

Web et théorie du document

Utopie des ingénieurs et appétit des entrepreneurs

Jean-Michel Salaün – Université de Montréal et Collégium de Lyon

3^{ème} Conférence Document numérique et Société – 15-16 Nov 2010 - IEP d'Aix en Provence

Résumé

Cet article illustre la pertinence d'une théorie du document le représentant en trois dimensions complémentaires : forme, texte, médium. Deux exemples sont proposés : l'évolution de la conception du web par son inventeur Tim Berners-Lee qui passe progressivement d'une dimension à l'autre ; le classement des stratégies des principales firmes investissant le web du document, *Amazon*, *Apple*, *Google* et *Facebook* et privilégiant chaque fois une des dimensions.

Théorie du document – web – Berners-Lee – Stratégies d'entreprises

Texte

Qu'est-ce qu'un document ? La question a été souvent posée sans réponse toujours satisfaisante. Niels Lund et Roswita Skare ont fait récemment une excellente recension (2009) des nombreuses approches et théories, jamais en réalité définitives, sur cet objet omniprésent dans nos vies individuelles et l'organisation de nos sociétés. Je m'en tiendrai ici à quelques éléments pour introduire mon propos.

Pour la plupart des textes réglementaires ou des normes, le document est un objet (matériel ou électronique) sur lequel est consignée une information, en anglais on dira un *record*, un enregistrement. L'objet a une forme et l'information est un texte, au sens large c'est-à-dire aussi bien de l'écrit que de l'image ou du son représentés par des symboles. De ces premières définitions, il faut surtout retenir le terme « consigné ». Les sciences de l'information se sont construites autour de cette notion d'enregistrement, sans toujours bien le concevoir. Quand on y parle d'information il s'agit implicitement d'une information consignée.

Mais ces définitions officielles, aussi opérationnelles soient-elles pour régler bien des situations, restent à la surface des choses. Un document ne se réduit pas à sa forme et son contenu, sinon tout écrit en serait un. Dire qu'un écrit est document, c'est lui donner un statut, une fonction particulière. Plus précisément, un document a deux fonctions complémentaires qui se sont affirmées au cours des siècles avec la mise en place de procédures spécifiques : transmettre et prouver. La fonction principale du document est donc mémorielle : on enregistre une information sur un objet pour pouvoir la transmettre ou s'y référer. L'ensemble du système documentaire est en quelque sorte notre mémoire externe pour paraphraser Michel Serres (2007). Comme le rappelle N. Lund, s'intéresser à cette fonction, c'est se demander comme Michel Foucault et bien d'autres, ce qu'un document « est » et « fait ».

Le RTP-DOC¹, au cours d'une réflexion collective (PÉDAUQUE, 2006), a proposé une représentation tri-dimensionnelle pour rendre compte de ces différentes facettes : forme, contenu ou texte et médium. Les deux premières dimensions sont les plus communément indiquées, la troisième, le médium, renvoie à sa fonction sociale. Je voudrais montrer ici combien cette représentation est utile pour décrypter les mouvements actuels sur et autour du web, y compris dans leurs dimensions stratégiques et politiques qui sont l'objet de ce colloque. Je ne prétends pas qu'il s'agisse de la seule grille d'analyse possible, mais que celle-là éclaire la situation sous

¹ Réseau Thématique Prioritaire Document et Contenu du CNRS qui a incité de 2002 à 2005 plus d'une centaine de chercheurs à réfléchir autour du document numérique.

un angle différent des analyses habituelles et donc ouvre d'autres portes, notamment pour les professionnels de l'information et les chercheurs en sciences de l'information.

1. Rappel de Pédauque

Pour suivre le raisonnement, il faut donc avoir en tête quelques-unes des propositions faites dans le premier texte collectif de Roger T. Pédauque (2003). On en trouvera ci-dessous un petit résumé rapide que j'ai un peu adapté.

La première dimension du document, celle de la forme, est anthropologique. Il s'agit du rapport de notre corps et de nos sens à l'objet document, quel que soit sa forme ou son support. Elle se traduit par l'équation $Document = Support + inscription$. L'inscription doit être lisible, c'est-à-dire déchiffrable. L'exemple le plus traditionnel est le livre imprimé. Pour bien des documents aujourd'hui, notamment les documents numériques, cette dimension passe par un appareillage spécial pour permettre leur lecture, et leur forme variera suivant le terminal de lecture. Cette dimension privilégie le repérage, le document doit pouvoir être vu.

