

Web 2.0 contre Web sémantique : un point de vue philosophique

Luciano Floridi^{1, 2}

¹ Research Chair in Philosophy of Information and GPI, University of Hertfordshire ;

² Faculty of Philosophy and IEG, University of Oxford.

Adresse de correspondance : Department of Philosophy, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK ; l.floridi@herts.ac.uk

Traduction française par Patrick Peccatte de

Web 2.0 vs. the Semantic Web: A Philosophical Assessment [format PDF]

<http://www.philosophyofinformation.net/publications/pdf/w2vsw.pdf>

Résumé

Cet article développe certaines des conclusions publiées dans Floridi (2007) concernant les futurs développements des Technologies de l'Information et de la Communication (TIC) et leur impact sur nos vies. Les deux principales thèses soutenues dans ce papier sont les suivantes : alors que la société de l'information se développe, la limite entre connecté et non connecté devient de plus en plus floue, et lorsqu'il n'existera plus de différence significative, nous allons progressivement nous re-conceptualiser nous-mêmes non pas comme des cyborgs, mais plutôt comme des *inforgs*, c'est-à-dire comme des organismes informationnels socialement connectés. Dans ce papier, j'examine le développement de ce qu'on appelle le Web sémantique et le Web 2.0 à partir de cette perspective et je tente de prévoir leur avenir. En ce qui concerne le Web sémantique, je soutiens qu'il s'agit d'un projet clair et bien défini, qui, en dépit de certains points de vue autorisés contraires, ne constitue *pas* une réalité prometteuse, et qu'il échouera probablement de la même manière que le projet de l'Intelligence Artificielle (IA) a échoué dans le passé. Concernant le Web 2.0, je soutiens que, même s'il est assez mal défini et qu'il lui manque une claire explication de sa nature et de sa portée, il a la capacité à devenir un succès (et en effet, c'est déjà un succès dans le cadre du nouveau phénomène du *Cloud Computing*), car il tire parti des seuls moteurs sémantiques disponibles à ce jour dans la nature, nous-mêmes. Je conclus en suggérant quelles autres modifications nous pourrions attendre dans le futur de notre environnement numérique.

Mots-clés

Cloud Computing ; TIC ; Inforgs ; Infosphere ; Web sémantique ; Web 2.0

Introduction

Quelle est la prochaine étape dans le développement du Web ? Depuis au moins l'époque de la bulle Internet, la question a contraint les experts et les techno-fans à rester vigilants. La récente réorganisation de l'industrie, avec la blogosphère qui vient à maturité (The Economist, 2008), a seulement augmenté la pression. Récemment, deux réponses ont progressivement émergées du marché assez bruyant et tapageur des idées : l'une, qui revient indubitablement à Tim Berners-Lee, plaide en faveur du Web sémantique, tandis que l'autre, facilement reconnaissable chez O'Reilly's, soutient ce que l'on appelle le Web 2.0. Comme d'habitude, les philosophes ont été plutôt silencieux sur la question, mais il est temps de briser le silence et de prendre position. C'est ce que j'ai l'intention de faire dans le présent article.

Je vais défendre dans les pages suivantes une thèse assez simple. Les applications du Web sémantique sont soit de la science-fiction excitante (lorsque le terme "sémantique" dans l'expression Web sémantique est pris au sérieux) soit des banalités réalistes (ce que j'appellerais le Web MetaSyntactique) ; par ailleurs, on ne sait pas bien ce que recouvrent réellement les applications du Web 2.0 mais elles représentent une réelle nouveauté dans l'actuel développement des technologies en ligne parce qu'elles prennent pleinement avantage de la sémantique et des capacités de collaboration des utilisateurs afin d'améliorer et d'élargir l'infosphère (<http://fr.wikipedia.org/wiki/Infosphere>). Plus précisément, dans la deuxième section, je soutiendrai que le Web sémantique est un projet clair et bien défini mais qu'il n'est certainement *pas* une réalité prometteuse, en dépit de certains points de vue autorisés qui affirment le contraire (par exemple, Berners-Lee et Fischetti, 1999). Je soulignerai quelques-uns des principaux problèmes que ce projet rencontre dans la section trois et soutiendrai que le Web sémantique, si on le prend au sérieux, échouera de la même façon que l'intelligence artificielle (IA) a échoué dans le passé. Dans la section quatre, je souhaite montrer que le Web 2.0 (O'Reilly 2006) est par contre une idée assez mal définie et qu'il manque une explication claire de sa nature et de sa portée. Cependant, je soutiendrai également que les critiques courantes envers le Web 2.0, telles que celles formulées par Berners-Lee (Anderson, 2006), peuvent être comparées à celles des détracteurs des solutions non-IA concernant des problèmes interprétés comme résolubles en IA, à l'instar de John McCarthy (1997) qui a été déçu par *Deep Blue* et sa capacité à gagner contre Kasparov bien qu'il possédait l'intelligence d'un grille-pain. Dans la cinquième section, je défendrai le point de vue selon lequel parce que le Web 2.0 n'est précisément pas le Web sémantique, c'est une des raisons pour lesquelles il est en train de réussir. Car une fois la nature ontologique du Web 2.0 rendue explicite et précise, on peut démontrer qu'il s'agit d'une réalité très prometteuse, qui reflète au mieux l'évolution future des TIC, car il s'appuie sur les seuls moteurs sémantiques disponibles à ce jour dans l'univers, nous-mêmes, et sur nos capacités sociales à collaborer de façon cumulative. Dans la sixième section, je commenterai la manière dont la philosophie de l'information peut nous aider à comprendre les développements technologiques actuels dans la société de l'information. Je terminerai enfin dans la septième section par un bref commentaire sur le processus de défragmentation de l'infosphère.

2. Qu'est-ce que le Web sémantique ?

L'idée d'un Web sémantique a été introduite par Tim Berners-Lee dans les années quatre-vingt dix. Une dizaine d'années plus tard, il est devenu difficile de démêler, au milieu d'un flot de promotions irréalistes et gonflées ou de publicités éhontées et non fiables, une définition simple et claire du Web sémantique également connu sous le nom de Web 3.0. J'espère ne pas ennuyer le lecteur si je mentionne ici une sélection assez longue de citations. Elles sont nécessaires afin d'illustrer par les mots eux-mêmes combien l'idée d'un Web sémantique est réellement boursouflée. Les passages suivants sont tous tirés de Berners-Lee et al. (2001, tous les passages ont été consultés le 31 octobre 2008; j'ai ajouté les italiques).

