



Vu, lu, su. Les architectes de l'information face au monopole du web

Jean-Michel Salaün

► To cite this version:

Jean-Michel Salaün. Vu, lu, su. Les architectes de l'information face au monopole du web. La Découverte, pp.152, 2012, 9782707173539. halshs-01979526

HAL Id: halshs-01979526

<https://shs.hal.science/halshs-01979526>

Submitted on 13 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vu, lu, su



Vu, lu, su



Jean-Michel Salaün

Vu, lu, su

Les architectes de l'information face à l'oligopole du Web

2012



Présentation

Ce livre propose d'observer le Web sous un angle inédit, en le plaçant comme un moment d'une histoire longue et méconnue, celle du document. Inventé par Tim Berners-Lee, le Web prend la suite des efforts d'indexation systématique lancés à la fin du XIX^e siècle, modifiant le document lui-même dans ses trois dimensions : la forme, le contenu et la fonction de transmission. Le Web est alors un média comme un autre, s'inspirant de la bibliothèque et de la radiotélévision pour répondre aux aspirations documentaires d'une société qui a changé en profondeur. Les anciens médias eux-mêmes élargissent leur vocation en devenant des « industries de la mémoire », par l'archivage numérique continu et public de leur production. Les nouveaux venus, comme Apple, Google ou Facebook, privilégient chacun une dimension différente du document pour prendre une position dominante dans la construction d'un « néodocument ».

Puisant ses références dans différentes disciplines et s'appuyant sur le travail d'un réseau de chercheurs francophones sur le document numérique, ce livre ouvre plus largement les possibilités d'interprétation du Web et propose à ses acteurs indépendants de devenir des « architectes de l'information » pour contrer l'hégémonie menaçante des géants de la Toile.

L'Auteur

Jean-Michel Salaün est professeur à l'École normale supérieure de Lyon. Il a été directeur de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal (2005-2010) et est l'auteur de nombreuses publications sur l'économie des bibliothèques et le document numérique.

Copyright

© Éditions La Découverte, Paris, 2012.

ISBN numérique : 978-2-7071-7353-9

ISBN papier : 978-2-7071-7135-1

Composition numérique : Facompo (Lisieux), janvier 2012.

Cette œuvre est protégée par le droit d'auteur et strictement réservée à l'usage privé du client. Toute reproduction ou diffusion au profit de tiers, à titre gratuit ou onéreux, de tout ou partie de cette œuvre est strictement interdite et constitue une contrefaçon prévue par les articles L 335-2 et suivants du Code de la propriété intellectuelle. L'éditeur se réserve le droit de poursuivre toute atteinte à ses droits de propriété intellectuelle devant les juridictions civiles ou pénales.

S'informer

Si vous désirez être tenu régulièrement informé de nos parutions, il vous suffit de vous abonner gratuitement à notre lettre d'information bimensuelle par courriel, à partir de notre site www.editionsladecouverte.fr, où vous retrouverez l'ensemble de notre catalogue.



Table

Introduction. Une approche documentaire du Web

1. Au commencement, la bibliothèque

L'affirmation d'un modèle

La valeur créée

Externalisations numériques

2. Les mutations du document

Petite histoire du mot

Petite histoire de la notion

Les trois dimensions

3. Réingénieries documentaires

Renouveau

Le document au ^{xxi}^e siècle

D'une modernité à l'autre

4. L'économie du document

Les trois modèles de valorisation du document

Articulation des modèles

L'économie du document

5. À la recherche du néodocument

Les paradoxes du néodocument

L'élaboration d'un modèle commercial

Conclusion. Architectes et archithèques

A Micha

Introduction

Une approche documentaire du Web

À la fin de sa communication dans un colloque sur les publications scientifiques numériques, un chercheur indiquait il y a quelques années : « Posez vos questions, j'ai déjà les réponses. Toutes les questions m'ont déjà été posées. » Jamais peut-être la naissance d'un média n'aura été suivie et débattue d'aussi près. L'histoire du Web s'écrit et se discute en direct sur le Web à partir de propositions d'essayistes majoritairement nord-américains¹. On y discute de la richesse des réseaux, des *digital natives*, de la neutralité du Net, de la longue traîne, de la fracture numérique, du *peer-to-peer*, des *creatives commons*, du piratage, de la gratuité, des archives ouvertes, du Web 2.0, du Web sémantique, de l'*open source*, l'*open acces*, l'*open data*, de l'influence de Google sur nos cerveaux, de celle de Facebook ou Twitter sur les révolutions, de la fin de la vie privée, de la bulle informationnelle, etc. Chaque innovation est immédiatement relayée, discutée, critiquée ou encensée. Chaque idée, chaque analyse est reprise, contredite ou confortée dans l'instant ou presque.

Ainsi, comme le précisait le chercheur, les réponses ont déjà été données avant même que les questions ne soient posées, ou plutôt l'ampleur et la rapidité du débat permettent de retrouver, toujours ou presque toujours, une réponse à la question que l'on se pose, car elle a déjà fait l'objet d'une discussion enregistrée préalable à notre réflexion. Le forum permanent a la grande vertu de documenter finement toutes les étapes du Web, mais il a aussi le défaut de saturer la réflexion autour de propositions issues d'une même source collective et d'interdire le recul par la vitesse du débat. Cette transparence est en phase avec son objet, comme nous le verrons dans ce livre. Mais par effet de miroir, par une mise en abyme du Web sur lui-même, elle fixe les postulats et fige les oppositions. Même plurielles, même apparemment contradictoires, ces analyses ont un socle commun de postulats implicites qui finit par constituer les fondations d'une réflexion collective sur le Web s'apparentant plus à une idéologie qu'à une pensée analytique.

Partant de prémices différentes, mon propos ouvre des perspectives nouvelles sans nécessairement s'opposer aux précédentes. Il s'agit d'observer le Web sous un angle jusqu'ici non considéré et d'essayer d'en tirer le maximum de leçons. Ce livre propose de placer le Web comme un moment d'une histoire longue et méconnue, celle du document. Le Web est issu de l'émergence progressive de la notion de document, qu'en retour il fait évoluer. La notion est encore mal étudiée aujourd'hui. Il est, en effet, des objets ou des phénomènes qui nous sont tellement familiers, tellement omniprésents, que nous ne les voyons plus, ou plutôt que nous ne nous interrogeons pas sur leur fonction, qui semble aller de soi. Tels sont les documents.

Nous avons des documents dans nos poches, nos sacs, nos portefeuilles, nos cartables, plein nos étagères, nos tiroirs, nos bureaux, nos disques durs. Nous en croisons dans la rue, nous en échangeons dans les magasins, nous en sommes submergés au travail, dans nos boîtes aux lettres, sur nos pages de réseaux, nous en consommons tous les jours dans les médias. Ils nous définissent, règlent nos actions, témoignent de nos transactions, nous engagent, nous renseignent, nous instruisent, nous distraient ou nous ennuiant. Ils accompagnent chaque moment de notre vie. Ils l'accompagnent de façon si étroite que l'on pourrait dire qu'ils en sont une part même. Faisant partie de nous, il nous est superflu de trop y réfléchir. Ils sont là, utiles ou inutiles, et nous faisons avec. Pourtant, les documents ont beaucoup changé au cours de l'histoire. Le mot même de « document », tellement employé aujourd'hui, n'est devenu d'usage courant qu'au cours du XIX^e siècle. Et nous assistons à la transformation radicale des documents, en direct pourrait-on dire, par le numérique et le Web, qui nous sont devenus en quelques années aussi familiers que leur support habituel précédent, le papier. Il est grand temps d'essayer de voir plus clair sur le sujet. Ce livre est donc une tentative pour relire l'histoire récente du Web à la lumière des travaux sur la théorie du document initiés en France, aux États-Unis et en Europe du Nord².

Le premier chapitre rappelle que la bibliothèque qui collecte les documents est, et de beaucoup, le plus ancien des médias. Si le numérique a ébranlé son modèle, il n'est pas le premier à l'avoir fait. L'imprimerie à caractères mobiles l'avait déjà privée de la fonction de reproduction et circulation des livres. Inversement, la bibliothèque a été, consciemment ou non, une des premières sources d'inspiration pour le développement du Web et, ironiquement, certains des acteurs du numérique ont réussi à faire de l'exploitation de son modèle une activité très profitable alors que celui-ci était depuis toujours fondé sur un écosystème non marchand.

Le deuxième chapitre revient sur la notion de document en montrant que si ses deux fonctions, transmettre et prouver, sont très anciennes, le document n'est devenu un objet familier qu'au XIX^e siècle. Son émergence est contemporaine du développement de la science moderne. Son apogée coïncide avec la systématisation du modèle de la bibliothèque. Comme souvent dans les sciences humaines et sociales, c'est au moment où l'objet semble s'échapper que les efforts pour le comprendre sont les plus grands. Analysant l'éclatement du document au travers du numérique, un collectif de chercheurs, signant du pseudonyme Roger T. Pédaque, a proposé d'éclairer la nature du document à partir de trois dimensions : la forme, le texte ou le contenu, et le médium ou la transmission ; le « vu », le « lu » et le « su ». Cette partition servira de canevas pour la réflexion dans la suite du livre.

Le troisième chapitre montre combien l'organisation documentaire, hiérarchique et systématique, issue du XIX^e siècle, a dominé le XX^e en se perfectionnant, miroir d'une société issue de la révolution industrielle et scientifique. La construction du Web sous l'impulsion du consortium W3C³ a conduit à une réingénierie documentaire radicale qui met aux commandes l'internaute et modifie le document dans ses trois dimensions. Poussé au bout, le processus renverse notre régime de vérité en faisant de l'individu lui-même un document. Cette réingénierie documentaire est en phase avec les valeurs émergentes d'une société postmoderne.

Pour les documents publiés, trois marchés, et trois seulement, qui correspondent aux trois dimensions du document, ont été ouverts : l'édition avec la vente de biens (forme), la bibliothèque avec la vente de services d'accès (texte) et enfin le spectacle avec la vente d'attention (médium). Le Web s'insère entre la radio-télévision et la bibliothèque, comme la presse s'était insérée entre l'édition et le spectacle au XIX^e. De l'édition à la radio-télévision en passant par la presse et le Web, on observe une modification progressive de la gestion de l'espace-temps de la production-consommation des documents publiés, régulée par l'organisation tarifaire. De plus, le Web s'impose aux médias et les transforme en industries de la mémoire, par leur accès transversal et leur archivage permanent. Dès lors, il est possible de reconsidérer les catégories traditionnelles de l'économie de la culture pour mettre en avant l'importance de celle du document. Cette relecture des industries culturelles à partir d'une entrée documentaire est l'objet du quatrième chapitre.

L'analyse tridimensionnelle du document permet enfin (cinquième chapitre) d'éclairer les stratégies des principaux acteurs industriels du Web et de les comparer avec les intentions des ingénieurs et les souhaits des internautes militants, soulignant la distance entre les utopies et les réalisations industrielles. Des verrouillages et des affermages ont été mis en place en privilégiant, là encore, l'une ou l'autre dimension du document : la forme, en maîtrisant les objets (Apple), le contenu, en naviguant au travers des textes par des « lectures industrielles » (Google), ou encore la relation (Facebook), de façon à accaparer avec plus ou moins de succès le maximum de profits. Le livre se conclut par le besoin de nouveaux professionnels de l'information, capables de comprendre les mouvements en cours pour construire et gérer les nouvelles infrastructures documentaires. Les *iSchools* en Amérique du Nord forment aujourd'hui des professionnels compétents aussi bien en informatique qu'en sciences de l'information. On appelle parfois ces nouveaux professionnels des « architectes de l'information ». En français, je suggère d'utiliser le terme d'« archithécaires » pour signifier ces nouvelles compétences, puisant leurs racines dans les savoirs bibliothéconomiques, mais les dépassant très largement.

En résumé, à partir de l'histoire du document et de son analyse tridimensionnelle, le livre propose un éclairage différent de ceux qui se discutent sur le Web. On y constate que cette invention du Britannique Tim Berners-Lee prend la suite des efforts d'indexation systématique lancés à la fin du XIX^e, modifiant le document lui-même dans ses trois dimensions : la forme, le contenu et la fonction de transmission. Le Web est alors un média comme un autre, s'inspirant du modèle de la bibliothèque et de celui de la radio-télévision pour répondre aux aspirations documentaires d'une société qui a changé en profondeur. Les anciens médias eux-mêmes élargissent leur vocation en devenant des « industries de la mémoire » par l'archivage numérique continu et public de leur production. Les nouveaux venus, Apple, Google, Facebook, privilégient chacun une dimension différente du document pour tenter de prendre une position dominante dans la construction d'un « néodocument ».

Ce livre plaide pour une prise en compte frontale de la réingénierie documentaire initiée par le Web ; certains parlent de « redocumentarisation » du monde pour souligner à la fois sa parenté avec les tentatives précédentes d'organisation documentaire systématique et son ambition globale. Reste que ce livre est un essai et non un compte rendu de recherches. Il en a les limites, les développements sont plus des propositions que des résultats à la méthodologie sans faille, le domaine aurait aussi pu être couvert plus largement. Au-delà des sciences de l'information, j'emprunte des notions à l'histoire, à la linguistique, à l'économie, à la sociologie et un peu à l'informatique. De tels emprunts sont toujours périlleux et les spécialistes des disciplines pourront sans doute me reprocher des approximations et des interprétations rapides. Mais il me semble que l'ampleur du phénomène décrit ne peut s'aborder lucidement sans l'éclairage croisé des expertises et j'espère que la cohérence générale du propos emportera la conviction, malgré les raccourcis.

Même si le propos et le raisonnement tenus sont originaux, ils s'appuient sur des réflexions et des recherches collectives, pluridisciplinaires et internationales menées depuis plusieurs années. À l'origine, la réflexion se nourrit de recherches sur les industries culturelles menées il y a déjà longtemps à l'Université Stendhal de Grenoble, puis à l'École nationale supérieure des sciences de l'information et des bibliothèques (ENSSIB) de Villeurbanne. Mais le propos aurait été sans doute moins ouvert sans de nombreux apports multidisciplinaires et internationaux. Je suis redevable aux très nombreux chercheurs français qui m'ont accompagné pendant trois ans (2002-2005) au sein du réseau thématique pluridisciplinaire « Document et contenu », et particulièrement à Catherine Garbay qui m'en a confié l'animation. Les discussions, analyses et réflexions, et notamment la dynamique de rédaction collective, sous le pseudonyme Roger T. Pédaque⁴, ont été décisives pour mûrir ma réflexion, comme en témoignent les nombreuses allusions et citations qui émaillent le livre. Il faut ajouter que la montée du Web 2.0 – le développement des outils de partage, blogs, wikis, plateformes communes, fils RSS, réseaux sociaux – a favorisé des échanges, facilité la veille systématique et l'émergence de blogueurs qui par leurs signalements, réflexions et provocations ont largement nourri ma propre analyse. Je voudrais en citer quelques-uns, car ils n'ont pas toujours l'honneur des publications académiques et pourtant, sans leur vigilance infatigable, ce livre serait moins documenté : Hervé Bienvault, Jose Afonso Futardo, Hubert Guillaud, Martin Lessard, Pierre Mounier, Sylvère Mercier, Fabrizio Tinti, et bien d'autres encore.

De même, les rencontres à Berkeley et à Tromsø de la « Document Academy », animée par Niels W. Lund et Michael Buckland, ont permis d'élargir mon point de vue à l'international. Cet élargissement a été consolidé par la direction pendant cinq années (2005-2010) de l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal. J'ai pu ainsi comprendre de l'intérieur le monde des sciences de l'information de l'Amérique du Nord, différent et complémentaire de celui de la France. Mon passage à l'EBSI a été l'occasion d'y construire un cours en ligne et un blog sur l'économie du document, dont bien des propositions sont reprises de façon synthétique et ordonnée dans ce livre. Les internautes qui ont suivi l'un ou l'autre en reconnaîtront des extraits le long de ces pages.

Enfin, ce livre n'aurait pu être rédigé sans l'année sabbatique que m'a octroyée l'Université de Montréal, sans les conditions de travail idéales fournies par le Collegium de Lyon qui m'a accueilli, et sans les encouragements et la relecture attentive de Michèle Guisset et Benoît Habert. Qu'ils soient tous remerciés.

Notes de l'introduction

1. Parmi ceux-là : Chris Anderson, John Battelle, Danah Boyd, Yochaï Benkler, Nicholas Carr, Manuel Castells, Stevan Harnad, Lawrence Lessig, Tim O'Reilly, Howard Rheingold, Clay Shirky, Richard Stallman, Peter Suber, Jimmy Wales, Tim Wu... sans parler du fondateur du Web lui-même : Tim Berners-Lee.

2. Niels Windfeld LUND, « Document theory », *Annual Review of Information Science and Technology (ARIST)*, Volume 43, 2009, p. 399-432.

3. La devise du World Wide Web Consortium (W3C) est « amener le Web à son plein potentiel ». Il s'agit d'un lieu ouvert où se discutent les protocoles et normes informatiques qui régissent le Web et où se construisent les consensus entre scientifiques et industriels. <<http://www.w3.org>>.

4. Pour simplifier j'écirai par la suite « Pédaque » pour signifier les « travaux du réseau de chercheurs RTP-doc ». Les textes collectifs ont été réunis dans un livre : Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, C & F Éditions, Caen, 2006. Ils sont par ailleurs accessibles sur <archivesic.ccsd.cnrs.fr>.

Au commencement, la bibliothèque¹

Le modèle de la bibliothèque est souvent mal compris. Elle est généralement présentée comme le dernier maillon de la chaîne de production-diffusion du livre, celui qui permet enfin sa rencontre avec le lecteur, au même titre que la librairie dont elle ne serait qu'une version non marchande. Cette présentation se retrouve régulièrement dans les manuels sur les métiers du livre et pourtant c'est une erreur de perspective qui nous fait voir l'histoire du livre par le petit bout de la lorgnette de celle de l'édition. La figure de l'éditeur est beaucoup plus récente, moins de deux cents ans, que celle du bibliothécaire, parallèle à l'histoire du livre et vieille de plus de deux millénaires.

Erreur bénigne tant que les logiques des bibliothèques se différencient radicalement du reste de la filière de production-distribution du livre, car elle n'affectait pas la compréhension d'ensemble. Erreur gênante aujourd'hui, car elle nous interdit de percevoir l'originalité documentaire et économique du Web et *a fortiori* la transformation des bibliothèques elles-mêmes et leur réintégration dans l'économie générale de la filière. La rapide mise en perspective de ce chapitre montrera l'absurdité de l'*a priori* d'une équivalence entre le libraire et le bibliothécaire, qui ne seraient séparés que par leur appartenance ou non à un régime commercial. S'il y a bien quelques similitudes (un fonds, une salle, un service à la clientèle), la valeur créée est fondamentalement différente. L'un, le libraire, n'est en effet que l'aboutissement de la chaîne industrielle actuelle du livre imprimé. L'autre, le bibliothécaire, gère un tout autre système, un écosystème pour reprendre un terme contemporain.