La seconde dimension est intellectuelle. Il s'agit du rapport de notre cerveau et de ses capacités de raisonnement au contenu du document, au texte donc, quelle que soit la façon dont il est représenté. Elle se traduit par l'équation $Document = Code + représentation$. La représentation est préjugée être fidèlement rendue par le code, ce qui suppose donc qu'elle soit transposable et qu'elle existe préalablement. Si l'on reprend l'exemple de notre livre imprimé, l'accent cette fois est mis sur le texte, sur son sens sans se préoccuper de son support. La productivité du code informatique autorise une manipulation inédite des documents sous forme numérique, jusqu'à parfois les faire apparaître à la demande. Cette fois il ne suffit plus de repérer par la forme, cette dimension met en avant la signification, le document doit pouvoir être compris ou lu.

La troisième dimension est sociale. Il s'agit du rapport de notre humanité, de notre position dans une société, à la fonction du document, à sa capacité de médiation donc, quel que soit sa forme ou son contenu. Elle se traduit par l'équation $Document = Mémoire + transaction$. Nous retrouvons alors les fonctions de transmission et de preuve. Dans le cas du livre imprimé, cette transmission passe par l'acte de lecture qui fait que l'information présentée est interprétée par le lecteur qui l'assimile. Le lecteur est transformé par l'information qui a été mise en mémoire sur le livre. La communauté des lecteurs, la société en général, utilise le document comme un élément plus ou moins partagé de la richesse mémorielle. Une fois de plus le numérique, cette fois principalement par la capacité des réseaux, modifie considérablement le rapport au temps et à l'espace, à la lecture au sens large et par la même sinon la fonction documentaire, du moins à la place du document dans le social. Cette dernière dimension insiste donc sur la fonction du document, la capacité de son contenu à être assimilé en dépassant le cercle intime et la barrière du temps, autrement à être su.

Chacune de ces dimensions a sa propre logique qui ne réduit pas aux autres et pourtant aucune n'est complètement indépendante des deux autres, et un document doit intégrer et coordonner les trois. Ses modalités anthropologiques (lisibilité-perception, forme-signe), intellectuelles (intelligibilité-assimilation, texte-contenu) et sociales (sociabilité-intégration, médium-relation) doivent non seulement être efficaces prises chacune séparément, mais encore être cohérentes entre elles.

Pour chacune de ces dimensions, les professions traditionnelles du document, archivistes, bibliothécaires et documentalistes, ont développé des outils adaptés : les classifications pour repérer les documents, l'indexation pour rendre compte du contenu et retrouver l'information, ce que l'on a appelé à la suite de Suzanne Briet (1951) « l'information secondaire » indispensable lorsque la masse des documents rend impossible le repérage par la forme, la conservation et les services d'accès pour les partager. Le numérique a fait surgir chaque fois des questions nouvelles qui ont été analysées et traitées par des expertises différentes, aussi bien en informatique que dans les sciences humaines et sociales.

Cette mesure en trois dimensions du document permet de relire bien des situations qui le concernent. On peut, par exemple, pointer et classer les recherches sur le document numérique et en repérer les apports ainsi que les oublis, comme cela a été fait dans les travaux du RTP-

DOC². Elle nous servira ici pour interpréter quelques débats d'aujourd'hui. Je prendrai rapidement deux exemples emblématiques en vous invitant à approfondir la réflexion sur ceux-là ou à l'élargir à d'autres encore : la fondation du web avec l'évolution de la pensée de son inventeur, Tim Berners-Lee ; la fin présumée du web ou son verrouillage, annoncée entre autres par Chris Anderson, rédacteur en chef de la revue *Wired*. La vie et la mort du web en quelque sorte.