La plupart des contenus du Web sont aujourd'hui conçus pour être lus par les hommes, pas pour être manipulés en tenant *pleinement compte du sens* par des programmes d'ordinateur. Les ordinateurs peuvent

analyser de manière experte les pages Web selon leur mise en forme et effectuer des traitements de routine - ici un en-tête, là un lien vers une autre page - mais en général, ils n'ont pas de moyen fiable de traiter la *sémantique*.

Le Web sémantique apportera de la structure au *contenu* des pages Web *du point de vue de leur sens*, en créant un environnement où les agents logiciels voyageant de page en page pourront facilement effectuer des tâches complexes pour les utilisateurs. [...] tout ceci *sans avoir besoin d'intelligence artificielle* du niveau de Hal dans le film 2001 ou de C-3PO dans Star Wars.

[Dans] le Web sémantique [...] les *information possèdent un sens bien défini*, [...] tandis que les machines deviennent de plus en plus capables de traiter et de « *comprendre* » les données qui sont seulement affichées à l'heure actuelle. [...] À ce jour, le Web s'est développé bien plus rapidement comme un médium de documents destinés aux humains plutôt que pour les données et informations qui peuvent être traitées automatiquement. Le Web sémantique vise à réaliser ceci.

Le défi du Web sémantique est donc de fournir un langage qui exprime à la fois des données et des règles de raisonnement sur les données et qui permette d'exporter sur le Web des règles de tout système de représentation des connaissances.

Ajouter de la logique au Web – c'est-à-dire les moyens d'utiliser les règles pour effectuer des inférences, de choisir des actions et de répondre à des questions - est pour le moment la tâche préalable de la communauté du Web sémantique. Un mélange de mathématiques et d'ingénierie des décisions complique cette tâche. La logique doit être suffisamment puissante pour décrire les propriétés complexes des objets, mais elle ne doit pas être si puissante que les agents puissent être leurrés quand on leur demande d'examiner un paradoxe. Heureusement, une grande majorité de l'information que nous souhaitons exprimer se situe dans la lignée d'une formulation telle que "un boulon à six pans est un type de boulon", qui est écrit lisiblement dans les langages existants qui possèdent un peu de vocabulaire.

Le Web sémantique permettra aux machines de *comprendre des documents et des données sémantiques*, et non pas des discours et des écrits humains.

Le sens est exprimé en RDF [Resource Description Framework] et encodé en ensembles de triplets, chaque triplet étant comme le sujet, le verbe et l'objet d'une phrase élémentaire. Ces triplets peuvent être écrits à l'aide de balises XML. En RDF, un document effectue des affirmations selon lesquelles des choses particulières (des personnes, des pages Web ou n'importe quoi d'autre) possèdent des propriétés (comme "est une sœur de", "est l'auteur de") avec certaines valeurs (une autre personne, une page Web). Cette structure se révèle être un moyen naturel de décrire la grande majorité des données traitées par des machines. Un sujet et un objet sont chacun identifiés par un Universal Resource Identifier (URI), de la même manière que l'on utilise un lien sur une page Web.

Le langage humain se développe en utilisant le même terme pour désigner parfois différentes choses, mais l'automatisation ne procède pas ainsi. [...] Utiliser un URI différent pour chaque concept spécifique résout ce problème. Une adresse postale peut être distinguée d'une adresse dans une rue, et les deux peuvent être distinguées d'une adresse discursive.

Toutes ces citations constituent des lectures passionnantes et pleines de promesses. Elles sont représentatives de la littérature sur le Web sémantique. Et pourtant, on est très loin de la perspective bien plus prudente et austère adoptée par le World Wide Web Consortium (W3C) qui décrit ainsi le Web sémantique (nous soulignons) :

Un cadre commun qui permet aux *données* d'être partagées et réutilisées à travers les limites des applications, des entreprises, des communautés. [...] Ce cadre est fondé sur le Resource Description Framework (RDF). "¹

Alors qui a raison ? Et pourquoi cette notable différence ?

1 - *W3C Semantic Web Frequently Asked Questions*, <http://www.w3.org/RDF/FAQ>, consulté le 31 octobre 2008.

3. Pourquoi le Web sémantique ne marchera pas

Les promoteurs du Web sémantique sont, au mieux, hypocritement naïfs quant à sa faisabilité et excessivement enthousiastes en ce qui concerne sa possible réalisation actuelle. Un véritable web sémantique est un problème d'IA complet pour lequel il n'existe pas de solution technologique prévisible². Un Web prétendument sémantique et techniquement réalisable est ainsi peu enthousiasmant, car il doit nécessairement manquer à ses engagements, à savoir

[...] [Un Web dans lequel les ordinateurs] deviennent capables d'analyser toutes les données du Web - les contenus, les liens et les transactions entre les personnes et les ordinateurs. Un "Web sémantique" qui devrait rendre ceci possible doit encore voir le jour, mais quand il émergera, les mécanismes quotidiens du commerce, de la bureaucratie et de nos vies quotidiennes seront traités par des machines communiquant avec d'autres machines. Les "agents intelligents" que l'on a pronostiqué depuis une éternité vont enfin se matérialiser. (Tim Berners-Lee et Fischetti, consulté le 31 octobre 2008).

