



dominique.cotte@ourouk.fr

**Dominique Cotte** est professeur à Lille (laboratoire Geriico). Il travaille également comme consultant pour les sociétés Aphanía et Ourouk sur l'ingénierie du document numérique et les systèmes d'information. Son dernier ouvrage *Émergences et transformation des formes médiatiques* est paru en 2011 aux éditions Hermès.

[ **positionner** ] Le mouvement général du web sémantique mobilise des compétences parfaitement intégrées par les documentalistes. Mais plus qu'une simple adaptation, cette évolution suppose de remettre en cause ses propres conceptions et comportements en matière de traitement de l'information.

## Les nouvelles perspectives du web sémantique pour les professionnels de l'information

L'appellation de « web sémantique » pourrait faire croire aux professionnels de l'information qu'ils sont moins concernés que les informaticiens ou les linguistes par ce nouveau domaine. Ce serait une erreur, liée notamment à la compréhension de ce qui relève du sémantique dans l'expression *web sémantique*, Tim Berners-Lee lui-même ayant entrepris de critiquer rétroactivement cette notion (voir à ce sujet l'excellent article d'Alexandre Monnin sur la définition du web sémantique<sup>1</sup> ainsi que le papier de James Hendler<sup>2</sup>).

De ce point de vue, parler de « web de données » ou de *linked data* (les deux concepts ne se recouvrant pas) permet de mieux comprendre comment le mouvement général du web sémantique mobilise des compétences parfaitement intégrées par les documentalistes, et ceci de longue date. Nous examinerons dans cet article la façon dont les différentes tâches nécessaires à une pratique de gestion de l'information dans la logique et les standards du web sémantique sont en parfaite continuité avec les savoir-faire documentaires. Néanmoins, il ne faut pas négliger le fait que ce transfert de compétences sur un terrain nouveau est plus qu'un simple déplacement. Il implique également un changement de paradigme et impose d'acquiescer de nouveaux comportements.

Travailler dans et avec le web de données suppose de se placer dans les perspectives suivantes : rompre avec la logique de clôture du « fonds documentaire » ; rompre avec la logique du travail sur « ses » données et s'impliquer dans une perspective d'ouverture ; se

confronter à une technicité plus importante que celle de l'usage de l'informatique documentaire classique ; dépasser le rôle de relais d'une information documentaire déjà élaborée pour réaliser (ou participer à l'élaboration) de supports d'information. Tous ces éléments posent bien sûr des questions d'identité professionnelle et interrogent les frontières avec les métiers existants, eux-mêmes déstabilisés par l'explosion du numérique. Comme dans tout grand changement technologique, les frontières entre les occupations, les tâches et les métiers ont tendance à être redéfinies. Autrement dit, dans le grand mouvement de « documentarisation » du Web qu'incarne le web sémantique, une place pour les tâches documentaires existe, mais il appartient seulement aux professionnels de l'information d'en prendre conscience et de s'approprier cette part de l'activité, sans toutefois chercher à aller trop loin.

À notre sens, on ne peut aujourd'hui éditer de l'information sur le Web sans faire appel à des pratiques documentaires, mais cela ne veut pas dire que les documentalistes doivent devenir des éditeurs et s'emparer de toute la chaîne de production. Au fondement des métiers de la gestion de l'information se retrouvent les besoins suivants, quels que soient les outils et les environnements techniques : structurer, normaliser, qualifier l'information. Le web sémantique, quelle que soit finalement l'interprétation plus ou moins extensive que l'on peut en donner, non seulement n'échappe pas à cette règle, mais il vient au contraire remettre sur le devant de la scène ce type de compétences que les évolutions du Web tout court (et les illusions informatiques en général) avaient un temps semblé reléguer à l'arrière-plan.

En effet, quelles sont les principales tâches de traitement de l'information dans le web sémantique ? Comme en

documentation, il s'agit de sélectionner, structurer, normaliser, qualifier et lier entre elles des ressources.

### Sélectionner des sources et des ressources

Le mouvement du *linking of data* (tout comme par ailleurs l'ouverture des données publiques) implique la forte structuration d'un existant qu'il n'est, par définition, pas nécessaire de recréer en d'autres endroits. De ce point de vue, il existe bien une rupture avec le paradigme documentaire basé sur la localisation ou la détention *in situ* de ressources pour pouvoir les communiquer à un public plus ou moins captif. Dans cette perspective, le documentaliste emprunte au veilleur les techniques de *sourcing*, d'identification de ressources pertinentes et, surtout, de qualification en fonction d'un projet et d'un contexte donnés. Créer, à des fins de publication sur un site web, d'alimentation de flux RSS ou de pages Netvibes, des agrégats de contenu suppose d'avoir défini, à titre de filtre intellectuel et humain, les critères de pertinence qui répondront au mieux aux besoins des utilisateurs.