Nous affirmerons dans ce livre que la bibliothèque est depuis l'origine un média au sens classique de ce mot, à l'instar de l'édition, la presse, la radio ou la télévision. Ce fut même, sans doute avec le spectacle, un des tout premiers médias de l'humanité. L'arrivée du numérique et du Web a fait sortir ce média ancien de la marginalité dans laquelle le développement de l'imprimé l'avait plongé, au point même d'inspirer des modèles de développement, commerciaux et non commerciaux. Une meilleure compréhension de la bibliothèque et de sa matière première, le document, est une clé essentielle pour comprendre le succès du Web. Mais n'anticipons pas, commençons dans ce chapitre par présenter la logique particulière du modèle de la bibliothèque.

L'affirmation d'un modèle

Pour qu'il puisse traverser les siècles en s'adaptant sans se départir de ses principes de base, il a fallu que le modèle de la bibliothèque soit à la fois simple et robuste. Sa logique générale est le partage, qui définit un écosystème autonome. L'autonomie l'a préservée des soubresauts de ses voisins plus jeunes et plus capricieuses, les industries culturelles.

Un succès qui compte

Un regard superficiel peut laisser croire que le succès des bibliothèques, variable suivant les pays, reste modeste comparé à l'explosion de la popularité des médias modernes. La réalité est bien différente. Suivant les chiffres compilés par un organisme américain de coopération entre les bibliothèques en 2003² :

- il y avait environ 1 million de bibliothèques dans le monde ;
- les dépenses annuelles étaient estimées à 31 milliards de dollars (un peu moins de la moitié aux seuls États-Unis) ;
- un habitant de la planète sur six était inscrit dans une bibliothèque ;

– les collections des bibliothèques comprenaient 16 milliards de livres (4,5 pour les bibliothèques publiques, 3,7 pour les bibliothèques scolaires, 3,5 pour les bibliothèques universitaires, 3,2 pour les centres de documentation, *special libraries*, 0,9 pour les bibliothèques nationales), c'est-à-dire environ 2,5 livres par personne.

Dans ces chiffres, la position des États-Unis est écrasante, sans doute du fait de la place que les bibliothèques tiennent dans le pays, mais aussi peut-être à cause d'un biais qui favorise le pays où les statistiques sont les plus complètes et fiables. En 2010, il y a eu 1,4 milliard d'entrées dans les bibliothèques aux États-Unis, plus qu'au cinéma (1,3) et beaucoup plus que pour les spectacles sportifs (218 millions)³. Un sondage a montré, par ailleurs, que 20,3 millions de ménages américains affirmaient avoir fréquenté l'une ou l'autre des bibliothèques publiques plus d'une vingtaine de fois dans l'année⁴. Même si l'implantation, l'activité et le succès des bibliothèques sont variables selon les pays, ces chiffres d'audience ou de financement suffisent à montrer qu'elles tiennent largement leur place dans le concert des médias.

Le partage

Le service rendu par la bibliothèque comprend deux moments. Il réunit tout d'abord dans une collection ordonnée des documents provenant de diverses sources : libraires, revendeurs ou encore administrations ou même particuliers. Le principe est de collecter, ranger, « emprisonner » en un lieu, organiser et classer des documents dispersés et jusque-là mobiles pour les mutualiser et les mettre à disposition des lecteurs d'une collectivité. Cette mise à disposition constitue le second moment. Un service de prêt ou de lecture sur place permet au lecteur de consulter les documents dans un temps volontairement limité pour ne pas priver d'éventuels autres lecteurs de leur jouissance. Les deux moments sont donc la constitution de la collection et sa mise en accès. Dans le vocabulaire de l'économie des services, on parlera de « base arrière » (*back-office*) pour le moment de la constitution de la collection, et de « face avant » (*front office*) pour celui de l'accès.

Ainsi, la bibliothèque construit un écosystème qui lui est propre. Son principe général est que le document reste à la bibliothèque, propriété de la collectivité concernée. La bibliothèque retire les documents de leur précédent circuit, commercial ou non, pour les intégrer définitivement à son propre circuit. Cet écosystème est original et autonome. Il a la faculté de pouvoir naturellement fonctionner en réseau, puisque les documents réunis dans une collection peuvent, par copie ou prêt entre bibliothèques, enrichir une autre collection fonctionnant sur le même principe. Depuis leur fondation, les bibliothèques communiquent entre elles, partagent les documents et enrichissent ainsi mutuellement leurs collections. Dans l'Antiquité par le travail des scribes⁵, puis par les moines copistes, plus récemment par le prêt entre bibliothèques et aujourd'hui par la numérisation et la mise en ligne des documents, les bibliothèques nourrissent leur écosystème.

La bibliothèque est donc une entreprise de service fondée sur le partage. Le cœur de son métier n'est pas, comme pour l'édition, de produire des biens (des objets tangibles ou intangibles, pour nous des documents) et de s'en départir, ou comme le spectacle de transmettre directement des informations au spectateur, mais d'enrichir des entités humaines par l'accès à des documents ou plus largement par l'accès au savoir contenu dans ces documents que la bibliothèque accumule et conserve. Fondée par principe sur le partage, une bibliothèque ne rapporte pas d'argent. L'entrée y est généralement gratuite, ou d'un prix symbolique. Il arrive que dans certaines organisations les prestations documentaires soient facturées, mais il s'agit de logiques comptables internes et non d'une économie de la bibliothèque proprement dite. Le financement de la bibliothèque ne s'effectue donc pas par le marché, mais par un subventionnement.

Le temps long

Si l'histoire des bibliothèques ne se confond pas avec celle de l'humanité, elle est néanmoins très longue, parallèle à celle de l'accumulation et de la transmission des connaissances depuis qu'elles sont consignées sur un support grâce à l'écriture. On trouve les premières traces de bibliothèques dans l'Antiquité en Assyrie, en Égypte, en Grèce, en Chine. Chaque fois, elles furent un lieu de conservation des documents tout autant qu'un lieu de leur production par la copie des exemplaires. La copie était nécessitée par la fragilité des supports qu'il fallait renouveler, par la volonté de diffuser les documents ou parfois simplement pour faire une offrande, comme dans les *scriptoria* bouddhiques de la Chine médiévale⁶. Les bibliothèques ont été aussi, plus largement, un lieu de production intellectuelle par les opportunités de lecture qu'elles facilitent et qui enrichissent l'écriture de ceux qui consultent les documents.

Beaucoup plus tard, vers les XII^e et XIII^e siècles pour la Corée et la Chine et le milieu du XV^e siècle pour l'Europe, l'invention de l'imprimerie à caractères mobiles a externalisé la reproduction matérielle des documents qui a quitté progressivement les *scriptoria*, c'est-à-dire le giron des bibliothèques, pour devenir

une des premières industries des temps modernes. Une activité économiquement autonome de production et diffusion des livres s'est organisée à partir, tout d'abord, des imprimeurs-libraires puis, vers la fin du XVIII^e siècle, de l'éditeur tel que nous le connaissons aujourd'hui⁷. Plusieurs symptômes, comme l'hyperconcentration des maisons d'édition et leur obligation de rentabilité, la montée des agents littéraires, et, bien sûr, l'arrivée du livre numérique, laissent présager que l'histoire de l'édition est à un nouveau tournant.

L'éditeur a pris progressivement une place dominante dans le processus de production-diffusion de l'objet-livre, et les bibliothèques ont perdu leur monopole sur l'ensemble de la filière. La production intellectuelle du livre et des connaissances en général n'a pas pour autant échappé complètement aux bibliothèques, qui sont restées un lieu familier pour les lettrés. Écrivains, professeurs, chercheurs, étudiants les fréquentent pour préparer leurs travaux et construire leurs œuvres. Par la suite, l'émergence de nouveaux médias et de nouvelles industries culturelles, la presse populaire au XIX^e siècle, les disques et la radio au début du XX^e, le cinéma, la télévision et la vidéo, ont élargi l'éventail de l'information et de la distraction pour le public. La place du livre s'est relativisée, mais chaque fois les bibliothèques ont intégré les nouveaux supports dans leurs collections sans pour autant modifier leur modèle, ni réduire leur rôle.

L'évolution des bibliothèques, comme celle de leurs confrères les centres d'archives et les musées, est en réalité parallèle à celle des sociétés et de leur rapport aux connaissances enregistrées, influant la croissance économique. Certains parlent à leur sujet d'infrastructures épistémiques (*epistemic infrastuctures*)⁸, c'est-à-dire d'institutions facilitant la préservation, l'organisation et la circulation des connaissances. Pour ne donner qu'un exemple emblématique, les travaux de Charles Darwin et la publication de son livre *De l'Origine des espèces* en 1859, et préalablement ceux de Carl Von Linné sur la classification systématique des espèces en 1738, ont été préfigurés par le succès des « cabinets de curiosités » très nombreux à Londres et Paris au XVIII^e siècle, les échanges dans les sociétés savantes, l'organisation de muséums et des bibliothèques académiques qui commencent alors à prendre de l'ampleur.

Les premiers services modernes d'archives et de bibliothèques sont nés au moment des Lumières et de la révolution scientifique. Le principe de provenance et de respect des fonds, fondement de l'archivistique moderne, est promulgué juste après la Révolution française à l'École des Chartes. De son côté, le classement des documents en catégories dans les bibliothèques commence à se systématiser à partir du XVIII^e siècle. À la fin du XIX^e siècle, au moment où les sociétés industrielles triomphent, les grandes bibliothèques universitaires sont en place, les classifications universelles s'imposent et le mouvement des bibliothèques publiques démarre pour démocratiser l'accès au savoir.

À partir de ce moment, on peut parler vraiment d'un modèle bibliothéconomique, une organisation stable systématique, qui s'étend sur l'ensemble des territoires de la planète. Parallèlement au développement explosif d'autres médias, comme l'édition de livres, puis la presse, et ensuite la radio ou la télévision, la bibliothèque a conforté son modèle, l'a précisé par touches successives, continue de le perfectionner, et ainsi a accompagné, parfois devancé, l'organisation et la diffusion des connaissances dans les sociétés. Dans une société où les connaissances circulent de plus en plus vite, cette force tranquille a un avantage. Média le plus ancien, c'est aussi celui où l'on peut s'abstraire du cycle trop rapide des médias plus jeunes qui tend à écraser les informations par leur renouvellement continu et à perdre l'attention du lecteur dans une surabondance. Le flot des médias contemporains est trop puissant, trop abondant pour autoriser un filtrage efficace. On va à la bibliothèque pour y découvrir, dans le calme, des documents que les autres médias noient dans le flux insatiable de leur production. On utilise les services d'un bibliothécaire ou d'un documentaliste afin de retrouver des informations utiles perdues dans le chaos général. Ainsi la bibliothèque est-elle le média du temps long, s'adaptant à l'évolution des sociétés et tempérant la précipitation des médias plus jeunes, plus tempétueux et plus éphémères. Cette qualité, loin de la rendre obsolète, est un levier sur lequel la bibliothèque peut appuyer pour répondre aux défis nouveaux du numérique.

La valeur créée

Qu'est-ce qui pousse une collectivité à financer une bibliothèque ? Pourquoi fonder et entretenir une bibliothèque si les documents sont disponibles et que chacun peut se les procurer individuellement ? Sans doute, il existe des pays encore nombreux ou des situations informationnelles particulières dans lesquels la disponibilité des documents ne va pas de soi et l'utilité de la bibliothèque est indiscutable. Mais la tendance

générale est plutôt à la surabondance d'informations. Alors la question mérite réponse, d'autant qu'aujourd'hui elle est posée de plus en plus souvent par les collectivités qui hésitent à investir dans les bibliothèques, arguant de l'accessibilité de l'information sur le Web.

Le retour sur investissement

La réponse classique à ce genre de question est le rapport entre les avantages et les coûts. S'il est positif, alors l'investissement serait justifié. Un grand nombre d'études empiriques ont été menées dans ce sens par des bibliothécaires soucieux de défendre leurs intérêts ou ceux de leur institution et de leur public auprès de leurs bailleurs de fonds. Mais les calculs ne sont pas simples et toujours critiquables, car les avantages sont difficilement quantifiables.

Deux rapports américains récents⁹ sont considérés comme des références pour les retours sur investissement des bibliothèques publiques américaines. Ils proposent une synthèse des études déjà réalisées et ont inspiré la plupart de celles qui ont suivi. Leurs auteurs soulignent la possibilité nouvelle d'intégrer des éléments plus intangibles et de mesurer la valeur de l'apprentissage (*learning values*) ou des bénéfices culturels (*cultural benefits*) grâce au concept de retour social sur investissement (*social return on investment – SROI*) qui gagne du terrain dans le monde de l'entreprise. Le premier rapport passe en revue dix-sept études réalisées au début des années 2000. Les résultats, sans surprise, sont très favorables aux bibliothèques avec des retours sur investissement allant de 3 à 6,54 \$ pour un dollar investi. Le second s'intéresse à l'impact des bibliothèques sur le développement urbain aux États-Unis. Après avoir souligné le rôle des bibliothèques publiques pour la formation des jeunes, pour la recherche d'emploi, pour l'information pour les petites et moyennes entreprises (PME) et très petites entreprises (TPE) et pour animer les zones de chalandise, il conclut : « [...] les bibliothèques publiques sont bien placées pour alimenter non seulement la nouvelle, mais aussi la prochaine économie par leur rôle dans la construction des compétences technologiques, et dans l'activité entrepreneuriale, et par leurs lieux vivants et accueillants. La combinaison entre un rôle plus important dans les stratégies de développement économique et leur omniprésence – 16 000 antennes dans plus de 9 000 systèmes – fait des bibliothèques publiques des outils stables et puissants pour les villes qui cherchent à construire une économie solide et résistante¹⁰. »

Les valeurs ajoutées

Mais ces calculs considèrent la bibliothèque comme une « boîte noire », sans s'intéresser à l'originalité de son modèle. Lorsque l'on ouvre la boîte, il faut s'interroger alors sur la valeur ajoutée par la collection que l'on a constituée et que l'on entretient à grands frais. Trois bénéfices peuvent être repérés.

Le premier, le plus immédiat, est le gain réalisé par la mutualisation des documents. Puisqu'un lecteur peut réutiliser un document déjà utilisé par un autre, on réalise des économies d'échelle en mutualisant les documents. Plus le prix d'achat initial est élevé (en argent, en temps ou en énergie) pour le consommateur, plus le gain sera fort. Ainsi, cette valeur est surtout appréciable dans des situations de pénurie. De plus, en période de crise économique, la bibliothèque permet aux catégories sociales les plus touchées de garder un contact avec la culture et le savoir enregistrés. Mais les études bibliométriques ont montré que la distribution de la demande dans une collection suit une loi de puissance¹¹. Pour une période et une population données, les consultations de documents se concentrent sur un petit nombre de documents très demandés et se dispersent sur un très grand nombre peu demandés. Pour satisfaire ses lecteurs, une bibliothèque se doit d'acheter rapidement plusieurs exemplaires des documents les plus demandés, sans se départir pour autant d'une large collection. Dès lors, l'avantage de la mutualisation est très relatif.

Le deuxième bénéfice de la constitution d'une collection est le gain d'opportunité. Une collection où les documents sont rangés et classés permet au lecteur de trouver plus facilement ou plus rapidement le document ou l'information qu'il cherche, puisque quelqu'un les a déjà collectés et rangés pour lui. Ensuite, en cherchant un document, il risque d'en trouver d'autres, proches de ses préoccupations, auxquels il n'aurait pas pensé, qu'il n'aurait pas trouvés tout seul et qui lui ouvriront d'autres possibilités, d'autres portes par rapport à son besoin initial. Aujourd'hui, ce deuxième volet a gagné un nom sur le Web, on y parle de découverte (*discovery*) ou de sérendipité, mais les lecteurs n'ont pas attendu le Web pour faire des découvertes en bibliothèque.

La dernière valeur ajoutée de la mise en collection est le gain en potentialité, ou valeur d'option. Il découle de sa fonction de conservation et s'apparente à une sorte d'assurance ou, pour prendre une notion plus contemporaine, au principe de précaution. Tandis que le libraire retourne au fur et à mesure ses invendus au distributeur, le bibliothécaire ne se défait qu'avec parcimonie et non sans réticence des documents qu'il accumule. Les élagages ou désherbages sont réguliers dans les bibliothèques publiques,

mais suivent un protocole précis. Il est ainsi préférable de garder, de façon parfois mutualisée, au moins un exemplaire d'un document plutôt que de devoir le rechercher sans garantie de succès, en reconstituer *a posteriori* le contenu ou, pire, de devoir s'en passer. Ainsi, collecter des documents, c'est aussi accroître la potentialité documentaire des lecteurs dans le temps, c'est prendre une assurance sur l'avenir. Même si le bibliothécaire constitue sa collection en fonction des besoins de la collectivité qu'il sert, personne ne peut connaître vraiment *a priori* et à coup sûr la valeur d'usage d'un document au moment de son acquisition et encore moins à l'avenir. Un livre, un document est un « bien d'expérience », c'est-à-dire que l'on ne connaît sa valeur qu'en le consommant et cette valeur est dépendante d'un contexte éminemment variable dans le temps.

Cette dernière valeur ajoutée est une évidence si l'on passe à l'échelle de l'histoire. Sans les bibliothèques, notre patrimoine documentaire serait perdu. Il est même sûr que, faute de bibliothèques à certaines époques et dans certaines collectivités, ou encore du fait de la destruction de bibliothèques, une bonne part a déjà été perdue sans qu'il soit possible de mesurer la valeur de cette perte. Sans remonter dans le passé, nous avons aussi des exemples très contemporains spectaculaires de l'importance de la valeur créée. C'est grâce aux collections conservées dans les bibliothèques que la firme Google a pu numériser des millions de livres et les mettre à disposition sur le Web. Un immense patrimoine documentaire a pu être ainsi revalorisé et exploité de nouveau, à un coût bien moindre que s'il avait fallu les réunir de nouveau, créant une richesse difficilement prévisible, même si bien des débats restent encore ouverts sur la forme et la répartition de la valeur récupérée par son exploitation commerciale.

Modèle bibliothéconomique

Au cours des siècles, les bibliothécaires ont développé des compétences adaptées à l'écosystème dont ils avaient la charge. Elles ont été déclinées selon des savoirs professionnels (plans de développement de collection, règles de catalogage, bulletinage, systèmes de prêt, mobiliers, architecture, marketing, etc.) et regroupées sous un terme : la bibliothéconomie. Parfois, les bibliothécaires ont confondu autonomie et enfermement, privilégiant le système pour lui-même plutôt que le service rendu, perfectionnant à outrance leurs outils et leurs règles, défendant le statut plutôt que la fonction. Mais ces crispations ne doivent pas occulter l'immense travail réalisé, l'adaptation des institutions et l'importance de ces infrastructures épistémiques pour le développement des sociétés. Pour souligner l'apport du modèle bibliothéconomique, on peut le représenter sur un tableau selon le niveau de son bénéficiaire.