Les principaux problèmes en ce qui concerne le Web sémantique tel qu'il est envisagé par des auteurs comme Tim Berners-Lee peuvent être résumés par la liste suivante :

1) Trop de paroles et trop peu de détails rendent le projet du Web sémantique confus sur le plan conceptuel. En effet, les concepts clés tels que "sémantique", "sens", "compréhension", "information", "connaissance" et "intelligence", qui sont généreusement saupoudrés dans la littérature concernant le Web sémantique, sont tous utilisés à mauvais escient, employés trop vaguement ou tout simplement métaphoriques. La situation actuelle est que les langages, les protocoles et les ontologies pour les métadonnées et la métasyntaxe peuvent permettre l'intégration, l'agrégation, le partage, la syndication et l'interrogation de données hétérogènes, mais néanmoins bien circonscrites à un sujet et à travers différentes bases de données. Il n'existe pourtant pratiquement pas de "sémantique" là-dedans. En 1997, par exemple, le W3C a défini la première version de la spécification *Resource Description Framework* (RDF), qui est devenue une recommandation W3C en 1999. RDF fournit un langage de représentation sous forme de triplets pour les *Universal Resource Identifiers* (URIs). Le sens et l'intelligence ne jouent aucun rôle dans ce domaine.

2) *Lorsqu'il est ambitieux*, le Web sémantique repose sur une version forte du projet de l'Intelligence Artificielle [IA], et, par conséquent, il est techniquement irréalisable. Mais quand ils essaient d'être plus réalistes, les partisans du Web sémantique confondent *faisabilité technique* (cela peut être réalisé en principe) et *possible succès* (les objectifs pour lesquels la technologie est déployée peuvent être atteints). Il suffit de considérer que l'aviation civile supersonique est effectivement parfaitement possible, mais le Concorde a été retiré en 2003 et il n'existe pas de plans sérieux pour ressusciter les vols supersoniques. L'argent peut ne pas constituer un problème (même si la crise financière actuelle n'est pas de bon augure pour les grands projets), mais nous devrions envisager très attentivement que si nous voulons investir dans une "sémantique à la Concorde", certaines des idées ne voleront pas, quelle que soit la quantité de ressources investies.

3) *Lorsque qu'elle est modeste*, l'idée d'un Web sémantique est bien plus ancienne. Comme Shadbolt et al. (2006) l'ont observé "Tim Berners-Lee l'a formulée à la première conférence du World Wide Web en 1994. Cette idée simple reste cependant largement non réalisée. "Elle est encore et demeurera de fait une *ancienne* idée qui provient directement du rêve de Leibniz de concevoir une *lingua characteristica* (une langue dans laquelle toutes les connaissances peuvent être formulées de manière non équivoque) et un *calculus ratiocinator* (un calcul sur le raisonnement, c'est-à-dire

2 - Pour un point de vue opposé et fort peu convaincant, voir:

http://semanticwebfaq.com/index.php?title=Is_the_Semantic_Web_artificial_intelligence%3F

essentiellement un moteur d'inférence sémantique) tels que la communication devrait être grandement améliorée et les désaccords pourraient être résolus plus facilement. Les choses ne se sont pas améliorées depuis l'époque de Leibniz, et l'ensemble du projet d'une véritable IA reste un rêve (Dreyfus, 1992) ainsi que le démontre le fait de ne pas passer ne serait-ce que des versions simplifiées du test de Turing (Floridi et al., à paraître). Le monde de l'informatique et des TIC s'est certes développé, mais ce que nous avons à la place de ce test, ce sont des ordinateurs qui testent les utilisateurs pour déterminer s'il s'agit d'êtres humains. Le lecteur a probablement subi et réussi le test connu sous l'acronyme de CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart"). Il s'agit de ces suites de lettres légèrement modifiées que l'on doit déchiffrer pour prouver que l'on est un humain et non un agent artificiel, par exemple lors de l'enregistrement d'un nouveau compte sur Wikipedia. Fait intéressant, une bonne stratégie que l'ordinateur A peut déployer pour tromper un autre ordinateur B (par exemple Wikipedia) et lui faire croire que A est un humain consiste à utiliser un grand nombre d'humains comme des moteurs sémantiques qui peuvent résoudre le CAPTCHA. L'ordinateur A se connecte à l'ordinateur B, collecte les informations nécessaires (par exemple une demande de nouveau compte sur Wikipedia), puis relaie le CAPTCHA à un opérateur ou à un groupe d'opérateurs humains leurrés par A afin de le résoudre moyennant une récompense sans savoir qu'ils ont été manipulés (Vaas, 2007). Cela m'amène au point suivant.

4) Dans le Web sémantique, le contenu sémantique est généré par les humains, ontologisé par les humains, et finalement consommé par les humains. Il n'est pas rare en effet que l'on entende certains se plaindre parce qu'il est difficile de trouver et de conserver de bons spécialistes des ontologies. RDF, XML, URI et toutes les autres solutions techniques ne sont qu'une syntaxe à mi-chemin entre un producteur humain en amont et un consommateur humain en aval. Par exemple:

Les éléments de métadonnées du Dublin Core constituent un standard pour la description des ressources d'information entre différents domaines. Le Dublin Core fournit un ensemble simple et standardisé de conventions pour décrire des choses en ligne de manière à ce qu'elles soient plus faciles à retrouver. Le Dublin Core est largement utilisé pour décrire des documents numériques comme la vidéo, le son, l'image, le texte, ainsi que des médias composites comme des pages Web. Les implémentations du Dublin Core font en général appel à XML et s'appuient sur Resource Description Framework (Wikipedia, article "Dublin Core", consulté le 31 octobre 2008).

Cela revient à une taxonomie muette. Aucune automatisation intelligente du traitement sémantique n'est ici envisagée et à juste titre.

5) Le Retour de l'IA Zombie. Il s'agit d'une erreur qu'il semble impossible d'éliminer une fois pour toutes. Elle consiste à confondre l'escalade réussie d'une colline avec le premier pas vers la lune, alors qu'il s'agit de la fin du voyage. Certes, les ontologies et les systèmes experts ont donné lieu à des applications couronnées de succès dans des contextes spécifiques, par exemple, des domaines particuliers de la e-science [les applications scientifiques qui utilisent massivement l'informatique] ou du commerce, mais il est irréaliste d'extrapoler à partir de cela une réussite qui soit applicable à l'ensemble du Web.