La même chose vaut pour les ontologies. Comme le souligne Jeffrey Heflin<sup>3</sup>, il convient de limiter la prolifération des ontologies sur un même sujet et d'essayer au maximum de réutiliser l'existant, en prenant évidemment en compte les spécificités locales, qu'elles soient nationales ou censées refléter le point de vue d'une organisation.

Une compétence documentaire sollicitée par le web de données consiste en la qualification des ressources. Les documentalistes n'auront plus (ou plus que partiellement) à constituer des « fonds documentaires » qui leur seraient spécifiques.

### Structurer de l'information sur des objets

Le travail documentaire portait le plus souvent, comme son nom l'indique, sur ces objets achevés que sont les documents. On sait qu'une première brèche a été ouverte dès lors que le document, devenu numérique, évoluait vers une structure ouverte, granulaire et reconfigurable<sup>4</sup>. L'idée du web de données est bien de repérer et de structurer des grains d'information à l'intérieur des structures documentaires pour les exploiter en les reliant entre elles de manière logique et standardisée. Les standards du web sémantique comme RDF créent du sens en typant des relations entre objets, qui permettent ensuite de relier des ressources de façon pérenne.

### Annoter les documents

Dans le vocabulaire du web sémantique, on parle souvent d'annotations, ce qui peut être source de confusions. Classiquement, l'annotation est un ajout de type commentaire, effectué pour soi-même ou pour autrui sur un document déjà existant. Elle vise dans ce cas à éclairer le texte et pas à le décrire. Dans le monde du web sémantique, *annoter* c'est taguer, autrement dit indexer en langage documentaire. L'activité

traditionnelle consistait à indexer des documents « du dehors » alors qu'ici, on indexe des données à l'intérieur d'un document ou d'une micro unité d'information.

### Structurer de l'information pour sa publication

Afin de communiquer correctement avec d'autres ressources, l'information, dans le web sémantique (mais ceci ne fait que poursuivre une des exigences du Web, et plus généralement, de l'informatique), doit être correctement structurée. Traditionnellement, les documentalistes sont habitués à indexer des documents à partir de formulaires de bases de données qui organisent leur description dans des champs. Il s'agit donc en quelque sorte d'une description externe, qui vient préciser du dehors des éléments sur le document, la fiche ainsi produite ne se substituant pas au document, et le document original (dit « primaire ») ne contenant pas tous les éléments de la fiche (par exemple son indexation ou les vedettes matières). La logique du document numérique en général et plus particulièrement sur le Web est différente puisque les données de description (ce que l'on appelle les métadonnées) sont directement intégrées dans le document. Cette logique d'encapsulation est poussée encore plus loin dans le web sémantique puisqu'elle se fait sous un format de données qui permettra d'enrichir le document original avec une information complémentaire. Typiquement, si un document mentionne un nom de personne et qu'il existe quelque part une ressource identifiable qui me permet de connaître des éléments biographiques de cette personne, je peux enrichir ma première ressource par l'ajout d'un lien vers ce deuxième élément.

### Le champ de l'entreprise

Dans l'expression web sémantique, le terme web ne désigne pas seulement la Toile ; il s'agit des technologies du Web, lesquelles ne se déploient pas seulement dans l'espace de l'Internet mais renouvellent également l'informatique d'entreprise. Le développement des intranets « 2.0 », des architectures Soa<sup>5</sup>, la mise en place de services web<sup>6</sup> sont autant de phénomènes qui favorisent et exigent la mise en place des standards du web sémantique. Au lieu de décrire *n* fois des données dans différentes applications, ou d'essayer de les concentrer et de les normaliser dans des ERP dont la mise en place est extrêmement lourde, le web sémantique permet d'organiser une circulation plus fluide entre les différents gisements d'information. La cartographie des données, la réalisation d'ontologies propres au micro domaine de l'entreprise sont autant d'exemples pour lesquelles des compétences documentaires sont requises. Quant aux entreprises publiques ou semi-publiques, aux services de l'État et des collectivités, leur présence sur le web de données et l'ouverture de leurs fonds au public exigeront un investissement croissant des professionnels de l'information. ●

1 <http://cblog.culture.fr/2011/09/07/web-semantique-iri-opendat>

2 <http://blogs.nature.com/jhendler/2009/06/16/what-is-the-semantic-web-really-all-about>

3 <http://www.cse.lehigh.edu/~heflin/pubs/heflin-thesis-orig.pdf>

4 Roger T. Pédaque, Le document à la lumière du numérique, C&F éditions, 2006

5 Service Oriented Architecture : architecture orientée services.

6 Programme informatique permettant la communication et l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués (source : Wikipédia).

7 Enterprise Resource Planning : progiciel de gestion intégré (PGI).