Tableau 1. *Le modèle bibliothéconomique*

Destinataire	Base arrière	Face avant	Produit
Individu	Collection	Accès	Mémoire
Collectivité	Information	Action	Capital
Société	Patrimoine	Transmission	Empreinte

La première ligne reprend les deux moments de la production du service de la bibliothèque pour un lecteur individuel : la configuration de la collection en base arrière et l'accès aux documents en face avant pour constituer à son profit une mémoire externe permanente dans laquelle il pourra puiser. La deuxième ligne confirme ces mêmes moments en raisonnant sur le service rendu non plus à un individu, mais à une collectivité. Il s'agit alors de constituer un capital par la collecte d'informations qui auront le potentiel d'alimenter l'action collective ou au moins une partie de celle-ci, c'est-à-dire la réalisation de quelques objectifs de la collectivité. Ces objectifs participent à la configuration de la collectivité qui réunit les individus. La collectivité investit dans ce capital informationnel pour que les individus répondent mieux à ses objectifs. La troisième ligne représente toujours le service rendu par la bibliothèque, mais cette fois à l'échelle de la société tout entière. La collection est devenue le patrimoine informationnel commun qu'il s'agit de transmettre. L'ensemble constitue l'empreinte informationnelle de cette société, autrement dit la part documentaire de la civilisation concernée, contribution de la collectivité et nourriture intellectuelle de l'individu. Les informations documentaires non collectées risquent d'être perdues à jamais pour la société.

On le voit maintenant clairement, la valeur créée par la bibliothèque est indépendante du modèle de l'édition, qui n'est qu'une source éphémère et ponctuelle pour alimenter le service.

Externalisations numériques

Le numérique, et plus précisément le Web, a ébranlé le modèle de la bibliothèque, à l'instar de celui de tous les médias. Comme au moment de la mise en place de l'industrie du livre avec l'imprimerie, une part du service rendu par les bibliothécaires a été externalisée. Mieux, le modèle bibliothéconomique lui-même a servi d'inspiration pour des industriels, qui ont trouvé le moyen d'en automatiser les services. Ils ont alors permis des performances documentaires jusque-là inconnues, notamment par le traitement du texte intégral, et ont réussi à les valoriser commercialement, parfois avec la complicité plus ou moins forcée des bibliothécaires eux-mêmes. Ironiquement, la transposition commerciale du modèle bibliothéconomique démontre *a posteriori* sa valeur économique. Sans entrer dans trop de détails, voici pour conclure ce chapitre quelques illustrations de cette nouvelle donne.

Collections numériques

La publication dans le monde scientifique est devenue, depuis la seconde moitié du ^{xx}e siècle, à la fois une modalité de diffusion de ses résultats et un mode de régulation de son économie générale. Le nombre de publications et leur qualité, mesurée à l'aune de la notoriété de la revue qui les accueille, sont pour le meilleur et le pire l'étalon du chercheur sur le marché de la recherche, tant pour sa carrière que pour l'obtention de subventions pour ses projets. Cette mesure a accéléré une explosion quantitative des publications et l'augmentation de leur valeur commerciale. Il s'est alors installé une relation étroite entre les éditeurs de revues scientifiques et les bibliothèques académiques, les dernières étant les principaux clients des premiers. Mais, au début des années 1990, l'inflation des titres et des tarifs est arrivée à un point de rupture.

Le Web a été perçu, par les éditeurs et les bibliothèques, comme une opportunité pour réviser leur position respective. Elsevier, l'un des principaux éditeurs de littérature scientifique, suite à une expérimentation dans des bibliothèques académiques américaines réalisée au milieu des années 1990, a proposé un modèle économique nouveau, fondé sur une licence forfaitaire, un droit d'entrée, payé par la bibliothèque, pour un accès en texte intégral de l'ensemble de ses usagers à l'ensemble des titres de l'éditeur. Compte tenu de coûts marginaux pratiquement nuls, le plus logique, en effet, était de faire payer un ticket d'entrée forfaitaire, d'autant que l'expérimentation avait montré l'intérêt des utilisateurs pour une plateforme d'accès unique. Elsevier, rachetant des titres de revues pour les mettre en ligne, a rapidement pris une position dominante sur ce marché, formant avec quelques autres éditeurs (Springer, Wolter-Kluwer, Wiley) un oligopole capable d'imposer son prix. Cette organisation s'est rapidement répandue comme la seule viable pour les revues scientifiques du point de vue des éditeurs et s'est déclinée de façon variable selon des licences négociées en position plus ou moins forte entre des éditeurs accumulant de plus en plus de titres et des bibliothèques réunies en consortiums. Elle revient en fait à déporter vers l'éditeur le modèle bibliothéconomique en lui conférant un statut commercial. La bibliothèque traditionnelle est réduite à un rôle d'agent, représentant la collectivité et négociant le ticket d'entrée à la collection pour ses membres. Reed-Elsevier est depuis quelques années le premier groupe d'édition mondial. Il doit cette place en grande partie à sa position de leader dans les revues électroniques scientifiques.

Recherche d'information

Google, on le sait, est d'abord un moteur de recherche sur le Web. L'idée première de ses fondateurs était d'adjoindre aux techniques habituelles de recherche d'informations une variable supplémentaire, la notoriété mesurée par le nombre de liens hypertextuels qui font référence à une page donnée. Cette idée n'est pas originale en soi, puisqu'elle reprend celle mise au point dès les années 1960 par Eugene Garfield pour fonder le classement des revues scientifiques : l'analyse des citations, à la base de leur régulation évoquée précédemment. Mais le génie des fondateurs de Google est de l'avoir appliquée à tous les documents numériques et ainsi d'autoriser un calcul et donc un classement en temps réel par le désormais fameux *Pagerank*. Afin de pouvoir effectuer ses calculs et le traitement automatique des textes, il a fallu copier dans d'énormes centres de données l'ensemble des sites Web et les reproduire en copie-cache. Pour cela, la firme s'est appuyée sur la notion d'usage équitable (*fair use*), c'est-à-dire la règle de droit sur laquelle se fondent les services de bibliothèque aux États-Unis. Ainsi, Google a repris le second moment du service de la bibliothèque, qu'il a financé par la publicité en vendant des mots clés aux annonceurs. Google est un des succès industriels et commerciaux les plus éclatants de ce début de millénaire. Nous analyserons plus en détail sa stratégie dans le dernier chapitre de ce livre.

Adaptations ou éclatement

Les bibliothécaires, habitués à s'adapter aux transformations des systèmes informationnels, ont aussi réagi très tôt à l'émergence de ce nouveau média, à la fois familier et inquiétant pour eux puisqu'il s'inspirait de leur modèle. Dès les années 1970, l'informatique est entrée dans les bibliothèques pour améliorer la gestion des documents et les outils d'accès, à commencer par le catalogue. Chaque innovation numérique (e-books, *cloud-computing*, Web de données, réseaux sociaux, etc.) trouve un écho dans les bibliothèques qui disposent de leurs veilleurs blogueurs et, contrairement à la légende, de professionnels entreprenants et innovateurs.

Aujourd'hui, dans les bibliothèques universitaires comme dans les bibliothèques publiques, un mouvement tend à faire du lieu physique le centre du service. L'accent est mis sur le caractère ouvert, convivial, flexible des bibliothèques. Ainsi, les bibliothèques universitaires deviennent des *learning centers*¹² où les étudiants travaillent confortablement, en groupe, interrogeant éventuellement différentes ressources documentaires, internes ou externes, mais surtout disposant de tous les outils de rédaction, de recherche, de calcul, de traitement et de communication. Les bibliothèques publiques quant à elles deviennent des « troisièmes lieux » (*third place libraries*¹³), s'éloignant d'une image de temples et de gardiens du savoir pour se rapprocher de la convivialité sociale du café où l'on échange des informations.

La concurrence de l'offre directe sur le Web est maintenant bien là et l'écosystème des bibliothèques s'est ouvert, autorisant des comparaisons. Bien des expérimentations, bien des batailles se sont tenues et se tiendront encore sur les droits numériques et sur les verrouillages techniques associés aux fichiers numériques, mais il est clair que, si elles concernent souvent des mises en collection et en accès, elles dépassent largement le monde des bibliothèques. Alors, faut-il conclure que les bibliothèques s'adapteront à ce nouveau média ou bien qu'au contraire elles s'éparpillent dans la Toile ? Deux raisons, au moins, doivent nous conduire à la prudence, sinon à l'optimisme, pour ces institutions. D'une part, le Web n'est pas stabilisé et, malgré quelques réussites spectaculaires, il n'a pas encore réellement trouvé un modèle économique en harmonie avec ses partenaires médiatiques. Aussi, on peut penser qu'avec le temps les valeurs bibliothéconomiques de mutualisation, d'opportunité et d'option vont refaire surface. D'autre part, le modèle de la bibliothèque n'a pas perdu sa pertinence pour les documents traditionnels. Aujourd'hui encore, et paradoxalement plus qu'hier, le livre imprimé reste la première image de marque de la bibliothèque publique en Amérique du Nord¹⁴. Le numérique ébranle la bibliothèque, mais chaque média émergent a obligé celle-ci à se repositionner, à commencer par la naissance de l'industrie du livre avec l'imprimerie. Il ne s'agit que d'une nouvelle étape de son histoire.

Notes du chapitre 1

1. Ce chapitre reprend quelques éléments du même auteur déjà publiés dans : « Immeasurable library economics », in Ilde RIZZO & Anna MIGNOSA (éds), *Handbook on the Economics of Cultural Heritage*, Edward Elgar Publishing, Cheltenham (UK) and Northampton, MA (USA), 2011.
2. ONLINE COMPUTER LIBRARY CENTER (OCLC), *Libraries : How they stack up, Report*, Dublin, Ohio, USA, 2003.
3. ONLINE COMPUTER LIBRARY CENTER (OCLC), *How Libraries Stack Up : 2010 Report*, Dublin, Ohio, USA, 2010.
4. AMERICAN LIBRARY ASSOCIATION, *State Of America's Libraries Report 2011*, Chicago, USA, 2011.
5. Jean-Jacques GLASSNER, « Des dieux, des scribes et des savants », *Annales. Histoire, Sciences sociales*, 3/2005 (60^e année), p. 483-506.
6. Jean-Pierre DRÈGE, « Les *scriptoria* bouddhiques dans la Chine médiévale (VI^e-X^e siècles) », in Christian JACOB (dir.), *Lieux de savoirs – Espaces et communautés*, Albin Michel, Paris, 2007, p. 515-536.
7. Jean-Yves MOLLIER, « Naissance, développement et mutations de l'édition de l'*Encyclopédie* de Diderot à Internet », Maison méditerranéenne des sciences de l'homme, 22 mai 2003, disponible sur <imageson.org>.
8. Margaret HEDSTROM et John Leslie KING, « Epistemic infrastructure in the rise of the knowledge economy », in Brian KAHIN et Dominique FORAY, *Advancing Knowledge and the Knowledge Economy*, MIT Press, Cambridge (Mass.), USA, 2006, p. 113-134.
9. AMERICANS FOR LIBRARIES COUNCIL, *Worth Their Weight : An Assessment of the Evolving Field of Library Valuation*, Report, New York, USA, 2007 ; URBAN LIBRARIES COUNCIL, *Making Cities Stronger : Public Library Contributions to Local Economic Development*, Report, USA, 2007.
10. *Ibid.*, p. 3 (traduction personnelle).
11. La bibliométrie a relevé cette distribution dès les années 1930. Pour une présentation détaillée : Thierry LAFOUGE, Yves LECOADIC et Christine MICHEL, *Éléments de statistique et de mathématique de l'information*, Presses de l'Enssib, Villeurbanne, 2002. Chris Anderson l'a présentée comme une caractéristique fondatrice de l'économie du Web : Chris ANDERSON, *The Long Tail : Why the Future of Business Is Selling Less of More*, Hyperion, New York, USA, 2006.
12. JOINT INFORMATION SYSTEMS COMMITTEE (JISC), *Designing Spaces for Effective Learning : a guide to 21st century learning space design*, Higher Education Funding Council for England (HEFCE) on behalf of JISC, Bristol, UK, 2006.
13. Harris CATHRYN, « Libraries with lattes : the new third place », *Australasian Public Libraries and Information Services APLIS*, 1^{er} décembre 2007.
14. ONLINE COMPUTER LIBRARY CENTER (OCLC), *Perceptions of Libraries, 2010 : Context and Community*, Report, Dublin, Ohio, USA, 2011.

Les mutations du document

Les bibliothèques se sont développées et adaptées à l'évolution des sociétés, collectant des documents pour les proposer à leurs usagers. Mais les documents eux-mêmes se sont aussi transformés. Leur nombre, leur forme, leur contenu et leur fonction ont changé en lien avec l'organisation des sociétés et aussi avec les performances quantitatives et qualitatives des technologies de représentation : imprimeries, enregistrements, calculs. Curieusement, il paraît plus difficile de cerner la notion de cet objet apparemment élémentaire, le document, que celle des institutions qui le collectent, bibliothèques, centres d'archives et musées.

Plusieurs ont tenté de la préciser, sans qu'il y ait encore aujourd'hui un consensus autour d'une théorie opérationnelle. En France, un courant récent a proposé quelques avancées, sur lesquelles ce livre appuie sa proposition. Pour mieux comprendre les changements du numérique, nous reprendrons l'éclairage selon trois angles complémentaires du document : sa forme, son texte ou son contenu et enfin sa qualité relationnelle ou transactionnelle de médium. Cette partition permet d'appréhender bien des changements en cours. Le numérique a donné une accélération supplémentaire à une documentation déjà considérée au siècle dernier comme quantitativement explosive. Celle-ci touche maintenant toutes les activités humaines, jusqu'aux humains eux-mêmes, non sans poser des questions sociales et éthiques à une échelle inédite.

Petite histoire du mot

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. Mais l'usage courant du mot « document » a moins de deux siècles, même si la réalité de sa fonction est bien plus ancienne.

Un enregistrement

Cette première définition met l'accent sur le support. L'histoire des documents serait alors parallèle à celle des supports sur lesquels l'information est consignée. C'est dans cette perspective d'abord que les premiers savoir-faire documentaires, la bibliothéconomie et l'archivistique, se sont construits. Il s'agissait de gérer des objets, en privilégiant la dimension matérielle. La « fluidification » des supports apportée par le numérique, avec la possibilité de basculer de l'un à l'autre, a justifié par la suite la priorité mise sur l'« information », c'est-à-dire le contenu du document. Les sciences de l'information sont nées dans les années 1970 de cette relativisation du support. On s'intéresse moins à l'objet, puisqu'il est changeant, qu'à l'information qu'il véhicule. Néanmoins, l'information reste enregistrée sur un support, passant éventuellement de l'un à un autre. Ainsi, quand on parle d'information dans les sciences du même nom, il s'agit, sans qu'il soit nécessaire de le préciser, d'une information consignée.

Depuis les débuts de l'écriture, quelque part 4 000 ans avant notre ère, jusqu'aux mémoires informatiques d'aujourd'hui, l'histoire de l'information consignée est longue et balisée par des inventions technologiques majeures en résonance avec l'esprit de l'époque qui les a vues naître. Il ne manque pas d'historiens et de penseurs pour proposer leur découpage et leur interprétation. Le sujet a été beaucoup discuté et est suffisamment connu pour qu'il ne soit pas nécessaire d'y revenir ici.

Contentons-nous de rappeler très schématiquement quelques étapes essentielles : le passage du rouleau (*volumen*) au livre relié (*codex*), au début de notre ère, contemporain en Occident de la progression du christianisme ; le passage du manuscrit au livre imprimé à partir du *xv^e* siècle en Occident (quelques siècles

plus tôt pour la Corée et la Chine), contemporain de la mise en place du capitalisme et de la popularisation de la Réforme – le livre est roi ; la rotative et le développement du journal au XIX^e siècle, avec la révolution industrielle et les démocraties parlementaires – revues et journaux se multiplient ; l'enregistrement analogique, photographie puis cinéma, disque au tournant des XIX^e et XX^e siècles et la vidéo cinquante ans plus tard, la communication de masse, la propagande et l'opinion publique – affiches et audiovisuel nous interpellent ; la machine à écrire, les machines à imprimer légères puis les photocopieuses au cours du XX^e siècle, la bureaucratie et le développement du commerce – la « paperasse » nous envahit ; et enfin l'ordinateur au milieu du même siècle et le Web à sa fin, l'économie de service et la mondialisation – c'est le règne des fichiers et l'addiction aux terminaux (micro-ordinateurs, téléphones intelligents, tablettes).

On peut faire une lecture de l'histoire du document à partir des nombreux travaux directs ou connexes sur l'histoire des supports, même si le terme « document » y est rarement employé. Les travaux sont fournis, riches, et donnent lieu à de nombreux débats et développements soulignant et affinant l'analyse des rapports étroits entre l'économique, le politique, le culturel et les différents états des technologies de communication. De ce point de vue, le numérique peut se lire dans la continuité d'une évolution de plusieurs millénaires. Récemment, un professeur danois, Thomas Petitt¹, a même suggéré que nous vivions une révolution à rebours. Passé la « parenthèse Gutenberg », nous retrouverions la relation à l'écrit du Moyen Âge. Ces approches laissent néanmoins souvent dans l'ombre des questions importantes, comme l'évolution de la signification et de la fonction du document.

En réalité, en employant le mot « document » dans le bref survol précédent, nous avons fait un anachronisme. Le terme n'est guère employé avant le XVIII^e siècle.

Des *documens* du Moyen Âge...

Issu du latin *documentum*, il apparaît de façon épisodique au XIII^e siècle le plus souvent sous la forme plurielle « *documens* » ou « *documenz* ». Le *Dictionnaire du Moyen Français* du laboratoire de recherches l'ATILF (Analyse et traitement informatique de la langue française) donne ces deux acceptions : « A. – Leçon, enseignement [...] ; B. – Acte écrit qui sert de preuve. » Ainsi, même s'il reste peu employé, le terme contient depuis le Moyen Âge deux sens complémentaires qui se sont affirmés au cours des siècles avec la mise en place de procédures spécifiques : transmettre et prouver.

Mais, différence importante pour la première acception, au Moyen Âge les *documens* sont d'abord des leçons qui ne sont pas nécessairement consignées par écrit. Le *doceo* latin signifie « enseigner », *documentum* est l'action d'enseigner. Sans doute les réticences de Socrate vis-à-vis de l'écrit² sont passées depuis longtemps, mais ce qui compte alors ce sont moins les livres dans lesquels sont consignés les savoirs que la mémoire humaine qui est exercée à les retenir et dont les livres ne sont qu'un adjuvant. Mary Carruthers, médiéviste spécialiste des arts de la mémoire, rappelle : « Le grec ancien ne possédait aucun verbe signifiant proprement “lire” ; celui dont on se servait, *anagignosco*, signifie “savoir de nouveau, se remémorer”. Il renvoie à une procédure mnésique. De même, le verbe latin signifiant la lecture est *lego*, littéralement “cueillir” ou “rassembler”³. » On pourrait presque dire que les « *documens* » du Moyen Âge étaient dans leur première acception les textes appris par cœur, les livres n'étant que des aide-mémoire, d'ailleurs peu sûrs car soumis aux erreurs ou réinterprétations des copistes. Les techniques pour exercer la mémoire, les *Ars memorialis*, étaient alors valorisées, perfectionnées et efficaces et la lecture elle-même se faisait à haute voix. Le jeu entre la mémoire orale et la mémoire écrite n'est pas l'apanage du Moyen Âge occidental⁴.