6) En relation avec le point précédent, il existe une sous-estimation étonnante des difficultés. Pour citer Shadbolt et al. (2006) à nouveau: « Les défis sont réels. Les ontologies qui fourniront la sémantique pour le Web sémantique doivent être développées, gérées, et approuvées par des communautés de praticiens. "Mais les ontologies ont un faible degré de résilience : le tagging, lorsque l'on se trompe, ne cause pas trop de problèmes, mais une ontologie est fragile. Les ontologies souffrent également d'un faible degré de modularité : chaque tag attribué de bas en haut est immédiatement utile, mais les descriptions d'entités effectuées de manière systématique, de haut

en bas, et qui se veulent exhaustives et fiables sont inutiles sans une grande économie d'échelle. Tagger un restaurant est déjà utile *en soi* ; par contre attribuer un URI à un restaurant ne donne pas grand chose, à moins que beaucoup d'autres restaurants ne soient également "URIsés". Si un lien est erroné (l'erreur 404), on obtient néanmoins quelques informations sur la page Web en question (elle n'est pas disponible). Si les bits d'une ontologie sont erronés (si un URI échoue), l'information manquante est invisible. L'utilisateur ne peut pas voir ce qu'une ontologie ne catalogue pas. Enfin, toute ontologie dépend du niveau d'abstraction (le choix d'une interface particulière, pour faire simple, voir Floridi 2008b) auquel le système est conceptualisé. Mais ces niveaux ne sont ni rigides ni statiques ; ils sont plutôt flexibles et en constante évolution. On peut souhaiter considérer un ensemble de restaurants, non seulement selon le type de nourriture qu'ils offrent, mais aussi pour leur atmosphère romantique, ou selon leur niveau de prix, ou la distance à partir des endroits fréquentés, ou encore selon les langues étrangères parlées ... la liste des desiderata est quasiment infinie, et ainsi en est-il également du nombre de niveaux d'abstraction que l'on peut adopter et aucune ontologie ne peut coder chacun de ces points de vue. Pour finir, les métadonnées sont également une solution fragile et ne fonctionnent que partiellement.

Compte tenu des éclaircissements et des objections précédents, la vérité est qu'une description techniquement précise d'un Web sémantique réellement envisageable possède très peu de ressemblance avec ce que l'on rencontre dans la publicité. Permettez-moi de citer une fois de plus le W3C (je souligne) :

Le Web sémantique est un *Web des données*. [...] Le Web sémantique comprend deux choses. Il s'occupe de *formats communs pour l'intégration et la combinaison de données* tirées de diverses sources, alors que le Web à l'origine se concentrait principalement sur l'échange de documents. Il s'agit également d'un *langage permettant d'enregistrer la manière dont les données portent sur les objets du monde réel*. Cela permet à une personne ou à une machine de commencer dans une base de données, puis d'aller ensuite dans une série illimitée de bases de données reliées non pas par des fils mais par le fait qu'elles portent sur la même chose. (<http://www.w3.org/2001/sw/> consulté le 31 octobre 2008).

Comme le lecteur peut s'en rendre compte, il s'agit tout au long de cette citation de *données* (pas d'*information sémantique* qui nécessite une certaine *compréhension*) et de *syntaxe* (pas de *sens* qui demande de l'*intelligence*). Nous devrions donc en réalité parler de *Web lisible par une machine* ou de *Web des données* comme le fait le W3C. Un tel *Web MetaSyntactique* fonctionne, et fonctionne même de mieux en mieux pour des contextes de plus en plus circonscrits, standardisés et formalisés (par exemple, un catalogue de films DVD pour des clients en ligne). C'est réellement ce que le W3C désigne ici. C'est peu enthousiasmant, et c'est tout simplement invendable en couleurs naturelles ; ce qui est dommage parce que le Web Metasyntactique est un développement véritablement utile.

4. Qu'est-ce que le Web 2.0 ?

Personne à ce jour n'a élaboré une définition du Web 2.0. Fournir une liste indiscutable de conditions nécessaires et suffisantes qui devraient qualifier quelque chose comme le Web 2.0 pourrait bien être une tâche plus que difficile et même se révéler impossible. Mais le fait que le Web 2.0 se réfère à une vague collection de technologies d'une grande variété, de services et de produits qui possèdent tous un air de famille, n'est pas une justification pour un manque de clarté philosophiquement frustrant. Le fait qu'un environnement soit recouvert de brouillard ne constitue pas une bonne raison pour en donner une photo floue. Certes, les tentatives qui visent à affiner ce que nous entendons par "applications Web 2.0" abondent, mais aucune d'entre elles n'a acquis le statut de standard *de facto*. Pour nos fins donc, nous pourrions faire pire que de nous appuyer sur une sorte d'auto-description. L'article correspondant dans *Wikipedia* (qui, normalement, doit elle-même être considérée comme une application Web 2.0) précise ainsi que :

Les concepts du Web 2.0 ont conduit à l'élaboration et à l'évolution de communautés basées sur le Web et ses services d'accueil tels que les sites de réseaux sociaux, les sites de partage de vidéos, les wikis, les blogs et les folksonomies. Le terme s'est imposé après la première conférence O'Reilly Media Web 2.0 en 2004. Bien que le terme suggère une nouvelle version du World Wide Web, il ne renvoie pas à une mise à jour de spécifications techniques quelconques mais à des changements dans la façon dont les développeurs de logiciels et les utilisateurs finaux utilisent le Web. Selon Tim O'Reilly: « Le Web 2.0 est une révolution dans l'industrie informatique provoquée par le passage à l'Internet comme plate-forme, et par une tentative de comprendre les règles de succès sur cette nouvelle plate-forme. » (Article "Web 2.0", consulté le 31 octobre 2008).

Pour être juste, O'Reilly a été un peu plus précis:

Le Web 2.0 est le réseau considéré comme une plate-forme et couvrant tous les périphériques connectés. Les applications Web 2.0 sont celles qui tirent le meilleur parti des avantages intrinsèques de cette plate-forme : la livraison des logiciels en tant que service continuellement mis à jour permettant au plus grand nombre de mieux l'utiliser, de consommer et remixer les données provenant de sources multiples, y compris des utilisateurs individuels, tout en fournissant leurs propres données et services sous une forme qui permette aussi leur remixage par d'autres, créant des effets de réseau par le biais d'une « architecture de participation », et allant ainsi au-delà de la métaphore de la page Web 1.0 pour fournir des expériences utilisateur riches.³

Ainsi, le Web 2.0 est vraiment le *web participatif* et comprend aujourd'hui des "classiques" tels que YouTube, MySpace, eBay, Second Life, Blogger, RapidShare, Facebook, etc. Il suffit juste de vérifier les vingt premiers sites Web mentionnés par Alexa (www.alexa.com). Quelle est alors la différence entre le Web 2.0 et le Web sémantique ? Une bonne façon de répondre à cette question consiste à essayer de comprendre le succès des applications Web 2.0.