2. Le web selon Tim Berners-Lee

Au départ, le web selon la conception de son inventeur reliait des documents entre eux. C'était un vaste système documentaire, conçu comme tel, et l'expression consacrée aujourd'hui par T. Berners-Lee lui-même pour parler des débuts du web est bien « le web des documents » (2009). Les documents étaient gérés comme des objets qu'il fallait retrouver par des processus de recherche ou par des liens, c'est-à-dire des citations. Ce type de raisonnement met l'accent sur la première dimension du document, la forme. Le repérage était, en effet, la question première et la réponse tenait dans une adresse (Url), un protocole d'échange (http), un format normalisé (Html) et un outil de navigation. Le génie de T. Berners-Lee est d'avoir su convaincre la communauté scientifique et les entreprises concernées d'embarquer dans cette normalisation de fait et le succès a été immédiat.

Du web sémantique au web des données

En 2001, le même dans un article célèbre propose de passer à une seconde étape pour mieux utiliser les capacités de calcul des ordinateurs reliés entre eux qu'il appelle alors le « web sémantique ». Le programme sera détaillé, en termes de langages, en couches successives représentées à l'aide d'un schéma sous forme d'un « cake » qui a été régulièrement amendé.

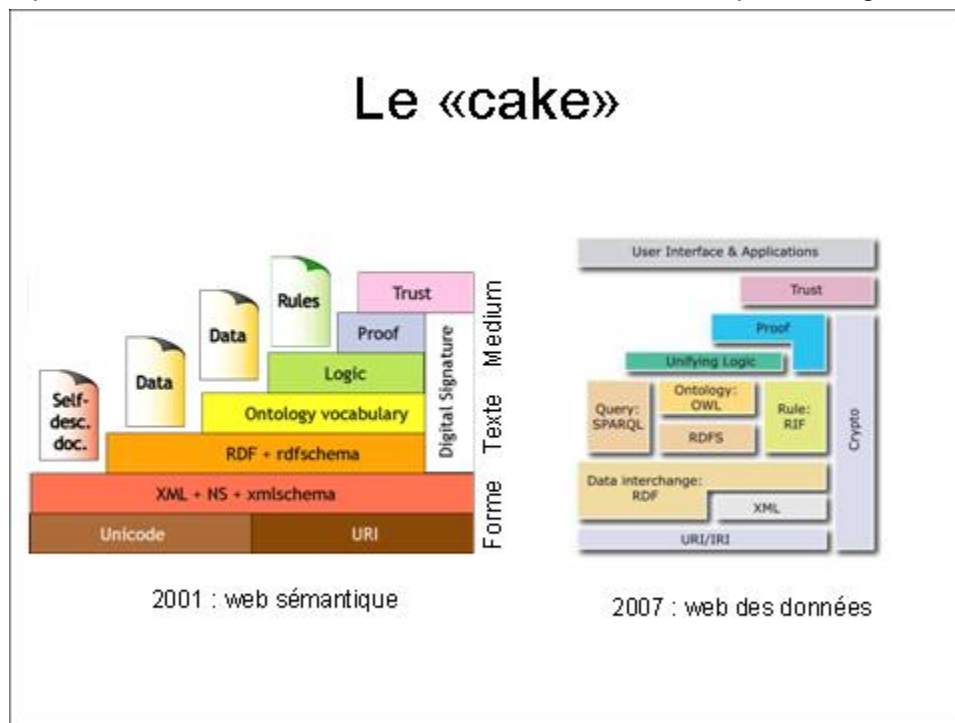


Figure 1 : Représentations du web sémantique ou web des données en couche d'après le W3C

² <http://rtp-doc.enssib.fr/>

Le web sémantique dans son ambition première dépasse largement les applications documentaires, mais il est aisé de retrouver dans son programme les trois dimensions du document. Les couches basses concernent le repérage et le format. Le milieu s'intéresse à la signification par le traitement et l'indexation, tandis que les trois dernières couches renvoient à la fonction sociale du document.

Le programme du web sémantique a évolué en se développant. Il a même changé de nom, on lui préfère maintenant l'appellation « web des données ». Le cake lui-même a été quelque peu modifié, on en trouve la dernière version sur la droite du schéma. Cette évolution mériterait une meilleure analyse d'un point de vue documentaire que je ne le saurais faire ici. Mais disons qu'en dépassant la première dimension celle du document comme objet, les développeurs du web ont rencontré des questions nouvelles dont les réponses n'étaient pas triviales. La notion de texte, par exemple, difficilement réductrice à un calcul algorithmique, ou celle des ontologies ont interrogé les linguistes (RASTIER 2008). Par ailleurs, la vitesse de publication des documents sur le web classique dépassait de loin la possibilité de les indexer et d'autres modes d'organisation sont apparus que l'on a réuni sous le vocable web 2.0 ouvrant la porte à d'autres modalités d'indexation moins formalisées comme les folksonomies.