La seconde acception, celle de preuve, faisait référence à des titres, souvent des titres de propriété. L'anglais « *record* », littéralement « enregistrement » que l'on traduit souvent par « document », rend bien compte de cette fonction. Voici ce que dit l'*Oxford Dictionary* de son étymologie : « Moyen anglais : du vieux français *record* “souvenir”, du latin *recordari* “se souvenir”, construit sur *cor*, *cord*, le “cœur”. Le nom a été très tôt utilisé dans le droit pour signifier la preuve écrite. Le verbe signifiait originellement “relater oralement ou par écrit” et aussi “répéter pour garder en mémoire”⁵. » Ainsi selon leur étymologie, le « document » français et le « record » anglais ont quasiment la même signification, mais proviennent de deux mots latins différents : le premier fait référence à la transmission et la leçon, le second à la mémoire et la preuve.

Il semble que les documents légaux, notamment pour fixer la propriété, aient été développés très tôt. Michael T. Clanchy a montré la très grande importance quantitative des chartes écrites en Angleterre, dès le début du premier millénaire, allant jusqu'à faire l'hypothèse que « huit millions de chartes pourraient avoir été écrites au XIII^e siècle seulement pour les petits propriétaires et les serfs⁶ ». Mais ces titres n'avaient pas vocation à être diffusés ; bien au contraire, ils étaient gardés sous clé et sous bonne garde par les notables et les seigneurs pour pouvoir être retrouvés et produits en cas de litige.

... aux documents de la révolution scientifique

Maintenues au premier Moyen Âge respectivement par l'Église et les couvents, d'une part, et par les princes et les notables, de l'autre, ces fonctions de transmission et preuve se sont autonomisées et professionnalisées, dans leur dimension sociale et dans leur dimension gestionnaire. Professeur et bibliothécaire pour la première, notaire et archiviste pour la seconde sont devenus les figures principales de la production et de la gestion du document dans l'une et l'autre de ses fonctions. Le professeur est celui qui transmet, tandis que le bibliothécaire gère les véhicules de la transmission. Le notaire est le « tiers de confiance » qui valide l'information inscrite, lui conférant valeur légale de preuve que l'archiviste va garder et préserver. Tandis que la production de documents s'élargit progressivement du fait de leur utilisation de plus en plus importante, tant pour la transmission que pour la preuve, et qu'elle profite des progrès des techniques de reproduction rappelées plus haut, la gestion des documents se rationalise progressivement avec la mise en place des institutions documentaires, bibliothèques et centres d'archives, évoquées au chapitre précédent. La notion moderne de « document », qui amalgame les deux acceptions, résulte sans doute du lent croisement des pratiques légales et confidentielles et des pratiques monastiques de transmission, plus nobles et publiques. Elle résulte aussi, et peut-être plus sûrement encore, de la transformation du rapport à la vérité qui n'est plus révélée, préservée et donnée par la religion et le Prince, mais construite par la raison et la démonstration.

Quoi qu'il en soit, le terme « document » ne semble pas avoir été très usité jusqu'à la fin du XVIII^e siècle, quelle que soit son orthographe en français comme en latin. Par exemple, une recherche des occurrences du mot dans l'*Encyclopédie* de Diderot et d'Alembert (1751-1772) donne un résultat décevant. On pouvait s'attendre à une multiplication d'un mot synonyme de « renseignement » et « preuve » dans ce genre d'ouvrage. On y trouve seulement cette très courte définition : « Documens, s. m. pl. (Jurisprud.) sont tous les titres, pièces, & autres preuves qui peuvent donner quelque connoissance d'une chose. (A) », ainsi que deux occurrences anecdotiques, l'une dans l'article « Langue » et l'autre dans l'article « Sensibilité, sentiment ». L'ouvrage manifeste pourtant le succès de l'esprit scientifique et son ouverture à la société, mais la science expérimentale dans ses fondements modernes n'émerge vraiment qu'au XIX^e siècle, celui de Charles Darwin et de Claude Bernard, et c'est alors que l'on aura vraiment besoin du document pour transmettre en prouvant. L'utilisation moderne du mot ne prend son envol qu'à partir de ce siècle pour entrer dans le langage courant aux siècles suivants.

Une interrogation de son occurrence dans la base de livres numérisés par Google⁷ donne une courbe très explicite. L'ordonnée représente le pourcentage de livres dans lesquels le mot est présent (par exemple, en 2000, la somme des occurrences du singulier et du pluriel indique que « document(s) » est présent dans 1,1 % des livres en français numérisés par Google réunis dans la base N-Gram). On observe l'extinction de l'usage de l'orthographe « *documens* » dans la première moitié du XIX^e, puis un envol de l'usage du terme « document », principalement pour son pluriel qui atteint son apogée au début du XX^e avec juste une chute, dans les années 1930, peut-être due à la crise économique. Un bref survol des titres contenant le mot montre qu'il s'agit principalement de revues, historiques puis scientifiques. Dès lors, il est probable que ce succès découle directement de la révolution scientifique et industrielle qui avait besoin d'outils puissants à la fois pour transmettre, convaincre et prouver. Le document s'affirme comme un véhicule privilégié pour la circulation du savoir, l'affirmation des techniques et certainement aussi la stabilité des organisations et le développement du commerce. Rappelons-nous que les « infrastructures épistémiques » évoquées au chapitre précédent se multiplient au XVIII^e siècle pour s'organiser de façon systématique au siècle suivant, notamment en muséums, bibliothèques académiques ou revues savantes. La popularité grandissante du mot reflète sans doute ce vaste mouvement, avec le retard et l'inertie propres à la langue écrite.

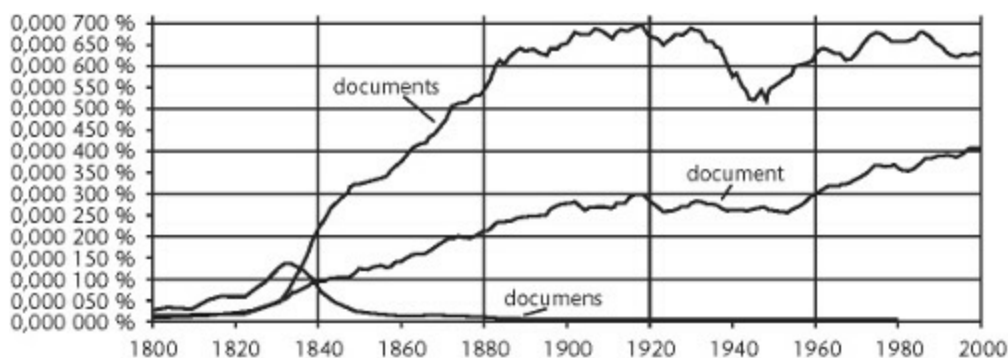


Figure 1. Proportion des livres comprenant les mots *documents*, *document* et *documents* dans la base de livres français numérisés de Google de 1800 à 2000

Pour conclure cette petite histoire du mot, signalons qu'on ne trouve pas moins de 15 281 occurrences de « document » au singulier ou au pluriel début 2011 dans la version française de l'« encyclopédie » Wikipédia⁸, qui *mutatis mutandis* joue pour les internautes aujourd'hui un rôle de référence et de vulgarisation comparable à celui de son ancêtre du XVIII^e. Nous sommes maintenant envahis par les documents, dans nos poches, sur et dans nos meubles, sur nos lieux de travail, à notre domicile comme dans les espaces publics et plus encore partout sur la Toile. Nous parlons de « documents » à propos de tout, au point que le mot s'est quelque peu vidé de sa substance. Le terme est banal, il s'est dévalorisé. Cela n'est pas anodin.

Petite histoire de la notion

La notion de « document » évolue logiquement à partir de l'arrivée de l'imprimerie, et donc de l'autonomisation de la reproduction des textes. Mais ce n'est qu'au début du XX^e siècle, quand il est devenu omniprésent, qu'on a cherché à le préciser et le définir. Il est impossible de décrire ici l'ensemble des mouvements qui ont contribué à la configuration du document moderne, il faudrait un livre entier, écrit par une palette d'experts des domaines les plus divers⁹. Tout le social est concerné, depuis la famille, l'école, les administrations, l'organisation du travail, celle de l'industrie et du commerce, les loisirs, la culture, la définition des États, le développement de l'espace public, les échanges internationaux et, bien sûr, la technique et la science. Les documents participent à la régulation de tous les rouages de la société. Je me contenterai ci-dessous de quelques allusions aux événements les plus importants et les plus significatifs pour mon propos.

Auteurs et chercheurs

Les possibilités ouvertes par l'imprimerie ont conduit à la diffusion de documents variés sous forme d'exemplaires à une échelle inconnue jusque-là. La reproduction des textes était auparavant confiée à des scribes et copistes qui parfois cherchaient à les « améliorer » ou les commenter. Ainsi, le texte écrit n'était pas stable, l'auteur pas toujours bien défini et la pensée plus rhétorique qu'analytique. L'imprimerie, en fixant les versions des textes et multipliant les exemplaires, a changé la donne. À la suite d'Henri-Jean Martin, Elizabeth Eisenstein ou encore Adrian Johns¹⁰, de très nombreux travaux d'historiens ont souligné combien l'arrivée de cette technologie a accompagné la transformation de la société. Diffusion de la Réforme, Renaissance des arts ou fondation d'une science moderne figurent parmi les changements, portés et encouragés par le développement de l'imprimerie à caractères mobiles. Je retiendrai surtout pour mon propos l'émergence de deux figures, celle de l'auteur et celle du scientifique.

La possibilité de reproduire les livres en grand nombre s'accompagne de l'organisation progressive d'un marché qui repose sur des textes susceptibles de rencontrer une demande. « On en vint donc à imaginer d'autres catégories et tout naturellement le frivole, l'inventé, le fictif commencèrent à prendre leur place aux côtés des bibles et des lourds traités de théologie ou de philosophie. [...] C'est dans le contexte de la valorisation de l'éphémère et du faux que l'auteur émerge », rappelle Jean-Claude Guédon. Et plus loin il ajoute : « L'écrivain, se décrivant à la façon romantique, se déclarait créateur, à l'instar d'un dieu, et, souverainement, pouvait ensuite transférer certains de ses droits – les droits dits « patrimoniaux » – à des

maisons d'édition ; cependant, son statut de créateur transcendantal lui interdisait de renier ou d'abandonner la paternité de sa création qui bénéficiait d'une inviolabilité totale¹¹. » D'une certaine façon, le document, quittant le giron de l'autorité transcendantale de l'Église et du seigneur, a trouvé dans l'auteur son nouveau mentor, du moins pour une partie de sa production. En deçà de la transcendance, la dimension économique du droit d'auteur est tout aussi essentielle. Les droits patrimoniaux vont permettre l'émergence d'une production riche et variée de titres, et, s'étendant aux spectacles, conduiront au développement d'un secteur économique à part entière, les industries culturelles.

Parallèlement, et selon un processus sensiblement différent, le savant se transforme progressivement en chercheur. Ici, c'est moins la pression d'un marché (qui conduit plutôt à la multiplication de mauvais livres scientifiques) que les possibilités de normalisation, l'économie de temps de transcription et les échanges entre savants qui sont déterminants. Ainsi, les *Philosophical Transactions of the Royal Society of London*, première revue scientifique moderne fondée en 1665, sont à la fois un moyen de diffusion des connaissances scientifiques et un premier outil de régulation de la science permettant d'attribuer la paternité d'une découverte à son auteur. Dès lors, le scientifique devient celui qui est reconnu par ses pairs grâce à ses publications dans une revue, il préfigure le chercheur moderne sous la pression du *publish or perish* (« publier ou périr »). Plus encore que l'auteur, le scientifique est à l'origine de la notion moderne de « document ».

Bibliothécaires et documentalistes

La première tentative systématique et ambitieuse de définition du terme « document » est sans doute celle du Belge Paul Otlet (1868-1944). Celui-ci a passé sa vie avec son ami Henri Lafontaine à tenter de réunir, classer et indexer tous les documents du monde, récupérant ce que l'activité du commerce de l'imprimé dispersait. Il donne dans son livre testament *Traité de documentation*, en 1934, la définition suivante : « Livre (Biblion ou Document ou Gramme) est le terme conventionnel employé ici pour exprimer toute espèce de documents. Il comprend non seulement le livre proprement dit, manuscrit ou imprimé, mais les revues, les journaux, les écrits et reproductions graphiques de toute espèce, dessins, gravures, cartes, schémas, diagrammes, photographies, etc. [...] Le Livre ainsi entendu présente un double aspect : a) il est au premier chef une œuvre de l'homme, le résultat de son travail intellectuel ; b) mais, multiplié à de nombreux exemplaires, il se présente aussi comme l'un des multiples objets créés par la civilisation et susceptibles d'agir sur elle¹². » La définition souligne la double multiplication des documents, en types différents et en exemplaires.

Quelques paragraphes plus loin, il s'exclame : « Nous devons former d'immenses bibliothèques, nous devons élaborer des répertoires puissants. Mais, de même qu'après de Jussieu et Linné décrivant des milliers d'espèces sont arrivés les Darwin et Claude Bernard qui ont créé la biologie, science théorique explicative, évolutive de l'ensemble des êtres vivants, de même les temps sont venus maintenant où il faut fonder la bibliologie, la science théorique, comparative, génétique et abstraite, embrassant tous les livres, toutes les espèces et toutes les formes de documents¹³. » La citation montre combien le premier théoricien du document est imprégné des fondements des sciences modernes. Par effet de retour, il souhaite faire profiter son objet des avancées de ces dernières. Le système documentaire, enfant de la science, doit à son tour s'étudier comme une science. Il ne suffit plus de repérer et classer les documents, il faut en comprendre la nature et les lois.

Au cours des années 1950, en France, à la suite de Paul Otlet, Suzanne Briet dans un petit livre passionné élargit et approfondit la réflexion sur le document en proposant la définition suivante : « Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel¹⁴. » La définition s'applique maintenant à tout objet porteur d'information, pour peu qu'il soit documenté. Si Paul Otlet privilégiait la transmission, Suzanne Briet insiste sur la preuve, au sens large de « représentation fidèle ». Les deux fonctions originelles se confondent de plus en plus dans la publication scientifique moderne. L'article scientifique est autant le lieu de l'exposé de la preuve que celui du partage des connaissances. Il deviendra même plus tard l'outil principal de régulation de la communauté scientifique dans une version de la bibliométrie que Paul Otlet n'avait pas envisagée.

Les deux pionniers ont élaboré, dans un mouvement de coopération internationale réunissant bibliothécaires et documentalistes¹⁵, des normes permettant les échanges et facilitant les procédures de travail. On leur doit la mise au point de langages documentaires, ce qu'on appellerait aujourd'hui des métadonnées ou un métalangage, pour retrouver par leur signalement les documents pertinents (ou les

informations pertinentes à l'intérieur des documents) dans une masse de plus en plus écrasante. Le système simplifié de métadonnées homogènes et combinables les unes avec les autres pour accélérer la recherche d'informations préfigurait l'informatique documentaire et plus tard le Web.

Journalistes

Parallèlement à ces efforts soutenus par les communautés scientifiques, le développement des journaux témoigne lui aussi de l'évolution de la notion d'« information ». On emploie peu le terme « document » dans ce monde-là, mais les préoccupations sont proches. Le journal populaire, à bas prix, à grand tirage et financé par la publicité, a été inventé dans la seconde moitié du XIX^e siècle conjointement des deux côtés de l'Atlantique. Progressivement et d'abord aux États-Unis, le contenu des journaux a cherché à s'émanciper des pouvoirs politiques et économiques. Au début du XX^e siècle, on se met à parler de « quatrième pouvoir » pour la presse et d'« objectivité » pour les journalistes. Paul Starr dans son livre *The Creation of The Media* rappelle la naissance de la notion d'objectivité prônée par Walter Lippmann en 1920 dans *Liberty and The News* : « Lippmann exhortait les journalistes à être plus “objectifs”, un mot qui venait d'apparaître pour décrire le journalisme. Les critiques aujourd'hui dénoncent l'objectivité comme une idéologie professionnelle, mais il est important de comprendre les pratiques professionnelles que Lippmann voulait faire adopter aux journalistes. Il voulait que les journalistes s'inspirent de la science en développant un “sens de la preuve” et en reconnaissant franchement les limites de l'information disponible : il les exhortait à démonter les idées reçues et les abstractions et à refuser de laisser de côté des nouvelles ou à mettre la morale ou n'importe quelle autre cause avant la véracité¹⁶. » Même si le terme « document » n'est pas employé, les notions de transmission et de preuve sont bien présentes. Elles sont mises en avant comme issues directement des valeurs scientifiques modernes qui se fondent sur les faits, l'expérience, et non sur une vérité révélée. L'article que le journaliste écrira à la suite de son enquête selon ces critères est bien un document au sens de Paul Otlet et Suzanne Briet.

En montrant que les professionnels des bibliothèques comme les journalistes s'appuient au début du siècle dernier sur la science moderne pour asseoir une conception du document qu'ils pensent porteur de progrès et de liberté, je n'adopte pas personnellement un jugement moral. Je constate simplement l'évolution du discours et des pratiques. Sans doute l'une et l'autre professions s'appuient sur une éthique forte, fondée sur la liberté d'expression, la liberté de conscience, le droit à l'information et à la culture, ce que l'on appelle aujourd'hui sous l'impulsion de Jürgen Habermas l'« espace public ». L'histoire du XX^e siècle, et tout particulièrement celle des totalitarismes, a montré malheureusement de façon dramatique combien les documents et les systèmes documentaires pouvaient être aussi des instruments de propagande et de contrôle social.

Philosophes, historiens, sociologues

Poursuivant les travaux de Michael Buckland¹⁷, qui le premier insista sur l'importance de la notion en rappelant son histoire, Niels Lund a publié récemment une recension des différentes théories sur le document¹⁸. Il y montre que, dans la seconde moitié du XX^e siècle, parallèlement aux efforts des professionnels pour améliorer les systèmes documentaires, un ensemble de penseurs se sont intéressés au document, moins pour le définir, le produire ou l'organiser que pour en comprendre la nature et la fonction. Il fait ressortir ainsi des convergences entre les travaux du sociologue Karl Mannheim, de l'anthropologue Harold Garfinkel, du philosophe Michel Foucault, du sociologue Bruno Latour ou encore des historiens John Seely Brown et Paul Duguid. Chacun à partir de sa discipline, de sa position et de son raisonnement particuliers propose une avancée dans la réflexion sur la notion. Les documents ne sont plus maintenant interprétés pour ce qu'ils disent explicitement, mais comme des traces à mettre en relation avec d'autres qui témoignent d'évolutions sans rapport nécessaire avec leur contenu. Ils sont pris dans un système signifiant plus vaste. Inversement, tout objet peut devenir document à partir du moment où il permettra, en relation avec d'autres, d'interpréter l'époque de sa circulation.

Ces penseurs désacralisent en quelque sorte le document, qui n'est plus cet objet extraordinaire porteur de savoir qui enthousiasmait les professionnels du début du siècle. C'est devenu un objet parmi d'autres, en relation avec d'autres, qu'il faut étudier comme tel, avec ses particularités. En même temps, et de façon un peu paradoxale, ils montrent combien les documents et l'organisation documentaire régulent, contrôlent, organisent les sociétés, participant à leur gouvernance, leur administration, leur évaluation. Ce faisceau de penseurs n'est pas un courant de pensée homogène, mais il marque une étape supplémentaire dans l'évolution de la notion.