5. Pourquoi le Web 2.0 marche-t-il ?

Le Web 2.0 fonctionne pour les raisons suivantes. Les métadonnées sont encore des données, quand bien même il s'agit de données sur des données ; c'est-à-dire que ce sont des différences identifiables qui fournissent et contraignent une interprétation sémantique mais qui en sont encore dépourvues par elles-mêmes (Floridi 2008A). Ces données ne doivent pas être confondues avec des informations sémantiques (qui nécessitent au moins la signification, Floridi 2005), sans parler de la connaissance (qui nécessite certainement la vérité et peut exiger en outre au moins une forme de justification ainsi que la compréhension). Cependant, les agents artificiels – y compris les logiciels quotidiens et, dans la mesure où les connaissances scientifiques actuelles sont en jeu, tout logiciel concevable qui peut être développé à l'avenir - sont des *moteurs syntaxiques* qui ne peuvent pas traiter des données pourvues de signification, c'est-à-dire des contenus informationnels, mais seulement des données à un niveau faible ou supérieur. Ainsi, le Web sémantique est largement un simple battage : nous avons vu, par exemple, que XML est un langage de description de *données* où aucune information n'est ou ne peut être impliquée. Les humains au contraire sont les seuls moteurs sémantiques disponibles, les fantômes dans la machine. Ainsi le Web 2.0 est le Web créé par des moteurs sémantiques pour des moteurs sémantiques, en s'appuyant sur la contribution de légions d'utilisateurs. À titre d'exemple, considérons les folksonomies.

Une folksonomie (mot construit à partir de folk et de taxonomie) est le résultat cumulé de la pratique sociale de production d'informations (de métadonnées, pour être précis) portant sur d'autres informations (par exemple, une photographie) par le biais d'une classification collaborative connue sous le nom d'étiquetage ou de tagging social (par exemple, la photo reçoit les tags "New York", "Hiver", "Statue de la Liberté"). Elle fonctionne de bas en haut, puisque est laissé au seul utilisateur individuel ou au producteur de la cible ainsi "tagguée" le choix de ce qui à classer, comment le

3 - Bien qu'elle semble ne plus être sur le site original, la citation O'Reilly (2006) peut encore être retrouvée sur de nombreux autres sites.

classer et quels mots-clés utiliser dans la classification. Les folksonomies sont devenues populaires depuis 2004 en tant que moyen efficace de personnaliser l'information et de faciliter sa concrétisation grâce à des outils de gestion de l'information. Il est maintenant presque trivial de remarquer que les folksonomies peuvent être extrêmement ambiguës, mais ce n'est pas un problème pour les moteurs sémantiques que nous sommes, capables de processus rapides de résolution des ambiguïtés. Il s'avère que le Web 2.0 constitue un objectif accessible et de plus en plus effectivement réalisé, représenté non plus par la création d'un autre espace extérieur, à l'instar du Web 1.0, mais par un écosystème bienveillant et habité par les humains considérés comme des inforgs (c'est-à-dire des organismes informationnels connectés).

L'explication précédente clarifie le fait que le Web 2.0 soit une partie d'un espace d'information, *l'infosphère*, où la vue platonicienne de la *mémoire conçue comme enregistrement et préservation éternelle* est remplacée par la vue de la *mémoire conçue comme accumulation et perfectionnement*. Il s'agit d'un environnement caractérisé par le fait qu'il est *favorisé par le temps* : le temps ajoute de la valeur et les applications Web 2.0 deviennent meilleures à l'usage, elles s'améliorent avec l'âge, notamment parce que le nombre de personnes impliquées augmente constamment. Ceci à son tour est fonction de la masse critique des "prod-umers" [producers/consumers], c'est-à-dire des producteurs et consommateurs d'information sémantique que j'ai défini plus haut comme des inforgs. Les articles de Wikipedia par exemple deviennent meilleurs plus longtemps ils sont en ligne et plus ils sont utilisés⁴, ceci également parce qu'une nouvelle génération de participants toujours plus nombreux intensifie l'effet de révision. Cela explique pourquoi le Web 2.0 est considéré comme faisant partie d'une évolution plus récente connue sous le nom de *Cloud computing* (http://fr.wikipedia.org/wiki/Cloud_Computing). Il s'agit là d'une autre métaphore et d'un mot à la mode pour désigner l'Internet de façon également assez floue et imprécise. Toutefois, comme dans le cas du Web 2.0, l'expression *Cloud computing* renvoie à un véritable nouveau paradigme quand elle est utilisée pour désigner une version plus évoluée de l'utilisation de ressources de calcul : outils logiciels, espace mémoire, puissance de calcul et autres services ou capacités informatiques sont tous fournis en permanence comme des services basés sur Internet (dans le "nuage") d'une manière entièrement transparente en ce qui concerne l'infrastructure et également transparente pour l'utilisateur. Il s'agit là de l'ultime défi à la localisation spatiale et, de ce fait, à la fragmentation des traitements de l'information. Le *Cloud computing* constitue un espace convivial en ce sens qu'il ne dépend pas du lieu où vous êtes mais seulement de ce dont vous avez besoin comme ressources informatiques.

Le Web 1.0 et le Web sémantique, au contraire, sont défavorisés par le temps et ne comptent pas sur le grand nombre de petites contributions qui peuvent être offertes par des millions d'inforgs. Par exemple, plus un article de la Britannica est disponible depuis longtemps, pire il est, jusqu'à devenir totalement dépassé à longue échéance ; il en est de même pour les vieux sites web qui fonctionnent de façon centralisée, ou de toute ontologie. Ainsi, un test simple pour savoir si quelque chose appartient au Web 2.0 consiste à se demander s'il s'améliore avec le temps, avec son usage, et par conséquent avec le nombre de personnes connectées. Les services qui passent avec succès ce test sont Flickr, YouTube, MySpace, etc.

