont dérivés de l'anglais. Ainsi, il faut songer aux implications multi-linguistiques des choix terminologiques que nous ferons. Actuellement, le langage Schema.org propose les termes tirés de l'anglais qui suivent pour décrire notre exemple.

Prénom deviendra *givenName*

Nom deviendra *familyName*

Contribution deviendra *Role*

Lieu de naissance deviendra *birthPlace*

Il faudra utiliser un standard pour décrire les lieux, par exemple les vocabulaires de Wikidata ou autres organismes « structurant » préférablement au niveau international.

Il importe donc de s'entendre sur les termes et leur sens, sur le dictionnaire que nous emploierons. Nous devons ensuite apprendre à parler MACHINE en choisissant la syntaxe, la structure que nous voulons employer pour faire comprendre à celle-ci que *Leonard Cohen est un auteur-compositeur né à Montréal*.

Si nous sommes sur le point de nous entendre sur une base et un dictionnaire communs pour parler à la chaîne de valeur et aux machines qui l'activent, nous n'avons pas encore abordé les enjeux du où, quand et pourquoi. Lorsqu'un nouvel album musical paraît, où, quand et pourquoi dois-je envoyer des métadonnées? Se pourrait-il aussi que les machines interlocutrices, les destinataires de l'information ne parlent pas tous la même langue? Bien évidemment.

Ainsi, de la même façon qu'il est possible de sauvegarder un texte en .doc, en .docx, en .txt ou encore en .odt, il faudra un jour pouvoir traduire les métadonnées liées à un objet culturel numérisé en code DDEX-XML, ou JSON-LD, selon des langages structurés divers, vers des destinataires nombreux et pour répondre à des besoins variés.

Plusieurs langages non-interopérables, tels DDEX et Schema.org, ou structurés et inter-opérables de type RDF-OWL, tels DOREMUS, CIDOC-CRM, MusicBrainz-NGS, The Music Ontology (mo), Dublin Core, peuvent être utilisés pour décrire ou exprimer une information en langage machine. Il s'agira alors de citer dans le code HTTP, XML, ou JSON-LD servant à communiquer l'information, le modèle utilisé afin de faire connaître aux moteurs automatiques le sens donné aux différents champs transmis.

Dans le langage du Web sémantique associé à un contenu, une balise d'ouverture indique toujours le langage structuré auquel se conformer. Par exemple, «@context»:«<http://schema.org/>» indique que nous avons affaire au langage Schema.org encodé en JSON.

Reprenons notre exemple et éclatons-le en utilisant plusieurs modèles sémantiques. Il est possible de construire un modèle sémantique qui convient à un besoin spécifique en ayant recours à plusieurs langages structurés, que l'on nomme aussi *ontologies*. Cette nouvelle ontologie agrégée visant à résoudre une problématique spécifique est construite sur plusieurs sous-langages ou sous-ontologies pour décrire ce qu'il importe de décrire. Par exemple, des événements (Event Ontology), des œuvres (ontologie FRBR), des personnes (ontologie FOAF), etc. Tant que nous indiquons à la machine, par une balise d'ouverture, à quelle ontologie se référer, il est possible de créer des modèles aussi complexes que la situation le requiert.

Ensuite, il s'agira, pour chaque ligne d'information qu'il est pertinent de transmettre aux robots, aux machines, de construire une instance codée, que l'on nomme un TRIPLET, composée du **champ** (*object*) (ex. prénom) suivi de la **valeur** de ce champ (*properties-subject*) (ex. Leonard) ainsi que de la **relation** qui lie les deux (*relationship-predicate*) (ex. *firstName*, *surname*, *name*, *roleName*, *dateCreated*, *startDate*, etc.).

Pour exprimer les attentes de la machine, voici des sujets et des valeurs qui servent à illustrer notre exemple... Nous vous invitons à cliquer sur les liens pour mieux comprendre les logiques à l'œuvre.

Leonard est un prénom :

http://xmlns.com/foaf/spec/#term_firstName

Cohen est un nom de famille :

http://xmlns.com/foaf/spec/#term_familyName

Leonard Cohen est un auteur-compositeur, le permalien de son identifiant VIAF est : <http://viaf.org/viaf/56703476> ;

et son permalien ISNI est : <http://www.isni.org/isni/0000000110273481>

Sa fiche wikidata est ici :

<https://www.wikidata.org/wiki/Q1276>

Sa présence (URI) dans Wikipédia : https://fr.wikipedia.org/wiki/Leonard_Cohen

2 <http://culturenumerique.mcc.gouv.qc.ca/102-deployer-une-approche-commune-relative-aux-donnees-numeriques/>

3 Pourvu que le jeu de données interrogé offre un port de terminaison SPARQL (SPARQL endpoint) (ex. <http://dbtune.org/musicbrainz/sparql>)