Les trois dimensions

Avec le nouveau millénaire, le document désormais numérique n'est plus qu'un ensemble de signaux dans un vaste flux. Chacun, publiant directement sur le Web, se ressent auteur. L'indexation n'est plus l'apanage des documentalistes et les métadonnées sont devenues indispensables à la « navigation » sur le Web. Les journaux sont concurrencés par les témoignages des internautes. Le document est aujourd'hui un ensemble de pièces éparses réagencées à la volée, une tête d'un réseau concrétisé par les liens hypertextuels, et les traces laissées par les internautes sont des outils de gestion du trafic et des stratégies des firmes. Bref, tous les éléments rappelés précédemment dans ce chapitre sont bousculés et la notion de document elle-même est ébranlée.

En France, un réseau de plus d'une centaine de chercheurs, issus de disciplines diverses aussi bien en sciences humaines et sociales qu'en informatique, a mené une réflexion sur le document numérique¹⁹. Pour mieux confronter leurs approches, ils ont rédigé collectivement des textes, signés d'un pseudonyme : Roger T. Pédaque²⁰.

Après avoir recensé les conceptions provenant de disciplines diverses, Pédaque a réalisé une synthèse et proposé une grille tridimensionnelle pour rendre compte des différentes facettes du document – forme, contenu ou texte, et médium –, qu'il nous faut détailler un peu car nous nous y référerons souvent. Les deux premières dimensions, forme et contenu, sont les plus communément admises. La troisième, le médium, renvoie à sa fonction sociale. Cette triade peut être résumée en français de façon mnémotechnique par trois participes passés : vu, lu et su. Selon cette conception, un document n'est efficient que lorsque les trois dimensions sont abouties et cohérentes entre elles. Je reprends à mon compte ci-dessous la proposition en l'adaptant à mon propos.

Vu : forme

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 le sens qu'il véhicule. Un document doit être immédiatement appréhendable comme tel. L'exemple le plus traditionnel est le livre imprimé que l'on repère d'un coup d'œil. Nous sommes capables de dire, même en l'entr'apercevant de loin, « cela est un livre », car nous avons rencontré cette forme depuis l'enfance, elle nous est familière et même souvent elle a été l'outil d'un apprentissage essentiel pour notre sociabilité : la lecture. Un lecteur averti sera capable de repérer la collection ou l'éditeur d'un livre de loin dans une bibliothèque. Sa couverture, son titre, son auteur, lui permettront de le ranger dans tel ou tel genre. L'intérieur du livre, sa structure logique en pages, en chapitres, avec une table des matières, etc., nous sont familiers et nous permettent de le feuilleter sans nous y perdre.

Cette habileté de repérage de documents est essentielle à notre vie quotidienne et n'est, bien sûr, pas exclusive au livre. Vous repérerez sans difficulté un passeport parmi des dizaines d'objets sur une table. Il vous est utile aussi de retrouver facilement factures, contrats, courriers. Ou encore, vous serez capable de pointer sans hésiter votre journal ou magazine favori dans un kiosque qui en présente pourtant des centaines lui ressemblant, rangés dans un ordre approximatif. Inversement, si nous ne sommes pas un minimum familiers avec la présentation du document, s'il ne correspond pas pour nous à un environnement de lecture suffisamment connu pour que nous ne nous y perdions pas, il nous sera très difficile de repérer le document, et celui-ci n'existera tout simplement pas pour nous. Cette première idée est un élargissement à la mise en forme de ce que les théoriciens de la réception²¹ appellent un « horizon d'attente » ou un « contrat de lecture », qu'ils appliquent généralement plutôt à la forme du texte.

Cette dimension de la forme privilégie donc le repérage, le document doit pouvoir être *vu*. Pédaque a résumé cette première dimension par une équation : « document = support + inscription ». Pour les documents imprimés, par définition l'inscription et le support sont solidaires. La solution de l'équation se résume donc à une simple question de perception. Celle-ci se complique avec l'audiovisuel, puisque l'enregistrement et la restitution supposent un appareillage. Le document est alors reconstruit à partir d'un signal enregistré avant de pouvoir être perçu. Reste que la solution de l'équation est encore unique. Le signal enregistré représente le document dans sa totalité et le décodage est une transposition d'une forme à une autre, avec simplement des risques de brouillages. Pour les documents numériques, l'équation se complique encore. À l'appareillage d'enregistrement et de restitution vient s'ajouter la possibilité du traitement et du calcul sur le signal. Les logiciels de gestion et d'affichage des documents, les formats de fichier manipulent la structure logique du texte, ou encore le protocole de partage et le système d'adressage autorisent une réinvention du document. À partir des mêmes éléments enregistrés, plusieurs documents

différents peuvent être restitués. Le contrat de lecture numérique se cherche encore et nombreux sont les travaux à ce sujet qui s'interrogent sur la meilleure forme de présentation selon le type de document et son usage. Pédaque propose pour la première dimension de la version numérique du document cette définition : « Un document numérique est un ensemble de données organisées selon une structure stable associée à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs²². »

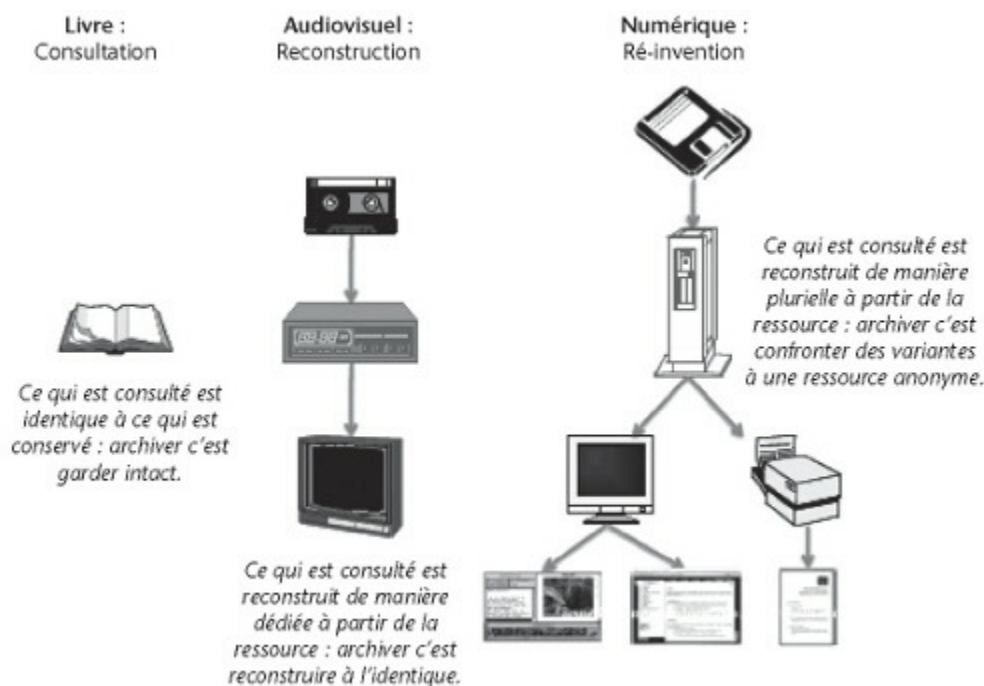


Figure 2. Archiver l'imprimé, l'audiovisuel et le numérique²³

Bruno Bachimont a représenté l'évolution de la relation du document à son support entre imprimé, audiovisuel et numérique sur un schéma illustrant les conséquences sur l'archivistique.

Dans cette dimension, le document n'est qu'une promesse. Sans connaître son contenu, nous savons déjà que c'est un document et nous pressentons qu'il sera pour nous utile, intéressant, distrayant ou au contraire sans attrait. Mais une promesse doit pouvoir être tenue, sans quoi elle n'est rien. On ne saurait donc se contenter de voir un document, il faut aussi pouvoir en consulter le contenu, même si cette potentialité n'est pas mise en œuvre.

Lu : texte

La deuxième dimension, celle du contenu ou du texte, est intellectuelle, ou si l'on veut « neuro-anthropologique²⁴ ». Il s'agit du rapport de notre cerveau et de ses capacités de raisonnement au contenu du document, quelle que soit la façon dont il est représenté, une écriture, une image animée ou non, un son ou encore la combinaison de tous ces éléments. Notre cerveau, par l'intermédiaire de nos sens, décode l'inscription proposée pour pouvoir interpréter la représentation qu'elle propose. Il existe deux étapes pour cette dimension : le déchiffrement des signes et leur interprétation par leur mise en contexte. Cette fois, il ne suffit plus de repérer par la forme, ici le document doit pouvoir être *lu* et compris. Cette dimension met en avant la signification. Le décodage de l'inscription peut être réalisé par nous-mêmes, par une machine ou par les deux. Dans le cas du document le plus traditionnel qui supporte un texte écrit, comme ce livre que vous êtes en train de lire sous une forme imprimée ou numérique, le décodage est réalisé par le lecteur, vous-même en l'occurrence. Il suppose que vous ayez effectué, sans doute dans votre enfance, un long apprentissage pour savoir lire, c'est-à-dire déchiffrer l'inscription faite de lettres et de signes typographiques ordonnés pour accéder à la représentation que je vous propose et que vous allez interpréter selon votre propre contexte.

Comme le repérage précédent, l'interprétation, c'est-à-dire le sens que vous allez donner à ce texte, suppose aussi un contrat de lecture implicite entre le lecteur et l'auteur, ici entre vous et moi. Je dois vous fournir dans mon texte suffisamment d'éléments pour que vous puissiez intégrer la représentation que je vous propose dans votre propre système de connaissances et, inversement, vous devez avoir préalablement intégré suffisamment d'éléments de contexte pour entrer dans le mien. C'est pour cette dimension du texte que les chercheurs ont mis en avant cette idée de contrat de lecture, montrant par exemple qu'elle sous-tend dans la littérature la notion de genre²⁵. Pour qu'un texte appartienne à un genre, il doit suivre certaines règles d'écriture qui, familières au lecteur, lui permettront d'entrer dans l'univers de l'auteur avec plus de facilité.

Cette deuxième dimension se décline aussi selon les types de représentation et on peut reprendre les enseignements de la figure 2. Dans le cas d'une représentation analogique – une photo, du son, de l'image animée, de l'audiovisuel –, l'effort de décodage de votre part sera dans un premier temps moins complexe, car la représentation sera plus « évidente ». *A contrario*, l'audiovisuel suppose l'usage de machines, des bien nommés « lecteurs » qui décodent à la place des humains les inscriptions sur les bandes enregistrées, les films ou les disques. Le décodage par les machines est passif, il s'agit d'une simple transposition d'une inscription « illisible » par l'humain vers une autre qu'il peut appréhender. La suite du processus intellectuel ne diffère guère entre l'audiovisuel et l'écrit. Pour l'un comme pour l'autre, il vous restera à interpréter le message proposé et là encore un contrat de lecture sera indispensable, qui passera par des genres audiovisuels variés et codés. L'informatique, à son tour, a fait un pas supplémentaire en autorisant le calcul, les manipulations logiques, par le code. Les documents numériques peuvent non seulement être décodés par les machines, mais « interprétés », ce qui ouvre la porte à bien des applications inédites pour le traitement automatique de la langue, la statistique ou encore l'ingénierie des connaissances, comme les tentatives de traduction automatique, les moteurs de recherche ou encore les applications du Web sémantique que nous aborderons au chapitre suivant. On trouvera facilement un renseignement par une simple question sur le Web ; on positionnera des objets sur une carte à l'intérieur de laquelle on pourra se déplacer ; on archivera ses photos que l'on pourra même retoucher, etc. Ces facilités ont modifié considérablement cette dimension du contrat de lecture numérique. Aujourd'hui, on feuillette, on navigue, on consulte, on recherche, on copie/colle sur le Web plutôt qu'on ne concentre son attention sur un texte long. Le texte est déconstruit au moins en partie et de nouveaux contrats se mettent progressivement en place, définissant de nouveaux genres : blogs, vidéos courtes, *tweets*, mélanges de contenus audios ou vidéos (*mashups*), etc.

Cette deuxième dimension peut se traduire également par une équation : « document = code + représentation ». Pour le document numérique, Pédaque a proposé la définition suivante : « un texte dont les éléments sont potentiellement analysables par un système de connaissances en vue de son exploitation par un lecteur compétent²⁶ ». Stanislas Dehaene, théoricien de la lecture, suggère implicitement dans la conclusion de son livre que cette dimension n'est toujours pas suffisante pour embrasser l'ensemble du phénomène. Il faut lui adjoindre une troisième dimension, sociale : « Au terme de notre voyage cortical, la lecture apparaît à la fois comme le fruit de l'évolution humaine et comme l'un des acteurs majeurs de son explosion culturelle. L'expansion de notre cortex préfrontal, cette cathédrale de l'esprit, a permis à notre espèce d'inventer l'écriture. À son tour, cette invention a dopé nos capacités mentales en nous dotant d'une véritable mémoire supplémentaire, externe et durable, qui nous permet d'« écouter les morts avec les yeux », comme le souligne le poète Francisco de Quevedo, et de réveiller, à plusieurs siècles de distance, la pensée des écrivains disparus. À ce titre, la lecture est la première « prothèse de l'esprit » – une prothèse que des générations de scribes ont adaptée au mieux aux circuits de nos cerveaux de primates²⁷. » Ce que Stanislas Dehaene dit de l'écriture peut s'extrapoler à l'ensemble des représentations documentaires, qui sont des « prothèses de l'esprit » dont il faut aussi mesurer la dimension sociale.

Su : médium

La troisième dimension est donc celle de la médiation. Quels que soient sa forme et son contenu, le document a une fonction sociale. Nous avons vu que les deux fonctions, transmettre et prouver, ont évolué en fusionnant et en s'élargissant. Aujourd'hui, on dit « informer » sans que l'acception du verbe soit très précise. Malgré tout, la partition initiale reste utile pour raisonner, car les deux termes ne renvoient pas aux mêmes univers. Transmettre se conjugue en espace et en temps, tandis que prouver se conjugue en véracité et en confiance. Je traiterai dans ce chapitre du premier, laissant le second pour le chapitre suivant.

Pour poursuivre l'exemple de la lecture de ce livre, vous-même, comme lecteur, êtes transformé par l'information qui a été mise en mémoire sur les pages imprimées. Ainsi, des informations quittent l'espace intime de l'auteur, moi-même, pour gagner celui du lecteur. Le véhicule de cette transmission est le livre

que vous avez entre les mains ou que vous parcourez sur un écran. Pour qu'il soit parvenu jusqu'à vous, il ne suffit pas qu'il ait été écrit et mis en forme, il faut encore qu'il ait été publié et que vous ayez pu y accéder. Toute une chaîne d'opérations a été suivie, mettant en relation toute une série d'acteurs, ici le système éditorial dans la version traditionnelle ou numérique.

Je vous fais part de mes réflexions que vous assimilez en les lisant. Cela ne signifie pas naturellement que vous les approuviez, mais bien que les représentations que j'ai construites sont parvenues jusqu'à vous. Nous ne nous connaissons vraisemblablement pas et pourtant j'ai pu vous transmettre des informations. Mieux, je construis ces représentations aujourd'hui, le jour où j'écris ces pages, assis dans mon bureau devant un paysage montagnard. Vous les lisez en dehors de ma présence au moment et dans le lieu qui vous conviennent, votre aujourd'hui à vous. Ni les jours ni les lieux ne coïncident. La seule chose que je puis dire ou espérer est que vous lirez effectivement ce livre plus tard, au plus tôt le jour de sa mise à disposition du public. En supposant que ce livre a été conservé dans une bibliothèque j'ai pu même imaginer que vous le lisez après ma mort, sans doute par hasard, car son contenu sera au mieux désuet. Cette dimension souligne la distance qui s'établit entre l'émetteur, dans notre exemple l'auteur du livre, moi, et le récepteur, le lecteur, vous.

L'exemple peut être étendu à toutes sortes de documents. Chaque fois, la caractéristique est la même, il s'agit de transmettre à d'autres, et au-delà de l'ici et du maintenant, un contenu à l'aide d'une forme. Le document, par le système de diffusion qui lui est associé, est donc un médium qui permet de faire passer un texte mis en forme au-delà du cercle intime et de la barrière du temps. Le document est un objet mémoriel : on enregistre une information (texte) sur un objet (forme) pour pouvoir la transmettre ou s'y référer, autrement dit pour qu'elle puisse être *sue* dans un autre lieu et à un autre moment. L'ensemble du système documentaire est en quelque sorte notre mémoire externe, pour paraphraser Stanislas Dehaene ou Michel Serres²⁸.

Pour fonctionner, ce système devra mettre en place un ensemble contractuel pour que les maillons de la chaîne s'articulent correctement et que les intérêts de chacun puissent être préservés. Le terme « contrat » est employé ici dans un sens moins métaphorique ou implicite que pour les contrats de lecture des deux dimensions précédentes. Il a des applications juridiques et suppose des transactions financières, comme un règlement de facture ou un paiement de droits, ou des obligations, comme le dépôt légal ou la propriété intellectuelle.

Le rapport au passé différencie le document du spectacle, qui se produit à un moment et dans un lieu où il s'écoute ou se regarde, et de la conversation en présence des interlocuteurs. Le spectacle et la conversation sont des performances. Le direct télévisé et radiophonique ou le téléphone ont permis de briser la barrière de l'espace, mais pour passer celle du temps il a fallu avoir recours à l'enregistrement, c'est-à-dire transformer ces performances en document. Nous verrons que, ce faisant, le document a rejoint son acception ancienne.

Une fois de plus, le numérique, tout particulièrement le Web, modifie considérablement le rapport au temps et à l'espace. La séparation entre la performance et le document, entre la communication et l'information, entre la conversation et la publication s'estompe. On enregistre, discute et publie d'un même clic. Dès lors, l'économie de l'ensemble est bouleversée et les négociations et polémiques ne manquent pas entre les acteurs de la publication traditionnelle et les nouveaux venus sur le Web, depuis des internautes autonomes et entreprenants jusqu'à des firmes commerciales à la croissance impressionnante.

Intégration

Chacune de ces dimensions a sa propre logique qui ne se réduit pas aux autres et pourtant aucune n'est complètement indépendante des deux autres. Un document doit intégrer et coordonner les trois. Ses modalités anthropologiques, intellectuelles et sociales doivent non seulement être efficaces prises chacune séparément, mais encore être cohérentes entre elles.

Si l'on reprend l'exemple du livre, on peut représenter l'aspect tridimensionnel sur un schéma triangulaire. Chaque sommet illustre une dimension : le livre comme une forme, en l'occurrence un codex qui nous est familier, comme un texte écrit sur des pages que l'on peut déchiffrer et comprendre et enfin un livre en train d'être lu, c'est-à-dire assimilé par un lecteur au lieu et moment souhaités par lui. Le livre n'existe en réalité qu'à la confluence des trois, il faut qu'il puisse être trouvé, déchiffré et assimilé pour exister. Bien entendu, le livre est ici un archétype que l'on peut remplacer par toutes sortes de documents.