4 - Un autre point de vue, qui n'est pas en opposition avec celui formulé ici, est exprimé par Duguid (2006). Ce n'est pas le lieu ici d'analyser en détail la critique informée et intelligente de Duguid, je vais donc mettre l'accent uniquement sur deux points. Tout d'abord, son objection selon laquelle les wikipédiens ne devraient pas être si dédaigneux envers la *Britannica* (ou envers toute autre source publiée d'information) est correcte, car après tout, comme il le remarque, des articles anciens et libres de droits d'auteur qui proviennent de la Britannica sont inclus dans Wikipedia. Mais on remarquera que ce fait va dans le sens que Wikipedia est favorisée par le temps, car elle s'améliore précisément parce qu'elle peut facilement cannibaliser toutes les autres ressources libres de droits disponibles. En second lieu, la structure éditoriale de Wikipédia est beaucoup plus complexe, articulée et "hiérarchisée" que certains semblent le penser. Les contenus auto-générés sont véritablement le résultat de processus rigoureux et hautement contrôlés. Le fait que n'importe qui peut contribuer ne signifie pas que toute personne en a la possibilité. Mais cette caractéristique est aussi favorisée par le temps, car elle repose sur des bénévoles et leur volonté de collaborer au sein d'une organisation.

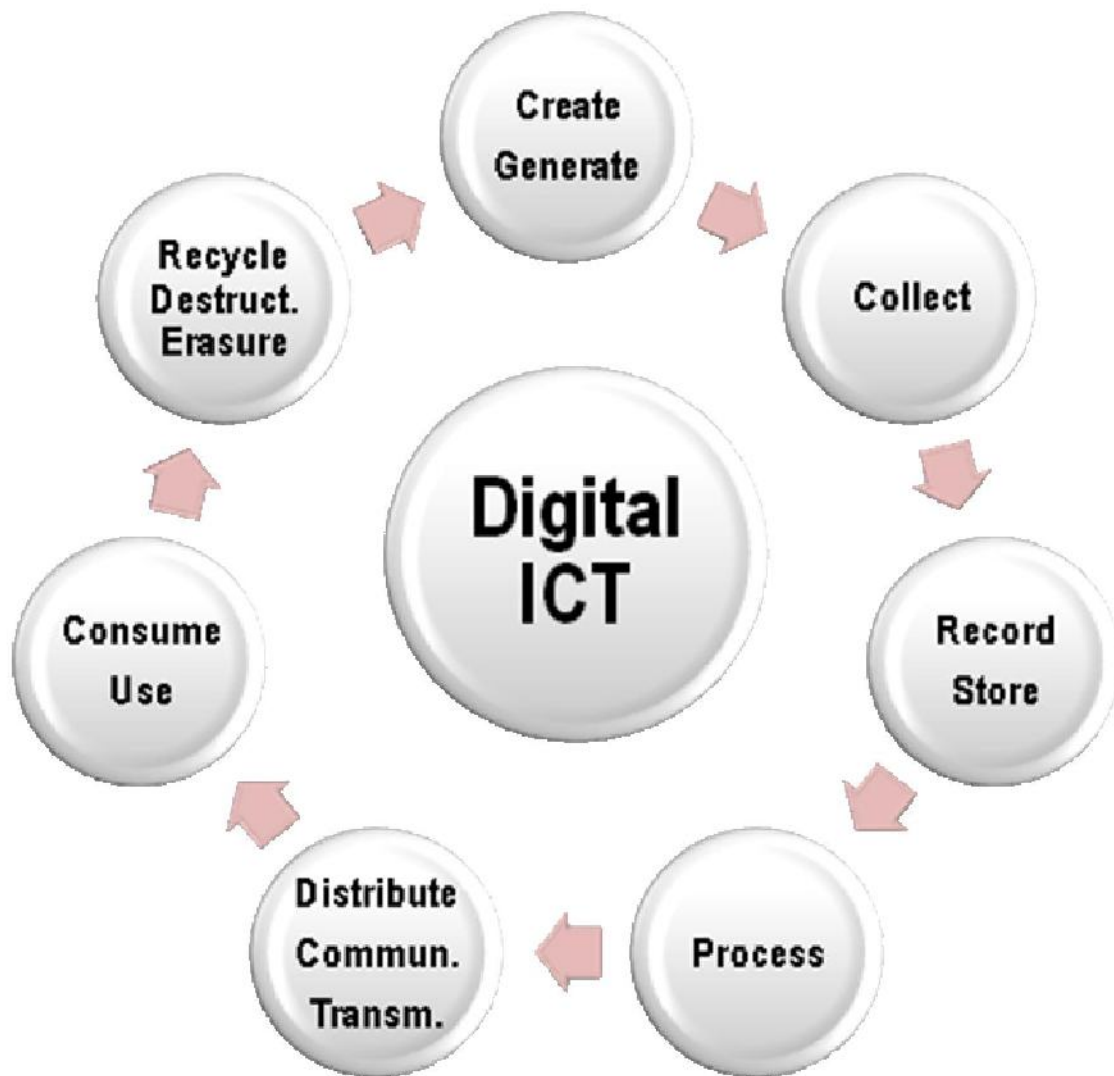


Figure 1. Le cycle de vie de l'information

6. L'évolution du Web : la défragmentation de l'Infosphère

Pour résumer, la totalité du Web sémantique est une erreur bien définie, tandis que le Web 2.0 est une réussite mal définie. Ils constituent tous les deux des cas intéressants d'un phénomène plus vaste qui peut être caractérisé comme la *construction et la défragmentation de l'infosphère*. Le Web 2.0, c'est-à-dire le Web participatif, efface les barrières entre la production et la consommation de l'information en introduisant moins de friction au sein d'une ou de plusieurs phases du cycle de vie de l'information (depuis sa production jusqu'à son utilisation en passant par son traitement et sa gestion, voir la Figure 1), ou efface les barrières entre les producteurs et les consommateurs de l'information. Le Web 3.0, le Web sémantique, compris comme il devrait l'être c'est-à-dire comme un Web MetaSyntactique, efface les barrières entre les bases de données. Nous pourrions alors qualifier le Web qui effectue le pont de Web 4.0, qui efface la fracture numérique entre celui qui est et celui qui n'est pas citoyen de la société de l'information au regard de la disponibilité et de l'accessibilité. Fait intéressant, cela se passe plus en termes de téléphones intelligents et autres appareils portables - par exemple en Chine et en Inde - qu'en termes de commercialisation des ordinateurs personnels. Le Web 5.0 ferait référence au *Cloud computing* et à sa capacité à effacer les