Les ingénieurs du W3C ont une attitude pragmatique. Ils ont étudié les difficultés, observé les évolutions, tenté de résoudre les problèmes ou de les contourner. Le programme s'est poursuivi et se poursuit encore. T. Berners-Lee le présente donc aujourd'hui comme le passage du « web des documents » au « web des données » et, faisant le parallèle entre les deux, il insiste sur la nécessité de rendre accessible les données de façon normalisée, comme ont été rendus accessibles les documents dans la première version du web (2009). L'intérêt des données, c'est qu'elles sont déjà ordonnées et structurées en bases de données qui peuvent être traitées par les ordinateurs, c'est leur vocation première. Il s'agit d'élargir le web en le transformant en une vaste base de données.

Sans nier l'intérêt et la productivité de cette proposition, ni même qu'elle était déjà présente dans le premier article comme le souligne Gauthier Poupeau sur son blogue (2008), on peut remarquer qu'il s'agit bien d'une évolution qui met de côté la question de la signification et donc celle du texte, ce qui pose des problèmes de tous ordres éthiques, politiques, économiques, parmi d'autres Hubert Guillaud en a fait une recension sur le site *InternetActu* (2009). Il n'est d'ailleurs pas sûr que la notion de document reste pertinente pour rendre compte du web des données, qui s'apparente plus à un vaste système de bureautique, un organisateur ou un traitement de texte intelligent où le document n'est créé qu'en bout de chaîne, comme le résultat des requêtes et calculs. Et inversement, sauf pour les bases de données simples celles justement qui permettent la génération de documents, il n'est pas sûr que le web soit l'outil le plus adapté pour le calcul et l'interopérabilité entre les très grosses bases de données complexes impliquant de nombreux niveaux de calculs comme le suggère John Wilbanks sur son blogue (2010).

La science du web

Récemment, Tim Berners-Lee et ses collègues ont fait une troisième proposition : construire une « science du web » (2006). On peut voir dans cette troisième étape la simple poursuite cohérente de sa feuille de route ou encore une réaction face au développement explosif et inédit de la toile, aussi bien d'un point de vue économique que sociologique. Quoi qu'il en soit du point de vue qui est le nôtre, l'inventeur du web aborde ici directement la troisième dimension du document, celle du médium. Il suffit pour s'en convaincre de citer un des articles prônant le développement de cette science nouvelle : *Le web est un espace construit, créé avec des langages formels et des protocoles particuliers. Cependant, puisque les humains sont les créateurs des pages web et des liens entre elles, leurs interactions engendrent de nouveaux comportements sur le web à l'échelle macroscopique. Ces interactions humaines sont, à leur tour, réglées par des conventions sociales et des lois. Par conséquent, la science du web doit être inévitablement interdisciplinaire ; son objectif est à la fois de comprendre la croissance du web et d'inventer des approches qui ouvrent la porte à de nouveaux modèles puissants et plus profitables* (2006, p.769, notre traduction). Cette troisième approche n'en est qu'à ses débuts et il est trop tôt pour préjuger de ses résultats.

Ainsi théoriquement, la boucle serait en train d'être bouclée et l'ensemble des dimensions du document seraient prises en compte. L'objectif de l'inventeur du web, qui reste un ingénieur, est constructiviste. S'il s'agit de mieux comprendre le web, c'est pour en améliorer le fonctionnement d'abord par des choix techniques dans le but général d'améliorer le bien être. Cette position très pragmatique fait sa force. Le consortium W3C est un lieu de négociations et de consensus non-contraignant mais très incitatif qui n'a plus à prouver son efficacité. Mais la position comprend une part d'utopie qui a aussi ses limites. Les acteurs concernés ne sont pas altruistes, et le cercle vertueux de la croissance du web à l'intérieur des principes édictés au départ peut se briser si les plus puissants considèrent que leurs intérêts seront mieux défendus s'ils s'en exonèrent.