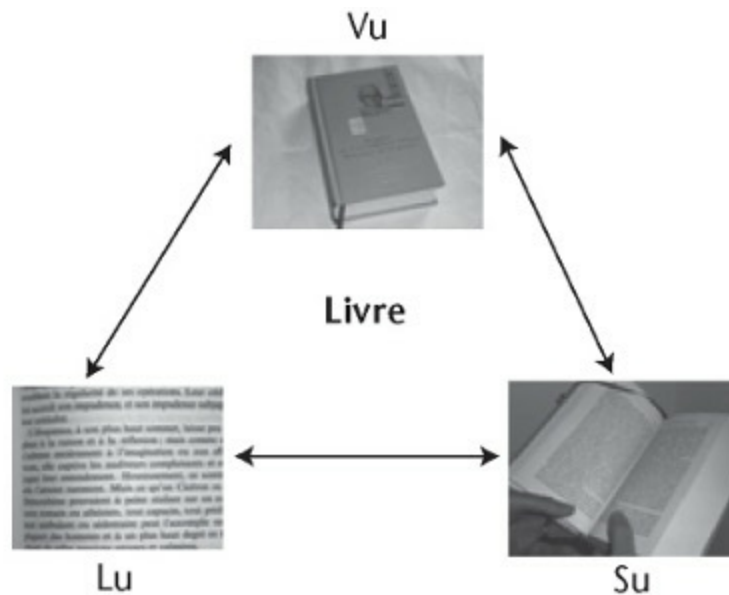


Figure 3. Les trois dimensions du livre

Pédauque, qui a suggéré cette partition en trois dimensions, conclut ainsi sa proposition : « Notons enfin que dans chaque entrée nous avons insisté sur l'idée de contrat de lecture traduit par la lisibilité dans la première, par la compréhension dans la seconde et par la sociabilité pour la troisième. Il est probable que ce contrat à trois facettes présente, dans toutes les nuances que nous avons exposées, la réalité de la notion de document. Un document ne serait finalement qu'un contrat entre des hommes dont les qualités anthropologiques (lisibilité – perception), intellectuelles (compréhension – assimilation) et sociales (sociabilité – intégration) fonderaient une part de leur humanité, de leur capacité à vivre ensemble. Dans cette perspective, le numérique n'est qu'une modalité de multiplication et d'évolution de ces contrats²⁹. »

Définition du document

Cette notion de contrat est intéressante, car elle insiste sur la construction du document qui permet de relier les hommes à partir de caractéristiques matérielles, intellectuelles et mémorielles négociées. Mais elle a un défaut : elle ne dit rien de l'objet du contrat.

C'est pourquoi, inspirée par Bruno Bachimont³⁰, je proposerai cette définition plus engageante : *un document est une trace permettant d'interpréter un événement passé à partir d'un contrat de lecture*. Nous retrouvons bien les trois dimensions, matérielle avec la trace (vu), intellectuelle avec l'interprétation (lu), mémorielle avec l'événement passé (su), ainsi que la nécessaire construction sociale avec le contrat. La notion de trace permet d'élargir la définition du document à toutes sortes d'objets comme l'avait proposé Suzanne Briet. Une étoile dans le ciel, une antilope, pour reprendre ses exemples, peuvent être des documents pourvu qu'elles soient les témoins d'un savoir inscrit dans un système documentaire. Ainsi les documents sont très divers, depuis les contrats, factures, bulletins de paie, bordereaux, circulaires, lois et règlements, cartes d'identité, permis variés, jusqu'aux romans, albums, films, photos d'actualité, en passant par les pièces archéologiques, données scientifiques, articles de revues, objets muséaux et aussi les cartes de visite, faire-parts, petites annonces, affiches publicitaires et l'on pourrait naturellement prolonger indéfiniment cette liste à la Prévert. L'interprétation de chacun passe par le régime documentaire auquel est rattachée la trace. Celui-ci peut être très varié : juridique, fictionnel, scientifique, coutumier, amical, etc. Pour interpréter un document correctement, il est nécessaire d'avoir assimilé les clés d'entrée de ces régimes. Enfin, la référence à un événement passé ne signifie pas nécessairement la relation de cet événement, mais bien que le document fait le lien entre quelque chose qui s'est déroulé dans le passé dont il est la trace et aujourd'hui. La trace peut être fortuite, un indice, ou construite, un texte. Le document est une façon de retrouver notre passé et, nécessairement, de le reconstruire en fonction de notre présent pour orienter notre futur. Ce livre que vous avez entre les mains est un document, il est la trace de l'analyse que j'ai construite à un moment donné et que vous réinterprétez par rapport à votre présent, à partir de nos habitudes partagées de l'écriture-lecture d'un essai.

Mais cette première définition ne permet pas encore de rendre compte d'une qualité essentielle du document ordinaire qui autorise sa mise en système : sa reproductibilité, sa plasticité, son traitement. Sans doute n'importe quel objet peut devenir un document, mais il reste alors unique. C'est en quelque sorte un prototype documentaire, disons un protodocument. Le document ordinaire est un texte, une représentation formelle de ce prototype sur un support maniable. Si le protodocument est déjà d'ordre textuel, il pourra être directement la matrice du document, comme dans le cas de la copie des scribes ou de l'imprimé. Dans le cas contraire, des textes viendront documenter le protodocument, jusqu'à parfois le remplacer, depuis les notices jusqu'aux enregistrements analogiques.

La définition du célèbre bibliothécaire indien Shiyali Ramamrita Ranganathan insiste, elle, sur ses qualités manipulatoires : « Un document est une micro-pensée enregistrée [*embodied micro thought*] sur papier ou sur un autre support, qui permet une manipulation physique facile, un transport dans l'espace et une préservation dans le temps³¹. » Mais cette définition a le défaut d'oublier les protodocuments et ne rend pas compte de la valeur sociale du processus documentaire. Nous pouvons alors articuler les deux définitions : *un protodocument est une trace permettant d'interpréter un événement passé à partir d'un contrat de lecture. Un document est la représentation d'un protodocument sur un support, pour une manipulation physique facile, un transport dans l'espace et une préservation dans le temps.*

Si l'on poursuit le raisonnement, la multiplication des documents et des genres dans toutes sortes de registres et leur transformation que nous avons décrites dans ce chapitre témoignent d'une relation fiévreuse à notre passé, une sorte d'interrogation existentielle sur notre présent face à un futur angoissant dont les termes se renouvellent sous nos yeux. Mieux ou pis, le numérique, par ses capacités calculatoires, permet de reconstruire des documents à la demande et nous donne l'illusion d'avoir toutes les réponses à nos questions avant même qu'elles ne soient posées, comme si notre futur était un destin déjà inscrit dans les machines. Nous allons en suggérer quelques interprétations dans les chapitres suivants.

Notes du chapitre 2

1. Thomas PETITT, *Before the Gutenberg Parenthesis : Elizabethan-American Compatibilities*, document non daté, disponible sur le site du Massachusetts Institute of Technology : <<http://Web.mit.edu>>.

2. « (Socrate) : Ce qu'il y a même en effet, sans doute, de terrible dans l'écriture, c'est, Phèdre, sa ressemblance avec la peinture : les rejets de celle-ci ne se présentent-ils pas comme des êtres vivants, mais ne se taisent-ils pas majestueusement quand on les interroge ? Il en est de même aussi pour les discours écrits : on croirait que ce qu'ils disent, ils y pensent ; mais, si on les interroge sur tel point de ce qu'ils disent, avec l'intention de s'instruire, c'est une chose unique qu'ils donnent à comprendre, une seule, toujours la même ! » PLATON, *Phèdre*, 275 (d).

3. Mary CARRUTHERS, *Le Livre de la mémoire. Une étude de la mémoire dans la culture médiévale* (trad. de l'anglais par Diane Meur), Macula, Paris, 2002, p. 51.

4. On en trouvera d'autres exemples en Inde, dans le monde romain ou pour la Torah dans le deuxième tome de Christian JACOB, *Lieux de savoir*, op. cit., p. 675-737.

5. Traduction personnelle.

6. Michael T. CLANCHY, *From Memory to Written Record : England 1066-1307*, Blackwell Publishers Ltd, Oxford, 1999 (6^e éd.), p. 2 (traduction personnelle).

7. Interrogation réalisée sur la base N-Gram le 11 janvier 2011. <<http://ngrams.googlelabs.com/>>.

8. Interrogation réalisée le 10 janvier 2011.

9. De ce point de vue, l'entreprise collective menée par Christian Jacob sur les « lieux de savoir » est exemplaire, même si elle n'aborde pas frontalement la notion de « document ». Cf. Christian JACOB, *Lieux de savoir*, op. cit.

10. Lucien FEBVRE et Henri-Jean MARTIN, *L'Apparition du livre*, Albin Michel, Paris, 1958 (rééd. 1999) ; Elizabeth EISENSTEIN, *La Révolution de l'imprimé dans l'Europe des premiers temps modernes*, La Découverte, Paris, 1991 ; Adrian JOHNS, *The Nature of the Book : Print and Knowledge in the Making*, Johns Chicago University Press, Chicago, 1998 (rééd. 2008).

11. Jean-Claude GUÉDON, « Le Créateur, le génie et l'éditeur », préface à la traduction française de Lawrence LESSIG, *L'Avenir des idées, Le sort des biens communs à l'heure des réseaux numériques*, Presses universitaires de Lyon, Lyon, 2005.

12. Paul OTLET, *Traité de documentation : le livre sur le livre. Théorie et pratique*, Editiones Mundaneum, Bruxelles, 1934, p. 9.

13. *Ibid.*, p. 9.

14. Suzanne BRIET, *Qu'est-ce que la documentation ?*, EDIT (Éditions documentaires industrielles et techniques), Paris, 1951, p. 7.

15. Pour une histoire de la coopération bibliographique, voir : Sylvie FAYET-SCRIBE et Cyril CANET, « Histoire de la normalisation autour du livre et du document : l'exemple de la notice bibliographique et catalographique », *Solaris*, décembre 1999-janvier 2000. <<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris>>.

16. Paul STARR, *The Creation of the Media : Political Origins of Modern Communications*, Basic Books, Cambridge (MA), États-Unis, 2004, p. 396-397 (traduction personnelle).

17. Michael BUCKLAND, « What is a "document" ? », *Journal of the American Society of Information Science* 48, n° 9, septembre 1997, p. 804-809.

18. Niels Windfeld LUND et Roswitha SKARE, « Document theory », *Encyclopedia of Library and Information Sciences*, Taylor and Francis, New York, 2010 (3^e édition), p. 1632-1639.

19. Réseau thématique prioritaire document et contenu (RTP-doc) du CNRS 2002-2005.

20. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, op. cit. Les textes sont par ailleurs accessibles sur <archivesic.ccsd.cnrs.fr>.

21. Hans Robert JAUSS, *Pour une esthétique de la réception*, Gallimard, Paris, 1978.

22. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, op. cit., p. 45.

23. D'après Bruno BACHIMONT, Support du cours « Archivistique audiovisuelle et numérique », École de bibliothéconomie et des sciences de l'information, Montréal (Canada), juillet 2011.

24. Stanislas DEHAENE, *Les Neurones de la lecture*, Odile Jacob, Paris, 2007.

- ²⁵. Pour une vue d'ensemble, voir : Karl CANVAT, « Pragmatique de la lecture : le cadrage générique », *Atelier de théorie littéraire : Genres et pragmatique de la lecture*, 2007, sur <fabula.org>.
- ²⁶. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, *op. cit.*, p. 59.
- ²⁷. Stanislas DEHAENE, *Les Neurones de la lecture*, *op. cit.*, p. 421.
- ²⁸. Michel SERRES, *Les Nouvelles Technologies : révolution culturelle et cognitive*, conférence INRIA (Institut public de recherche en sciences du numérique), 11 décembre 2007, disponible sur <<http://interstices.info>>, rubrique « Débattre ».
- ²⁹. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, *op. cit.*, p. 78.
- ³⁰. Bruno BACHIMONT, Support du cours « Archivistique audiovisuelle et numérique », *loc. cit.*
- ³¹. Cité par Michael BUCKLAND, « What is a "document" ? », *loc. cit.* (traduction personnelle).

Réingénieries documentaires

L'histoire de l'Internet et du Web, pourtant bien courte, a fait déjà l'objet de nombreuses relations. Je me propose d'en présenter ici une brève interprétation documentaire. L'épisode numérique contemporain n'est pas, en effet, sans ressemblance avec les efforts des bibliothécaires de la fin du ^{xix}^e siècle. À cette époque comme aujourd'hui, les procédures se sont rationalisées pour réaliser un rêve aussi vieux que les bibliothèques elles-mêmes : récolter, classer et donner accès à tous les documents du monde. Le réseau numérique a sans doute radicalement transformé nos systèmes documentaires et leur périmètre, mais il s'est inspiré des modèles anciens, sans toujours le savoir et même parfois en croyant s'y opposer. Ce renouveau a ouvert un chantier de réingénierie documentaire infiniment plus vaste que le précédent. Cette fois, il ne concerne plus seulement un nombre limité de professionnels spécialisés et d'institutions à but non lucratif bien identifiées, mais des populations entières d'internautes et un très grand nombre d'institutions, dont quelques-unes y trouveront l'opportunité d'un commerce très lucratif. Si l'échelle et les acteurs ont changé, l'objectif est toujours de retrouver la bonne information et de maintenir un ordre documentaire face à l'abondance et au chaos, et nous pouvons tirer quelques leçons du rapprochement et des divergences des deux épisodes. Dans ce chapitre je m'en tiendrai à l'histoire du Web du côté des ingénieurs, réservant l'observation des industriels pour le chapitre suivant.

Renouveau

Nous avons vu que le ^{xix}^e siècle s'est conclu par la mise en place d'un ordre documentaire systématique. Le numérique est l'occasion de le renouveler. J'emploie à dessein « renouvellement » plutôt que « rupture », plus souvent choisi pour caractériser l'arrivée du Web, car, si la Toile a bien constitué une innovation radicale, ses promoteurs ont suivi peu ou prou un raisonnement comparable à celui des pionniers de la rationalisation de la bibliothéconomie.

L'ordre documentaire

Le premier ordre documentaire s'est construit d'abord en Amérique du Nord à partir du schéma de classification encyclopédique conçu par Melvil Dewey en 1876 et de l'énoncé des principes de l'organisation du catalogue de la bibliothèque par Charles Ami Cutter en 1889. L'objectif était d'identifier et de repérer des livres et des revues de plus en plus nombreux dans des collections de bibliothèques, qui elles-mêmes se multipliaient. On organisait donc la bibliothèque tout en facilitant l'accès des documents au lecteur. Dans le même temps et toujours sous l'impulsion de Melvil Dewey, les premières écoles de bibliothécaires se sont ouvertes en Amérique du Nord, dont les principes enseignés alors sont toujours opérationnels dans les bibliothèques d'aujourd'hui.

La première théorisation de cet ordre documentaire, qui fut en même temps une tentative pour son application universelle, est due à un Européen, Paul Otlet, déjà présenté au chapitre précédent. Celui-ci rêvait de classer le monde au sein d'une institution coopérative, le Mundaneum¹.

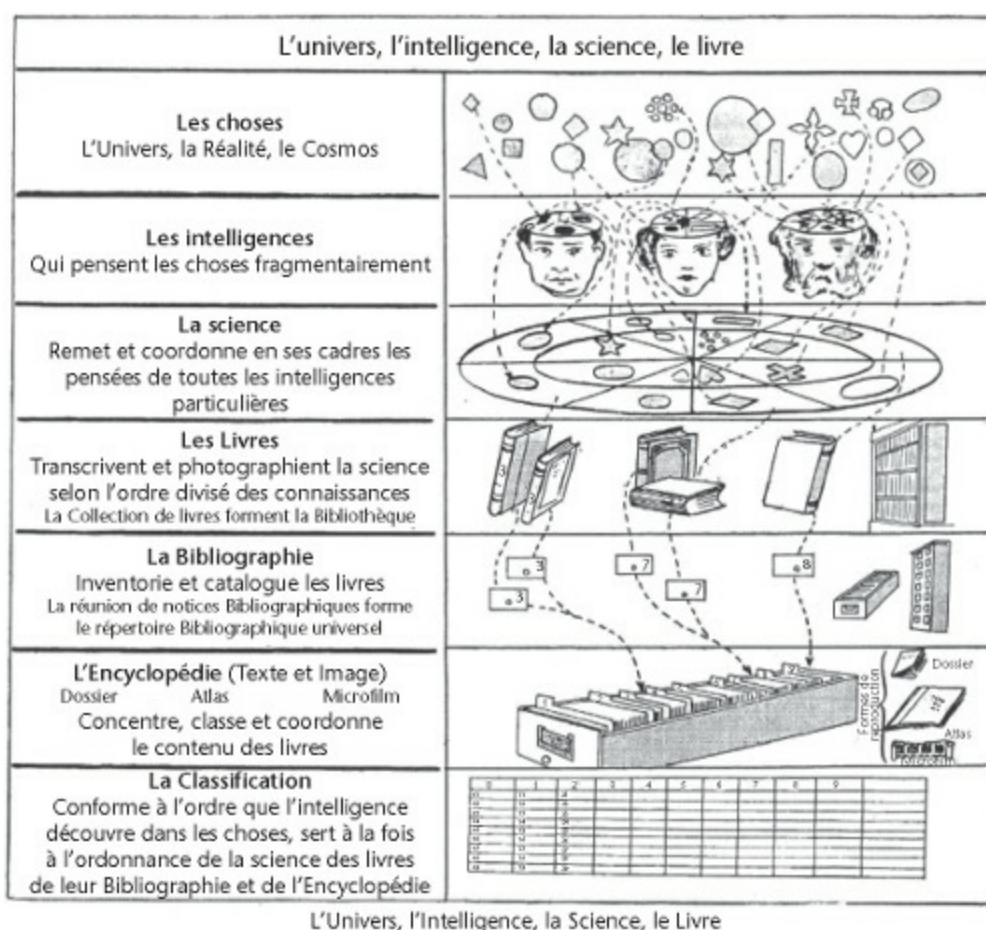


Figure 4. L'organisation documentaire selon Paul Otlet²

Ce tableau, issu de son livre testament de 1934, résume le raisonnement général. Les quatre premières cases en partant du haut figurent la construction des documents. Partant de l'univers, des représentations sont élaborées par les intelligences humaines particulières qui se confrontent, se précisent et se structurent grâce à la science. Le résultat est consigné dans des objets, les livres, eux-mêmes réunis dans des bibliothèques. Les trois lignes suivantes présentent les principaux éléments de l'ordre documentaire nouveau. Il s'agit de notices bibliographiques réunies dans un répertoire bibliographique universel. Une notice bibliographique est une description précise, rigoureuse et formelle des éléments permettant l'identification d'un document, y compris des mots clés, résumant le contenu sur une fiche cartonnée normalisée. L'ensemble de ces fiches normées, toutes de même taille, stockées dans des meubles à tiroirs identiques, a constitué le catalogue de toutes les bibliothèques importantes jusqu'à l'arrivée de l'informatisation dans les années 1970. La notice bibliographique est un substitut du document qui le remplace avantageusement dans le système documentaire, car on peut la manipuler facilement et logiquement du fait de sa taille, de son homogénéité et de son formalisme. Aujourd'hui nous dirions qu'il s'agit de ses métadonnées. Le système documentaire est piloté par ces catalogues normalisés, qui peuvent échanger entre eux. Pour Paul Otlet, il devait donc même être centralisé dans un répertoire universel. Les bibliothécaires ont œuvré et continuent d'œuvrer à ce répertoire aujourd'hui informatisé, de façon centralisée ou distribuée. Le principe de son élaboration est d'éviter dans la mesure du possible les doublons, donc du travail inutile, et de coopérer pour additionner travaux et compétences et faciliter le réseautage.