barrières physiques et à globaliser le local. Enfin, le Web 6.0 serait le Web intégré à la vie (Onlife), qui efface le seuil entre ici (le hors-ligne, l'analogique, basé sur le carbone) et là-bas (le en ligne, le numérique, basé sur le silicium). Dans ce dernier cas, d'autres qualificatifs sont envisageables tels que "l'informatique omniprésente", "l'intelligence ambiante", "l'Internet des objets" ou "les objets augmentés par le Web". Tous ces différents Web se développent en parallèle et, de ce fait, ne figurent qu'en partie selon leur ordre chronologique d'apparition. Leur numérotation n'implique pas d'ordre hiérarchique, mais constitue seulement une question pratique d'étiquetage. Ils devraient plutôt être considérés comme des forces convergentes qui poussent l'évolution du web dans le sens d'une meilleure infosphère.

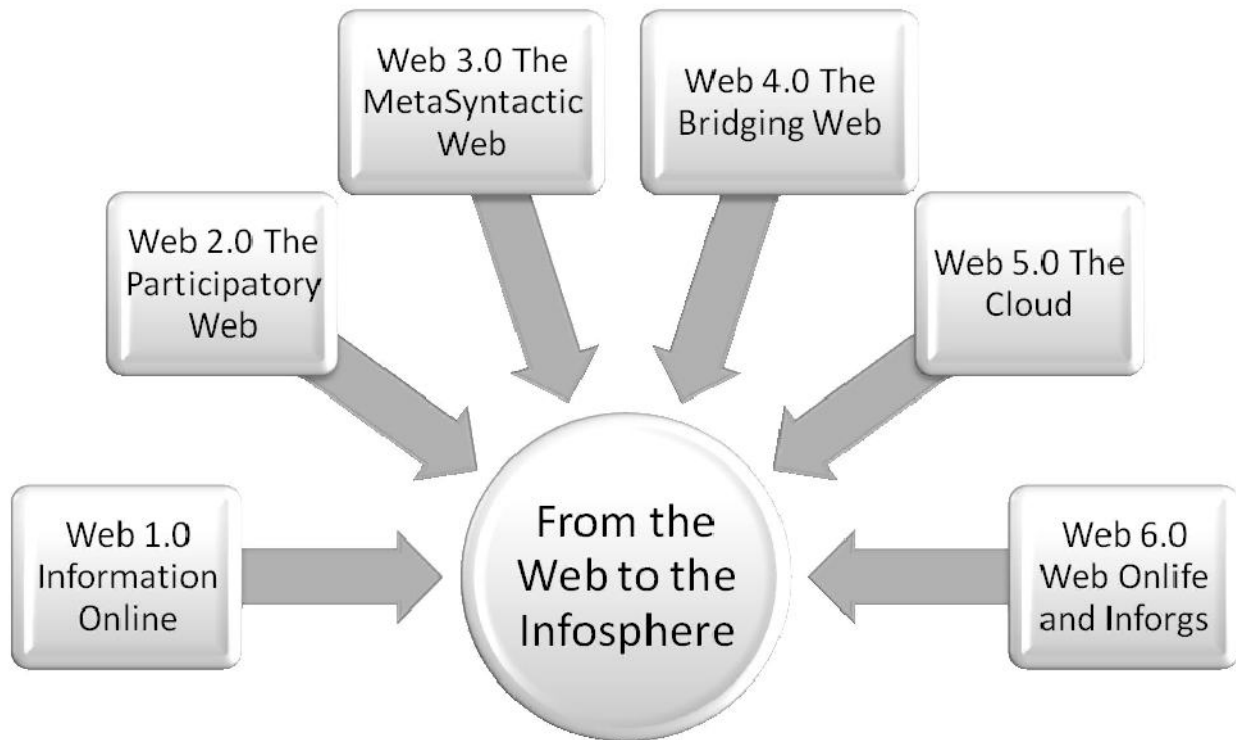


Figure 2 : Cartographie de l'évolution du Web

7. Conclusion

L'interprétation précédente de l'avenir du Web (voir Figure 2) décrit un vaste scénario selon lequel les hommes en tant qu'inforgs sociaux vont habiter une infosphère de plus en plus illimitée, sans jointures, synchronisée (selon le temps), délocalisée (selon l'espace) et corrélée (selon les interactions). C'est un environnement basé sur l'accumulation progressive et la transmission de la sémantique à travers le temps par des générations d'inforgs. Un effort collaboratif pour sauver et améliorer le sens pour de futurs raffinements et réutilisations : cette "politique verte" est le dernier point sur lequel je souhaite faire un commentaire.