3. Le web selon les entrepreneurs

Chris Anderson, rédacteur en chef de la revue *Wired* et célèbre notamment par la popularisation de la notion de « longue traîne », a publié récemment avec un de ses collègues un article annonçant ou dénonçant la mort du web (2010). La thèse n'est pas vraiment nouvelle et bien des observateurs avaient avant lui souligné la montée des verrouillages sur un réseau dont l'ouverture est pourtant le principe de base. C. Anderson constate que le média internet est arrivé à maturité et que le web (c'est-à-dire ici l'accès par un navigateur) n'en est qu'une des applications dont la rentabilité n'est pas toujours au rendez-vous. Dès lors d'autres joueurs, comme *Apple*, *Amazon*, *Google* ou *Facebook* développent sans se soucier toujours d'interopérabilité d'autres services dédiés et fermés qui rencontrent le succès par leur simplicité d'utilisation, notamment sur les appareils mobiles dont la croissance explosive laisse présager des profits conséquents. Tous ces services sont des services documentaires.

On peut discuter de l'opportunité de la distinction faite entre web et internet dans l'article. Mais le constat général d'une reprise en main est difficilement contestable. La concentration s'est brusquement accélérée ces toutes dernières années. Une étude a récemment montré, par exemple, que la moitié du trafic de l'internet s'était concentrée sur 150 fermes de serveurs contre 30.000 deux ans plus tôt (LABOVITZ 2009).

Une fois le constat fait, il reste à comprendre quelles sont les logiques du média qui s'installe. Ses prédécesseurs ont construit des modèles économiques robustes (édition, presse, radio-télévision). Quel sera celui du petit dernier ? Chacune des firmes est à la recherche du Graal. Rapidement, je voudrais montrer ici que les trois dimensions du document, forme, texte et médium, fournissent là encore une grille commode et instructive pour analyser les stratégies des principales firmes concernées, très marquées par leur métier et leur culture propre. Attention, l'ensemble des dimensions est chaque fois présent dans toutes les stratégies, sinon elles n'auraient aucune chance d'aboutir. Mais, chacune a tendance à privilégier l'une des trois, en fonction de ses compétences particulières, autrement dit de son avantage concurrentiel.

Maîtriser l'objet

Amazon est au départ un libraire en ligne. Il vend des livres imprimés et maintenant aussi toutes sortes d'autres objets. Lorsqu'il s'est lancé dans le livre électronique, c'est-à-dire l'économie du document numérique, il a proposé une tablette, le *Kindle*, qui lui permettait de distribuer et vendre ces documents numériques comme des objets. Poursuivant son métier d'origine en le déclinant sur le web, *Amazon* commercialise donc le document numérique comme un objet et tout naturellement c'est un autre objet, un terminal dédié, qui lui sert de point d'entrée et facilite le verrouillage. Nous sommes bien prioritairement dans une relation anthropologique au document.

Apple qui vient d'un tout autre monde, la bureautique, mais réalise ses profits par la vente de machines, aboutit au même point. Il verrouille le marché du document numérique en les vendant comme des objets au travers de terminaux dédiés : l'*iPod* pour la musique, puis l'*iPhone* et aujourd'hui l'*iPad*. Certes, il faudrait nuancer. La stratégie de chacune des firmes n'est pas exactement comparable, mais il est frappant de voir qu'elles ont construit leur économie dans ce domaine en privilégiant la première dimension du document, sa forme et donc sa stabilité.

Les négociations avec les éditeurs ne sont pas faciles mais elles ne remettent pas en cause le modèle lui-même. Elles portent sur les prix ou sur les risques de monopoles dans la distribution. Cette stratégie, en effet, poursuit l'ordre documentaire antérieur. Elle adapte au numérique le

modèle traditionnel de l'édition qui vend des objets et protège le contenu par la propriété intellectuelle.

Rechercher par le texte

Google a construit sa fortune sur le web des documents en privilégiant la seconde dimension, le texte. Linguistique computationnelle et lexicométrie statistique sont les compétences de bases de la firme qu'elle a appliquées non pas à la manière du web sémantique, mais directement sur le web documentaire en le considérant comme un vaste texte, organisé par les liens entre les documents entre eux et la demande des internautes. Le génie de la firme est d'avoir su utiliser son savoir de base pour bâtir de toutes pièces un marché bi-face (internautes et annonceurs) s'inspirant des médias plus anciens (presse, radiotélévision), mais cohérent avec son service par la vente de mots-clés, associés à la signification de la requête et donc au texte, mais dissociés des documents comme objets contrairement aux médias précédents.