Paul Otlet suggère un instrument supplémentaire dans son tableau, l'Encyclopédie, qui n'aura pas vraiment de succès à l'époque, mais qui mérite d'être rappelée, car elle préfigure le Wikipédia d'aujourd'hui. L'Encyclopédie est en effet composée d'une série de dossiers de synthèse sur tous les sujets constituant le savoir humain, réalisés et actualisés par les documentalistes à partir des documents existants et diffusables à la demande. Reste la dernière case, essentielle à l'ordre documentaire : la classification. La classification décimale universelle de Paul Otlet s'inspire de celle de Melvil Dewey. Comme elle, son principe est de

regrouper les documents les plus semblables par leur contenu dans des classes et sous-classes numérotées. Il s'agit d'un outil que nous qualifierions aujourd'hui de sémantique, puisqu'il s'appuie sur le contenu du document, et qu'il permet de définir un ordre, régi par un système décimal, un code, pour permettre d'organiser aussi bien les répertoires que les collections. La classification joue pour Paul Otlet le rôle central, organisant et reliant l'ensemble des instruments. L'Encyclopédie et la classification soulignent que les organisations documentaires sont pilotées à partir du début du xx^e siècle sur la base de l'analyse de la deuxième dimension du document, le contenu, le lu.

La volonté de Paul Otlet de construire une science nouvelle, une bibliologie ou même une documentologie, peut nous paraître désuète et naïve aujourd'hui. Mais elle l'a conduit à la première analyse fine du domaine, intégrant les différents supports, y compris microfilms et audiovisuels, et les différents modes de diffusion. C'est ainsi que, dans la conclusion de son livre, il aura cette vision prophétique du système documentaire à venir : « Ici, la table de travail ne serait plus chargée d'aucun livre. À leur place se dresse un écran et à portée un téléphone. Là-bas au loin, dans un édifice immense, sont tous les livres et tous les renseignements... De là, on fait apparaître sur l'écran la page à lire pour connaître la réponse aux questions posées par téléphone, avec ou sans fil. [...] Utopie aujourd'hui, parce qu'elle n'existe encore nulle part, mais elle pourrait bien devenir la réalité pourvu que se perfectionnent encore nos méthodes et notre instrumentation³. » Rappelons que nous sommes en 1934, les stations de radiodiffusion n'ont qu'une quinzaine d'années, la BBC⁴ vient tout juste de lancer les premiers programmes de télévision, les mots « ordinateur » et « informatique » n'existent pas.

Même si cette vision s'approche de ce que sera le Web quelque soixante années plus tard, Paul Otlet reste dans un raisonnement documentaire issu du xix^e siècle, marqué par le modèle hiérarchique industriel de la division du travail. Ainsi, constatant que le développement des sciences interdit désormais à un savant de dominer l'ensemble du savoir, il ajoute quelques pages plus loin : « De grands moyens sont devenus nécessaires et on doit noter les suivants : 1) la division plus grande du travail ; 2) le travail en coopération ; 3) l'établissement de centres d'informations spécialisés où l'on aura le droit de s'adresser pour toutes les questions spéciales ; 4) la systématisation ou synthèse qui remplace des millions de détails par quelques centaines de lois ou propositions générales ; 5) la mathématique qui fournit avec ses formules des moyens de condensation puissants ; 6) la visualisation par le développement des moyens instructifs de représentation et notamment schématique ; 7) le développement des machines intellectuelles ; 8) le livre irradié fait pour la lecture pour tous soit par la lecture individuelle et l'audition d'un livre désiré, soit par la demande radiophonique de renseignements individuels ; 9) la télévision, le livre, le document que sur demande on présentera à la lecture sur le téléviseur, soit pour tous, soit pour chacun⁵. » Nombre des étapes de son programme seront scrupuleusement appliquées par les bibliothécaires et documentalistes au xx^e siècle et la classification, qui est pour lui la « Clé de voûte de la Pensée et du Document⁶ », va grandement faciliter l'informatisation des systèmes documentaires. Les fiches des catalogues de bibliothèques, classées et codées, constituent des bases de données avant la lettre. Les premiers ordinateurs entrent dans les bibliothèques dans les années 1950 et le premier format bibliographique à destination d'un système informatique, le format MARC (*Machine Readable Cataloguing*) a été développé par la Bibliothèque du Congrès, à Washington, au début des années 1960. Suivront rapidement les systèmes intégrés de gestion des bibliothèques (SIGB) pour piloter et coordonner l'ensemble des opérations, depuis les acquisitions de documents jusqu'à leur prêt, en passant par leur traitement.

Après la Seconde Guerre mondiale, l'enjeu des débats internationaux s'est déplacé. De grands laboratoires ont été fondés, certaines entreprises ont développé une activité de recherche. Tous ont constitué pour leurs besoins propres des bibliothèques spécialisées, ou centres de documentation, comme on préfère les nommer en France. Au début des années 1960, les États-Unis, concurrencés par l'Union soviétique dans la course à l'espace, vont mettre en place une organisation nationale de la circulation de l'information scientifique et technique. Face au développement rapide de l'activité des chercheurs et de leur besoin d'échanger et de publier, les revues et les congrès scientifiques se sont multipliés. Les articles et communications en représentent la valeur documentaire réelle, mais ils sont inclus dans des volumes plus vastes, revues et actes. Le catalogage du volume ne renseigne plus vraiment sur son contenu, d'autres outils documentaires sont alors nécessaires. L'objectif de normalisation et de coopération est toujours le même, mais il privilégie maintenant ce qu'on appelle les « produits documentaires ». L'accent est mis sur les revues signalétiques ou analytiques, la littérature dite « secondaire » qui recense et résume les publications dans un domaine et permet ainsi de repérer et de retrouver les articles scientifiques, soulignant la prééminence du traitement du contenu dans les systèmes documentaires. L'informatisation transforme rapidement ces produits en bases de données bibliographiques. La première base de données bibliographiques publique expérimentale accessible en ligne date de 1964, le premier serveur est ouvert l'année suivante⁷. Même si le traitement automatique de l'information a fait lui-même de considérables progrès, la technicité des

normalisations et la rusticité des premiers outils rendent l'interrogation des bases de données complexe au point de nécessiter un savoir spécialisé. Tout un savoir sophistiqué de recherche documentaire a ainsi été constitué et les sciences de l'information, qui émergent sous ce nom après la Seconde Guerre mondiale, sont issues de ces changements.

Le document comme tête de réseau

On considère généralement que Vannevar Bush a posé, juste après la guerre, les premières bases de ce qui deviendra le Web dans un article célèbre, « As we may think⁸ ». Il y propose, en effet, un système documentaire, le *Memex*, constitué d'une base de microfilms donnant accès directement aux documents sur un bureau. Le *Memex* ressemble beaucoup au rêve de Paul Otlet, pourtant sa logique de repérage se veut résolument différente : « Le cœur du problème de la sélection va cependant plus loin qu'un retard dans l'adoption de mécanismes par les bibliothèques ou un manque de développement d'outils pour leur utilisation. Notre inaptitude à obtenir un document découle largement de l'artificialité du système d'indexation. [...] L'esprit humain ne fonctionne pas de cette façon. Il fonctionne par association. À partir d'un élément récolté, il passe immédiatement au suivant qui est suggéré par association d'idées, en conformité avec un réseau de chemins intriqués les uns dans les autres reliant les cellules du cerveau⁹. » En réalité, la proposition de Vannevar Bush reste métaphorique. Pour permettre de retrouver des documents microfilmés par association, ainsi qu'il l'imaginait, une indexation est nécessaire, une « indexation à facettes¹⁰ », mais une indexation tout de même. Ironiquement, malgré quelques tentatives, le *Memex* ne put jamais être mis au point, faute notamment d'une telle indexation¹¹.

L'exemple le plus illustratif de la convergence entre ces réflexions et les efforts documentaires de l'après-guerre est peut-être l'aventure d'Eugene Garfield, qui commence au début des années 1950 et prend toute son ampleur en 1960 par la fondation de l'Institute of Scientific Information (ISI). À la demande de l'industrie pharmaceutique, Eugene Garfield réalise d'abord une revue de sommaires, qu'il élargit progressivement aux principales revues scientifiques, les *Current Contents*. Mais surtout, s'appuyant sur les premières études de scientométrie des sociologues de la science, il construit un index des citations dans les articles scientifiques qui lui permettra de démontrer que les articles les plus cités se concentrent dans un petit nombre de revues. L'ISI classe alors les revues selon un « facteur d'impact » qui servira d'indicateur et de régulateur de l'économie des sciences de la nature. L'histoire d'Eugene Garfield est pour nous pleine de leçons. Elle démarre par un produit documentaire classique (*Current Contents*), se poursuit sur la première application à grande échelle de l'intuition de Vannevar Bush (les liens entre les documents par les associations d'idées des chercheurs, c'est-à-dire les citations), débouche sur une régulation de la circulation des articles scientifiques qui préfigure celle des moteurs de recherche sur le Web (le *Pagerank* de Google compte le nombre de liens pointant vers un document et s'inspire directement du facteur d'impact) et devient enfin l'outil principal et controversé d'une économie internationale de la science académique, par des classements des chercheurs et des universités. Nous verrons que le Web permet, lui aussi, aujourd'hui d'orienter les décisions des internautes par le suivi de leurs traces.

La véritable bascule de l'ordre documentaire découle des liens directs entre les éléments contenus dans les documents, sans passer par une classification ou un index. Depuis toujours, citations ou notes renvoyaient d'un texte à un autre, mais il était pour le moins laborieux de quitter le texte lu pour aborder le texte cité, faute d'avoir l'ensemble des volumes sous la main. Le numérique, qui s'installe progressivement dans la production des textes eux-mêmes à partir des années 1970, va changer la donne. La possibilité de poser des ancres sur un texte numérique permettant de basculer directement à un autre, l'hypertexte, a été théorisée dès les années 1960 par Ted Nelson. Dans cette conception, le document n'est plus seulement un objet, mais aussi un nœud ou une tête de réseau qui permettent de passer d'un document à un autre.

Les bibliothécaires, à l'origine de l'ordre documentaire précédent, n'ont pas été associés à cette bascule. On pourrait même presque dire, comme l'a montré la citation de Vannevar Bush, qu'elle a été réalisée contre eux, et pourtant elle peut être aussi comprise comme la suite logique de leur histoire.

Le Web

En 1989, Tim Berners-Lee propose à l'organisation européenne pour la recherche nucléaire, le CERN, de construire un *mesh* (filet) qui deviendra vite un *web* (toile). Au départ, le Web, selon la conception de son inventeur, reliait des documents entre eux à partir de l'activité des internautes. C'était un vaste système documentaire autogéré. « L'objectif serait de donner une place à chaque information ou référence que l'on pense importante pour qu'elle puisse être retrouvée, et un moyen pour la retrouver ensuite. Le résultat devrait être suffisamment attrayant pour que l'information déposée augmente une fois passé un seuil

critique, de sorte que l'utilité du système encouragera à son tour un usage plus fort », indique Tim Berners-Lee dans la conclusion du premier papier qui décrit son projet¹². Pour que le projet soit opérationnel, il fallait que les différents documents soient potentiellement reliables et appartiennent à un même réseau. C'est une version simplifiée de l'hypertexte par le protocole de transfert HTTP¹³ qui permettra la connexion directe et à grande échelle entre les fichiers de texte des ordinateurs sous la forme des liens que nous connaissons et qui sera à la base du Web, avec la fixation des adresses (URI¹⁴) et la compatibilité des formats (HTML¹⁵). Les documents sont gérés comme des objets que l'on peut retrouver grâce à leur adresse et auxquels on peut accéder par des liens. Les premiers standards du Web reflètent cette orientation. Le génie de Tim Berners-Lee est d'avoir su convaincre la communauté scientifique et les entreprises concernées d'embarquer dans cette normalisation de fait en montant un lieu virtuel de négociations continues sur les protocoles informatiques, le consortium W3C, en 1994. Le succès a été immédiat.

En 2001, le même Tim Berners-Lee, dans un article célèbre¹⁶ rédigé avec deux collègues, 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 ». Les discussions et polémiques autour des développements du Web sémantique et son passage au Web des données dépassent l'approche qui nous intéresse ici, mais l'orientation du projet marque bien un pas supplémentaire vers la conception d'un système documentaire global et peut-être vers la conception d'un nouveau type de document. « Cette philosophie ressemble à celle du Web classique : dès le développement du Web, ses détracteurs ont souligné qu'il ne pourrait jamais être une bibliothèque bien organisée, que sans base de données centrale et sans structure arborescente on ne pourrait jamais être sûr de tout trouver. Ils avaient raison. Mais la puissance d'expression du système a mis à la disposition du public des quantités importantes d'information et les moteurs de recherche (qui auraient paru tout à fait irréalisables il y a dix ans) permettent de trouver des ressources. Le défi du Web sémantique, cependant, est de fournir un langage qui exprime à la fois des données et des règles pour raisonner sur les données et pour que les règles de n'importe quel système de représentation de la connaissance puissent être exportées sur le Web¹⁷ », écrivent les promoteurs du Web sémantique. Cette citation est pour nous pleine d'enseignements. Tout d'abord, elle souligne les limites documentaires du Web classique en le comparant à l'organisation d'une bibliothèque, pour indiquer immédiatement que les moteurs de recherche, dont la dynamique rapidement commerciale s'est développée parallèlement au Web, auraient résolu le problème de la recherche d'information. En résumé, l'autogestion du Web et les moteurs remplaceraient les services d'une bibliothèque. Ensuite, les auteurs proposent d'aller plus loin en intervenant sur les textes eux-mêmes, à partir d'unités documentaires de base, les « données », qui pourraient se réagencer à la demande pour répondre aux questions posées aussi bien par des internautes que par des machines. En simplifiant, on pourrait dire que le Web sémantique est pour ses promoteurs une bibliothèque de données (ou d'unités de contenu) permettant de reconstruire tous les documents possibles à partir de liens intelligents grâce aux capacités de calcul des machines interconnectées, inexploitées dans le Web classique.

La présentation initiale, un peu effrayante dans ses conséquences sur la signification reconstruite par les machines, a été depuis nuancée. On parle moins de Web sémantique et plus de « Web des données » et Tim Berners-Lee exhorte aujourd'hui¹⁸ les institutions à rendre directement accessibles sur le réseau les « données brutes », comme les internautes ont rendu accessibles les documents sur le Web classique, afin qu'elles puissent servir de sources aux calculs des ordinateurs pour pouvoir notamment construire à la volée de nouveaux documents.

L'inversion du flux

Les différentes couches informatiques qui forment le Web sémantique ont été présentées sur un schéma, baptisé le « *cake* » car il les empile les unes sur les autres, comme un gâteau, sorte de feuille de route pour l'avancement et l'agencement des différents niveaux de langages et protocoles en fonction des tâches à effectuer. La dernière version du schéma est reproduite ci-dessous. L'organisation en tâches successives sous forme d'un empilement n'est pas sans rappeler l'organisation documentaire présentée par Paul Otlet, pourtant nous ne sommes plus dans une division humaine du travail, dans une manufacture ou une usine, mais dans une robotisation du processus. L'ordinateur a remplacé l'homme, ce qui rend la lecture du schéma moins immédiate. Il ne s'agit pas d'expliquer à des humains le travail à effectuer, mais de se mettre d'accord sur les langages nécessaires pour piloter des calculateurs. Nous avons changé de siècle.

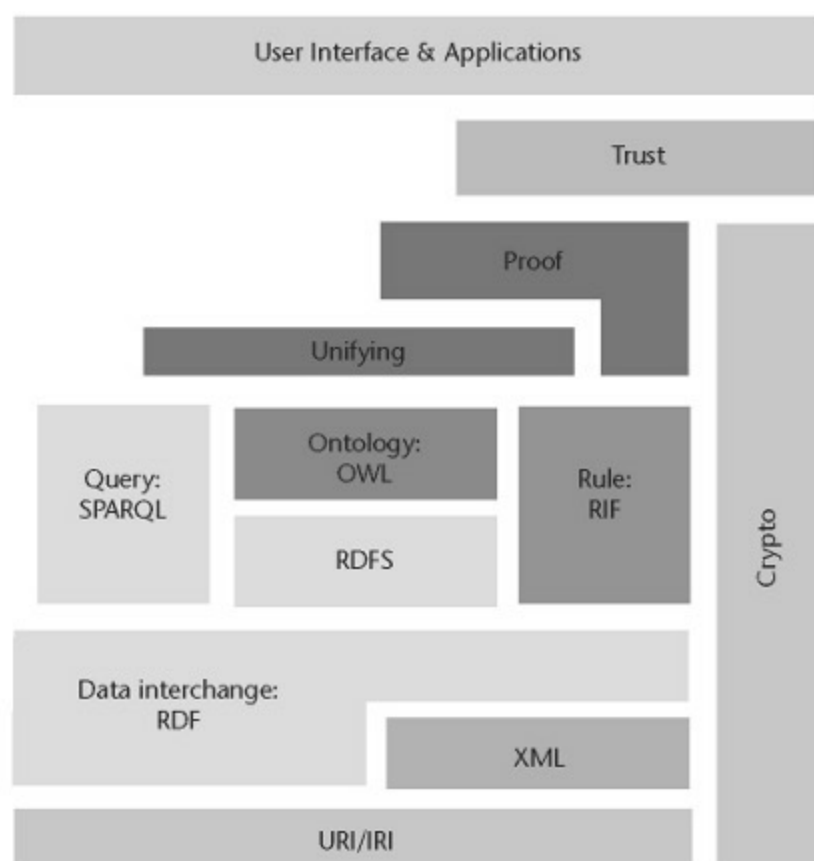


Figure 5. *L'organisation documentaire du Web sémantique en couches selon le W3C*¹⁹

Tentons une interprétation en regard de la période précédente en comparant les deux figures. Dans la proposition de Paul Otlet (figure 4), on trouvait tout en haut les auteurs qui pensaient le monde, le représentaient en concepts grâce à la science et le consignaient dans des documents. Le schéma du W3C met à leur place des utilisateurs qui, plutôt que représenter le monde, vont construire selon leurs besoins des réponses à leurs questions à partir des ressources documentaires existantes. On pourrait dire en raccourci que le monde n'est plus représenté par un travail scientifique préalable, mais que chacun se représente le monde à partir de données récoltées préalablement, sans préjuger de la façon dont elles ont été récoltées. En réalité, il n'y a même pas de récolte, comme l'indique la première citation de Tim Berners-Lee, le Web s'auto-alimente par son propre succès ou, dit autrement, publication et récolte sont confondues.