Le lecteur doit se souvenir des scènes troublantes du film *Matrix* lorsque l'on découvre finalement des batteries d'élevage d'êtres humains considérés comme de simples sources biologiques d'énergie. Il s'agit d'une histoire irrésistible, mais aussi d'un gaspillage idiot de ressources. Ce qui différencie les hommes ce ne sont pas leurs corps qui ne sont pas beaucoup mieux que ceux de nombreux autres animaux, mais c'est la coalition de capacités que l'on peut appeler intelligence ou esprit. Nous pourrions avoir des queues, des cornes, des ailes ou des plumes, être ovipares ou vivre sous la mer: le meilleur usage que l'on pourrait faire de notre humanité serait encore en termes d'inforgs, des organismes qui sont sémantiquement omnivores, capables de traitement sémantique et d'interactions

intelligentes. Nous générons et utilisons du sens un peu comme les larves du ver du mûrier produisent et utilisent la soie. Il s'agit d'une disposition extraordinaire, peut-être unique dans l'univers, et que nous n'avons exploité que partiellement dans le passé. Les civilisations, les cultures, les sciences, les religions, les traditions sociales, les langues, les récits, l'art, la musique, la poésie ... en bref, toute la vaste sémantique en entrée et en sortie de milliards d'infos a été stratifiée lentement durant des millénaires, comme une mince couche d'humus sur le lit rigide de l'histoire. Elle a été emportée trop souvent par les catastrophes naturelles ou d'origine humaine, ou bien encore rendue stérile par son inaccessibilité ou son indisponibilité. Sans elle, la vie humaine est une vie de brute, celle d'un corps stupide. Sa présence, sa préservation, son accumulation et utilisation optimales ont pourtant été très limitées par rapport à ce que l'humanité a été capable de réaliser dans le domaine de la gestion des ressources matérielles et énergétiques ainsi que dans le modelage de l'environnement physique. La révolution de l'information (Floridi, à paraître) que nous vivons aujourd'hui peut en partie être expliquée comme un rétablissement de ce déséquilibre. Les technologies de l'information et de la communication ont atteint un stade où elles pourraient assurer la présence stable, l'accumulation, la croissance et la disponibilité de plus en plus grande de notre humus sémantique. La bonne nouvelle est que la construction de l'infosphère comme un environnement convivial pour les générations futures est devenu plus facile. La mauvaise nouvelle est que, pour ce qui concerne l'avenir prévisible, la responsabilité de cette gigantesque tâche demeure totalement humaine.⁵

À propos de l'auteur

Luciano Floridi (www.philosophyofinformation.net) est professeur de philosophie à l'Université de Hertfordshire où il est titulaire de la Chaire de recherche en philosophie de l'information ; il est également Fellow of St Cross College, à Oxford University, où il est le fondateur et directeur de l'Information Ethics research Group. Il est actuellement président de l'International Association for Philosophy and Computing (www.ia-cap.org) et Gauss Professor à l'Académie des Sciences de Göttingen. Son prochain livre à paraître est *The Philosophy of Information* (Oxford University Press).

5 - J'ai présenté une première version de ce papier à l'Université Loyola en juillet 2007 au cours de PANAC 2007, la conférence annuelle de philosophie de l'informatique en Amérique du Nord. Je suis reconnaissant à Selmer Bringsjord, Marvin Croy, Anthony Beavers, Matthew Butcher, George K. Thiruvathukal et Thomas Wren pour cette opportunité. Une version révisée a été l'objet d'un exposé que j'ai donné à l'University College de Londres en janvier 2008, et je tiens à remercier Jon E. Agar et Hasok Chang pour leur aimable invitation. Dans les deux cas, je suis très reconnaissant aux participants pour les discussions animées que j'ai apprécié pendant et après les réunions ; elles sont à l'origine de nombreuses améliorations. Un remerciement spécial revient à Joanna Gillies pour la révision du texte, et à Matteo Turilli pour ses commentaires utiles sur une avant-dernière version de ce document. Enfin, je suis redevable à Don Fallis pour son aimable invitation à participer à ce numéro spécial, pour son encouragement et sa patience durant la rédaction, et pour ses nombreuses et précieuses suggestions éditoriales.

Références

- Anderson, Nate.** 2006. "Tim Berners-Lee on Web 2.0 : "Nobody Even Knows What It Means"", *Ars Technica*. Consulté le 2 novembre 2008 sur <http://arstechnica.com/news.ars/post/20060901-7650.html>
- Berners-Lee, Tim and Fischetti, Mark.** 1999. *Weaving the Web: Origins and Future of the World Wide Web*. San Francisco : HarperCollins.
- Berners-Lee, Tim, Hendler, James, and Lassila, Ora .** 2001, "The Semantic Web", *Scientific American* , 284(5), 34-43. Consulté le 2 novembre 2008 sur <http://www.sciam.com/article.cfm?id=the-semantic-web>
- Dreyfus, Hubert L.** 1992. *What Computers Still Can't Do: A Critique of Artificial Reason* 3rd ed. Cambridge, Mass; London: MIT Press
- Duguid, Paul.** 2006. "Limits of Self-Organization: Peer Production and Laws of Quality", *First Monday*. Consulté le 2 novembre 2008 sur http://www.firstmonday.org/issues/issue11_10/duguid/
- Floridi, Luciano.** 2005. "Information, Semantic Conceptions Of." In Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/information-semantic/>
- Floridi, Luciano.** 2007. "A Look into the Future Impact of Ict on Our Lives." *The Information Society* 23(1): 59-64.
- Floridi, Luciano.** 2008a. "Data." In William A. Darity (ed.), *International Encyclopedia of the Social Sciences* . London: Macmillan.
- Floridi, Luciano.** 2008b. "The Method of Levels of Abstraction." *Minds and Machines* 18(3): 303-329.
- Floridi, Luciano.** Forthcoming. "Artificial Intelligence's New Frontier: Artificial Companions and the Fourth Revolution." *Metaphilosophy* .
- Floridi, Luciano, Taddeo, Mariarosaria, and Turilli, Matteo.** Forthcoming. "Turing's Imitation Game: Still a Challenge for Any Machine and Some Judges." *Minds and Machines* .
- McCarthy, John.** 1997. "AI as Sport." *Science* 276(5318): 1518-1519.
- O'Reilly, Tim.** 2006. "Web 2.0 Compact Definition: Trying Again". Consulté le 4 mars 2007 sur http://radar.oreilly.com/archives/2006/12/web_20_compact.html.
- Shadbolt, Nigel, Berners-Lee, Tim, and Hall, Wendy.** 2006. "The Semantic Web Revisited." *IEEE Intelligent Systems* 21(3): 96-101.
- The Economist.** 8 novembre 2008. "Blogging – Oh Grow UP".
- Vaas Lisa.** 2007. "Striptease Used to Recruit Help in Cracking Sites". Consulté le 31 Octobre 2008 sur <http://www.pcmag.com/article2/0,2817,2210674,00.asp>