Bien des éléments pourraient être ajoutés pour conforter l'accent mis sur cette dimension : La nécessité pour la firme de disposer de l'ensemble du web en mémoire-cache ; l'utilisation du *fair-use* aux États-Unis pour se dispenser du copyright ; les relations difficiles avec les ayants-droits; la distribution gratuite du système *Android* sur les téléphones portables ; ou, tout récemment, la « recherche instantanée » qui suggère le texte de la requête au fur et à mesure de la frappe et construit un ordre documentaire à partir d'un contenu, etc. La dimension qui prime ici est la dimension intellectuelle, celle du texte et du contenu des documents.

Google a ouvert la possibilité de commercialiser un espace qui jusqu'ici se trouvait hors-marché : celui des bibliothèques, des archives, des institutions documentaires. En se positionnant sur la dimension texte pour capter l'attention par la recherche, il a participé à la modification de l'ordre documentaire ancien grâce à une intense « lecture industrielle » pour reprendre l'expression d'Alain Giffard (2008). La capacité de chercher transversalement dans l'ensemble des textes réduit l'importance de leur classement antérieur et, modifiant notre système de mémoire externe collectif, participe au décadre de doxas. Il est alors naturel que, nonobstant son caractère fortement capitaliste et les ambiguïtés du nouvel ordre documentaire qu'il installe, la firme ait une certaine connivence avec les militants d'un web libertaire dont le ressort est la remise en cause de l'ordre ancien fondé sur la maîtrise des objets. Inversement, les relations avec les médias plus anciens sont particulièrement difficiles.

Relier les humains

Google construit aussi sa stratégie sur la troisième dimension du document, le médium, puisque le classement des documents réalisés par le moteur considère les liens posés par les internautes et leurs requêtes. La firme détient ainsi par son moteur la plus grosse base de données sur les intentions des internautes. Il reste que le savoir-faire de *Google* est d'abord contextuel : il repose sur une expertise de traitement des textes, aussi bien les requêtes des internautes que les documents publiés sur le web, et la dimension comportementale, c'est-à-dire l'analyse des réactions des personnes, n'est pris en compte que vis-à-vis des textes. C'est dans ce contexte que l'on peut relire le slogan de la firme : *don't be evil*, on pourrait chez *Google* manipuler les mots, pas vraiment les personnes. Cette posture a évolué ces dernières années sous la pression des annonceurs et de la concurrence, néanmoins les débats internes à *Google* sur l'utilisation des cookies par exemple ont été très vifs ainsi que le rapporte le *Wall Street Journal* (VASCELLARO, 2010).

D'autres firmes inversent la problématique : ce n'est plus l'ordre documentaire ancien, ni même le contenu qui prime, mais bien les lecteurs. Les capacités documentaires du web sont alors renversées. Les documents ne sont plus que des vecteurs qui relient entre eux des humains. On pourrait dire comme Olivier Ertzscheid « L'homme est un document comme un autre » (2009), les techniques du web documentaire vont servir à indexer les individus pour les décrire et les repérer, non seulement dans leur « forme », c'est-à-dire leur identité et caractéristiques, mais aussi dans leur « texte », leurs comportements et intentions. Cette dernière expertise est évidemment celle de *Facebook* ou d'autres jeunes firmes qui récoltent les traces des internautes pour revendre leurs connaissances des comportements aux entreprises intéressées (STEEL, 2010). *Facebook*, en voulant utiliser sa maîtrise du graphe social comme un avantage concurrentiel décisif pour valoriser la vente d'attention, polarise l'ordre documentaire sur sa

troisième dimension, le medium. Aujourd'hui le débat éthique est très vif autour de ces questions. En effet, jusqu'à présent seules les organisations policières confondaient l'ordre documentaire et l'ordre social pour contrôler ce dernier. Pour autant, les polémiques ne semblent pas avoir grand effet sur la stratégie des entreprises.

Du point de vue de mon analyse, on peut aussi remarquer que la radicalisation de l'homologie individu-document, modifie encore une fois l'ordre des documents qui est alors soumis à celui des individus. La versatilité et la réactivité des clivages sociaux s'introduisent dans les arcanes des logiques documentaires, qui, si elles ne les ignoraient évidemment pas, avaient auparavant plus de recul et d'inertie. On pourrait dire que l'ordre documentaire s'apparente ici à celui de la mémoire vive, avec ses emballements et ses oublis.

Reste que la valeur économique de cette stratégie n'est pas encore vraiment démontrée. On en perçoit les potentialités qui paraissent immenses, font rêver les uns, cauchemarder les autres. Certaines rentabilisations pointent, mais il n'est pas évident que l'homologie individu-document soit à terme éthiquement supportable et économiquement rentable.

Ainsi la théorie du document nous offre une grille de lecture de l'évolution d'un web, pris aujourd'hui entre l'utopie des ingénieurs et l'appétit des entrepreneurs. Ces deux illustrations mériteraient sans doute une analyse plus fouillée et rigoureuse. Mais mon objectif n'était que de montrer la pertinence de la grille de lecture et d'inviter le lecteur à l'appliquer à celles-là ou à d'autres thématiques du document numérique.

La domination de l'ordre ancien du document, celui de la documentarisation du début du XX^e siècle, est contestée. Il s'appuyait prioritairement sur la forme du document considéré comme un objet que l'on pouvait repérer et que l'on s'échangeait. Aujourd'hui la bataille qui se mène sous nos yeux est celle de la constitution d'un nouvel ordre documentaire différent, par une redocumentarisation. Il sera sans doute radicalement différent du précédent, mais il nous appartient de l'analyser et, si possible, de l'améliorer.

Bibliographie

ANDERSON C. et WOLFF M., (2010). The Web Is Dead. Long Live the Internet. *Wired Magazine*, Septembre.

BERNERS-LEE, T., (2009). The next Web of open, linked data. Conférence TED 2009 vidéo, février.

BERNERS-LEE, T et al. (2006). Creating a Science of the Web. *Science* 313, n°5788 (Août), p. 769-771.

BERNERS-LEE, T et al. (2001). The Semantic Web. *Scientific American* (Mai).

BRIET S., (1951). *Qu'est-ce que la documentation*, Paris : EDIT. (accessible en ligne : <<http://martinetl.free.fr/suzannebriet/questcequeladocumentation/>>).

ERTZSCHEID O., (2009). L'homme est un document comme les autres : du World Wide Web au World Life Web. *Hermès*, n° 53, p. 33-40.

GIFFARD A., (2008). Lectures industrielles. Billet *Ars Industrialis*. (Document en ligne : <<http://www.arsindustrialis.org/node/2879>>).

GUILLAUD H. (2009). Critiques du Web² (1/4) : Quelles données libère-t-on ? *Internet Actu*, Septembre 15.

LABOVITZ C. et alii, (2009). Atlas Internet Observatory 2009 Annual Report. Conférence NANOG 47, Octobre 19, Michigan USA.

LUND N. W. et SKARE R., (2010). Document Theory. In *Encyclopedia of Library and Information Sciences, Third Edition*, Taylor & Francis, p. 1632-1639.

PÉDAUQUE R. T., (2003). *Document : forme, signe et médium, les re-formulations du numérique*, (Document en ligne : <http://archivesic.ccsd.cnrs.fr/sic_00000511.html>).

PÉDAUQUE R. T., (2006). *Le Document à la lumière du numérique*, C & F Éditions.

POUPEAU G. (2008). Du Web sémantique au web de données, 2ème partie : retour sur un des articles de Roger T. Pédauque. Billet *Les Petites Cases*, mars 4.

SERRES M., (2007). Les nouvelles technologies : révolution culturelle et cognitive, Conférence *INRIA*, Vidéo, Lille, décembre 20

STEEL E. et ANGIN J., (2010). On the Web's Cutting Edge, Anonymity in Name Only. *Wall Street Journal*, Août 4.

VASCELLARO J. E., (2010). Google Agonizes on Privacy as Ad World Vaults Ahead. *Wall Street Journal*, Août 10.

WILBANKS John, (2010). Documents and Data..., Billet *ScienceBlogs*, Septembre 10..