Cette différence dans le processus de publication et de récolte implique une autre, tout aussi fondamentale, dans celui de l'indexation et du catalogage. Dans la division du travail précédente, les documents, avant d'être intégrés à une collection de bibliothèque, n'étaient repérés que succinctement par quelques éléments, titre, auteur, résumé de contenu peu normalisés. Après leur publication, les bibliothécaires « traduisaient » les documents dans un langage documentaire commun pour les retrouver et les échanger facilement dans l'écosystème bibliothéconomique. Sur le Web, les deux moments étant confondus, il devient nécessaire pour un fonctionnement correct des moteurs de recherche de construire conjointement et soigneusement le document et son indexation ou catalogage, que l'on appellera dorénavant ses métadonnées, dès le dépôt du document sur un site pour qu'il puisse être retrouvé, c'est-à-dire exister. Données et métadonnées, document et indexation forment un tout. Bien des difficultés du fonctionnement du Web comme système documentaire viennent de la méconnaissance ou l'oubli, par les producteurs de documents, de ce changement dans les procédures. Inversement, des professions nouvelles sont apparues qui proposent aux responsables de sites Web d'améliorer leur visibilité par une meilleure indexation et aussi par diverses stratégies de repérage plus ou moins conformes à une logique documentaire, par exemple en augmentant artificiellement le nombre de liens pointant vers le site.

On pourrait discuter longtemps de la pertinence épistémologique de l'une ou l'autre posture. Là n'est pas mon propos, je voulais simplement souligner que, d'un point de vue documentaire, les flux sont inversés :

l'un part des producteurs de documents et classe ces derniers pour construire un système qui permettra au lecteur d'accéder au savoir ; l'autre part des lecteurs qui retrouvent ou reconstruisent les documents à partir de ressources déjà existantes et que l'on va classer. Pour Otlet, la confiance et la preuve sont du côté des auteurs ; pour le W3C, elles sont du côté des lecteurs.

Bien entendu, *in fine*, les processus sont comparables. Le dispositif d'Otlet s'adresse au lecteur et le *cake* du W3C ne peut fonctionner que si des auteurs ont préalablement déposé des ressources. Néanmoins, les priorités sont bel et bien inversées. Il n'y a plus d'effort de collecte de documents publiés, puisque la publication se fait directement sur la Toile. Les documents sont là et l'effort consiste à les rendre accessibles. C'est un peu comme si on avait effacé le premier moment du service de la bibliothèque (la constitution de la collection) pour se consacrer au second (l'accès au lecteur).

Les trois dimensions du Web sémantique

De plus, sans discuter les détails d'un schéma qui n'est, pour ses auteurs mêmes, qu'illustratif, on peut remarquer que l'on retrouve dans la succession des couches les trois dimensions du document que nous avons repérées au chapitre précédent et appliquées au livre (figure 3). Ainsi, le schéma, malgré sa complexité apparente, est aussi éclairant par la prise en compte, peut-être inconsciente, des différentes dimensions du document.

Les couches les plus basses (URL/URI, XML, RDF) concernent les adresses et les formats des ressources, le repérage par la forme, le « vu ». Les couches intermédiaires (SPARQL, OWL, RDFS, RIF) s'occupent de la recherche, de l'indexation, de la sémantique, de la représentation des connaissances, un traitement à partir du contenu, du texte, le « lu ». Enfin, les couches supérieures concernent des règles sociales (Unifying logic, Proof, Trust), celles-là mêmes qui supportent la fonction du document de transmission et preuve, le « su »²⁰.

Pour une meilleure lecture selon les trois dimensions, j'ai découpé le *cake* en tranches horizontales que j'ai redistribuées en triangle selon les trois dimensions du document. Cette représentation triangulaire a la vertu de casser l'empilement et sa lecture linéaire en montrant notamment les liaisons fortes qui existent entre les formats et les adresses d'un côté, et la confiance et la preuve de l'autre. Comparée à celle du livre (figure 3), elle met en valeur l'ampleur de la réingénierie documentaire en cours. Précédemment, nous trouvions une représentation du livre sur chacun des sommets du triangle, même si la différence de perspective soulignait les différences de dimensions. Cette fois, le document n'apparaît plus qu'au centre, le navigateur le reconstruit à la demande de l'internaute.

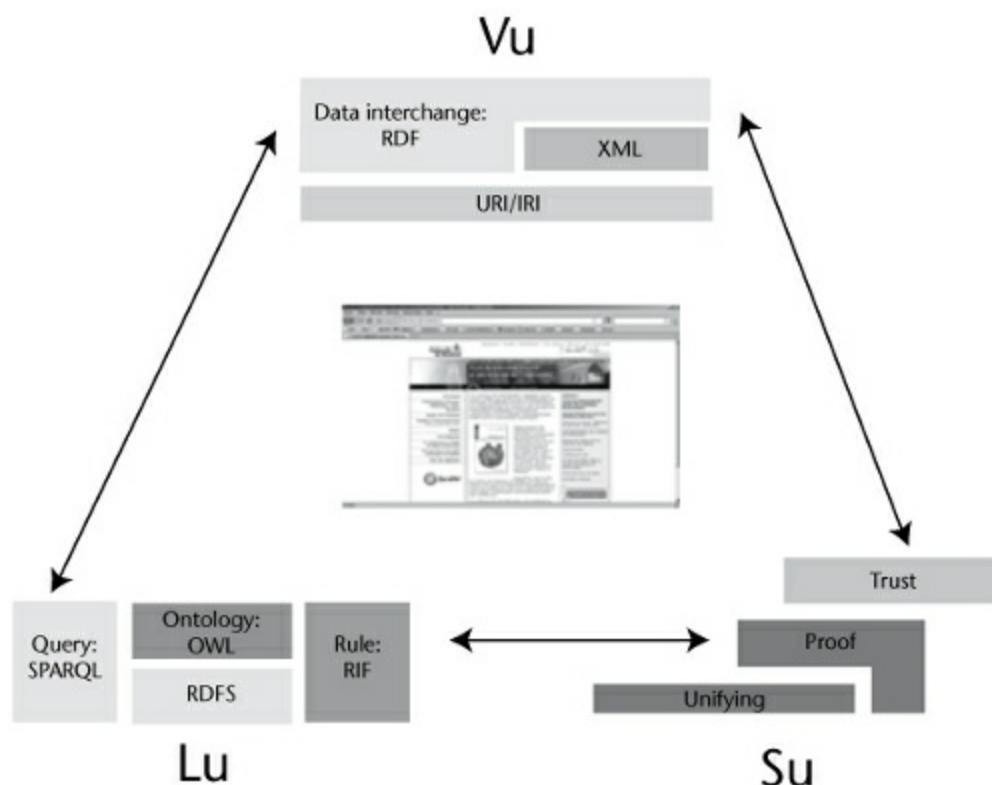


Figure 6. Les trois dimensions du document selon le Web sémantique

Cette succession est aussi une feuille de route dont la chronologie n'est pas anodine. Après avoir beaucoup travaillé sur les couches les plus basses, celles des formats et des adresses, l'accent est mis aujourd'hui sur les couches intermédiaires, celles de la recherche et de l'indexation. Sans trop pousser le parallèle, on peut remarquer que les bibliothèques ont suivi le même cheminement : l'accent était d'abord sur la récolte des documents (vu), puis avec Paul Otlet sur la classification et l'indexation (lu), et enfin avec Suzanne Briet sur la confiance et la preuve (su).

On pourrait dire enfin que le système documentaire a réintégré la construction du document. L'idée de « parenthèse Gutenberg » évoquée au chapitre précédent prend alors une tout autre ampleur. L'imprimerie avait sorti la production documentaire des bibliothèques, des infrastructures épistémiques de l'époque. Le numérique réintègre la production documentaire dans l'infrastructure épistémique contemporaine constituée par le Web. Mais la construction documentaire n'est plus réservée à une catégorie à part de la population (scribe, moines copistes). Au contraire, chaque internaute est invité à cette construction, en écrivant ou tout simplement en agencant les ressources documentaires entre elles, construisant des documents comme monsieur Jourdain faisait de la prose sans le savoir. Dans cette perspective, le Web 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. Les fonctions traditionnelles du document, transmettre et prouver, sont pilotées par l'action de l'internaute.

Le document au **xxi^e** siècle

Au moment où j'écris ces lignes, le Web a moins de vingt années. À l'échelle documentaire, il est encore dans l'enfance et la réingénierie est loin d'être terminée. Celle-ci n'est évidemment pas la simple concrétisation d'une feuille de route, même élaborée par un dispositif aussi consensuel que le W3C. Elle est soumise aux réactions diverses et parfois divergentes, spontanées ou stratégiques, des acteurs. Mais une chose est sûre, pour le document le changement est radical, au point que l'on peut sans doute parler de l'émergence d'un « néodocument ».

Dans l'esprit du W3C, la matrice de ce néodocument est sans doute le navigateur. L'histoire des navigateurs reste à écrire et serait sans doute pleine d'enseignements²¹. On peut déjà remarquer que le très

petit nombre de navigateurs disponibles et l'énorme succès de quelques-uns soulignent l'importance de la normalisation de fait. Le néodocument remplace de plus en plus les anciens documents qui règlent notre vie en société. Il a naturellement envahi notre quotidien et suppose un contrat de lecture pour être repéré, décrypté et éventuellement transmis. Il nous serait insupportable d'avoir à apprendre plusieurs contrats de lecture. Passer d'un navigateur à un autre, aux fonctionnalités pourtant très proches, est souvent agaçant, sans compter que nous devons aussi apprendre à gérer d'autres services concurrents (messageries, applications des mobiles, etc.). Mais la construction du néodocument ne se résume pas à celle du navigateur. J'en rappelle ci-dessous quelques brèves illustrations, sans prétendre à l'exhaustivité, en passant en revue ses trois dimensions.

Vu : des sites multiformes

Un collègue blogueur, Olivier Ertzscheid, a proposé récemment de classer les différentes formes des sites Web actuels²². Les formes ont beaucoup évolué en quelques années et ne sont sans doute pas encore stabilisées, toute classification est donc discutable. Mais la sienne a le mérite de souligner la diversité et multiplicité des formes, néanmoins limitées par le code, la perception humaine et la nécessité de ne pas rompre trop brutalement les contrats de lecture précédents. Il distingue six formes différentes qui constituent en quelque sorte une nouvelle typographie des sites Web :

- Les espaces « de surface ». Les pages Web classiques, bâties en HTML, incluant le Web statique et le Web dynamique (horaires de trains ou d'avions, par exemple), comprenant quelques liens internes et externes.
- Les espaces « profonds » qui utilisent des services de syndication ou d'agrégation de contenus qui récupèrent et signalent du contenu réalisé ailleurs.
- Les espaces d'« encapsulation », les zones intégrant des vidéos, des livres ou documents à feuilleter, des diaporamas, etc.
- Les espaces « marginalisés », les bandeaux pour faciliter la navigation dans le site qui peuvent être construits à l'initiative de l'auteur ou du lecteur.
- Les espaces « relationnels » pour passer à un autre site (sans nécessairement quitter le site d'origine), liens hypertextes, mais aussi signalements divers sur les réseaux sociaux par exemple.
- Les espaces « dispositifs » qui externalisent la fabrique de la page. Ils permettent de l'enrichir en fonctionnalités, avec par exemple les *plug-in* des navigateurs. Ils se chargent et sont utilisés de manière transparente pour le lecteur (ou sans que l'auteur en ait connaissance).

D'autres catégories pourraient être proposées, mais le plus frappant dans cette énumération est de constater combien la forme du document sur le Web a perdu sa stabilité et son universalité. C'est devenu, comme l'écrit Olivier Ertzscheid, un palimpseste où le texte se réécrit par l'adjonction d'apports divers, à l'encontre du principe de l'enregistrement à la base du document moderne qui était de reproduire sur un support des perceptions d'une même matrice, d'un protodocument.

Lu : homothétie et néodocuments

Le numérique et le Web en particulier ont l'ambition de conquérir la totalité de l'espace documentaire pour faire du terminal relié au réseau le seul point d'accès à l'ensemble du savoir. Cela signifie à la fois intégrer les documents traditionnels, prénumériques, ou ceux qui reprennent simplement sous un codage numérique l'organisation textuelle ancienne, et permettre l'émergence de documents nouveaux, originaux, inédits par l'organisation de leur contenu.

Pour que l'on puisse accéder aux documents antérieurs à la généralisation de la bureautique, il faut qu'ils aient été préalablement numérisés. Cette opération de transposition d'un support à un autre est une modification de la forme et ouvre la voie à la possibilité ultérieure de manipulation par les outils informatiques et donc d'autres modifications potentielles. On parle souvent de document homothétique pour signaler une transposition qui serait très semblable au document original dans son aspect, par exemple une page à l'écran (PDF) ou encore la simulation d'un codex. Mais cette notion reste vague et n'est qu'une facilité d'« effet diligence²³ ». La question plus fondamentale qui se pose est jusqu'à quel point le document est conforme à l'original. Plus précisément, c'est moins la notion de conformité formelle que celle d'invariant qui importe. Jusqu'à quel point de changements la transposition d'un document juridique, par exemple, aura-t-elle la même valeur de preuve que l'original, celle d'un document d'archives pourra-t-elle témoigner de l'action qu'il représente, celle d'une œuvre ne trahira-t-elle pas le projet de son auteur ? La question posée pour l'écrit s'adresse de plus en plus à l'image, dont les pixels peuvent être manipulés, et à l'audio et l'audiovisuel, dont le codage passe depuis longtemps par un signal électrique, mais dont le codage numérique facilite maintenant grandement la transformation.

Par ailleurs, la popularisation du Web a autorisé des initiatives documentaires inédites, qui s'appuient sur le renversement du flux documentaire. Ces néodocuments, dont nous discuterons quelques enjeux au chapitre 5, ne ressemblent plus que de loin aux documents traditionnels et en subvertissent les genres, même s'ils s'en inspirent souvent très directement. Parmi bien d'autres, citons :

- Wikipédia reprend la forme d'une encyclopédie, mais pas sa stabilité. Elle s'appuie au contraire sur la possibilité d'une remise à jour permanente, d'une extension sans limites et multilingue.
- Les blogs imitent les journaux intimes dans leur forme et leur individualité. Ils sont pourtant proposés à la cantonade, souvent reliés entre eux, accueillent et attirent les commentaires.
- Les réseaux sociaux retrouvent les échanges épistolaires, mais à une vitesse et une échelle qui les transforment radicalement.

Su : la vérité est dans la trace

Bien des analystes du Web considèrent que celui-ci est fondé sur la contestation des élites et des institutions. On pourrait rétorquer que c'est sans doute l'apanage de tout média nouveau d'être porté par ceux qui se considèrent à tort ou à raison comme exclus des anciens et trouvent là un vecteur d'expression et d'affirmation. En réalité, comme toujours, le nombre de ceux qui s'expriment en ligne est très minoritaire, constituant une sorte de nouvelle « élite branchée ». L'originalité du Web tient moins dans les possibilités nouvelles d'expression directe que dans la prise en compte des actions et réactions de l'internaute par leur traçabilité. Les suivis et calculs des traces laissées par les internautes ouvrent la voie à de nouveaux rapprochements et une nouvelle hiérarchie des documents. Google, on le sait, répond aux requêtes en calculant le nombre de liens et de clics. Facebook et les réseaux sociaux s'appuient sur un « graphe social » construit sur la base des relations entre leurs membres et propulsent des images et des textes en fonction des clics positifs. Les sites de partage permettent l'échange de photos, de vidéos, de recommandations et sont souvent hiérarchisés par leur nombre de connexions. Il y aurait ainsi une sorte de « sagesse des foules²⁴ » documentaire où la prise en compte souterraine des recherches individuelles ferait ressortir sur le réseau le savoir à acquérir.

Cette idée, même si elle n'est pas toujours exprimée aussi crûment, est très présente dans la construction du Web. Je ne prendrai qu'un seul exemple, emblématique me semble-t-il. Suite à l'impulsion du W3C vers un « Web des données », une nouvelle spécialité journalistique est apparue, le journalisme de données, qui consiste souvent à retraiter les données disponibles sur le Web sous forme graphique pour faire ressortir des tendances dans toutes les dimensions de la société. Voici ce que répond Michael Cross du *Guardian*, l'un des principaux représentants de ce nouveau journalisme, au fait que les mêmes données brutes pourraient alors appuyer deux thèses opposées : « [...] Plus les données sont disponibles, plus il y a de chances pour que les gens en parlent, les analysent, les croisent : la discussion offre une chance de réinterpréter les données de manière efficace et fiable, même si elles sont peut-être utilisées pour servir un certain agenda. Les données sont accessibles et, avec la puissance du Web, nous pouvons faire en sorte que l'interprétation la plus fiable se retrouve mise en avant²⁵. »

D'une modernité à l'autre

Par le passé, l'ordre documentaire ne s'est pas installé par hasard. Nous avons vu au premier chapitre qu'il s'est appuyé sur le modèle ancien et stable des bibliothèques et au deuxième chapitre qu'il s'est construit sur l'affirmation au XIX^e siècle du rôle du document. L'ordre documentaire construit alors était l'empreinte de son époque : celle de l'esprit scientifique, de la raison déductive, de la représentativité des institutions, du scientifique, de l'auteur romantique, etc. Bien des transformations ont depuis affecté les sociétés, bouleversant bien des vérités ou des rapports de forces que l'on croyait établis. Entre le début du siècle dernier et aujourd'hui, on peut faire un inventaire à la Prévert des changements majeurs : la famille, le rôle de la femme, l'espérance de vie, le rapport à la religion, l'instruction publique, l'organisation du travail, les temps de loisir, les transports, la géopolitique, les armes de destruction massive, le rapport à la nature, et, tout récemment, la raréfaction des ressources énergétiques, la mondialisation, la financiarisation de l'économie et le climat. La réingénierie documentaire reflète ou tente de refléter une organisation postmoderne de notre rapport à un monde changé. Elle est repérable aussi bien dans les sphères privée, collective et publique, qui se superposent elles-mêmes de plus en plus. Le croisement et parfois le choc

entre les productions documentaires autrefois réservées à l'une des sphères sont l'une des expressions de cet ordre nouveau. Aujourd'hui, cet ordre fait plutôt référence à la réflexivité, à un savoir neutre et limité, à une raison statistique, aux opinions, à l'individu.

Objectivité et réflexivité

Pédauque a consacré un chapitre complet aux relations entre document et modernités²⁶. On y lit notamment : « Une bonne part des relations dans les sociétés dites “modernes” sont fondées sur et cimentées par la stabilité du document papier et sa reproductibilité industrielle à l'identique (effet de série) ou encore sa permanence sécurisant les contrats, les règles et les identités. Plus encore, cette modernité est à mettre en relation avec une certaine façon d'écrire et de penser dans la linéarité et l'argumentation. [...] La seconde modernisation introduit l'exigence d'une réflexivité, d'un savoir qui se déclare lui-même et qui se contrôle en connaissant ses limites et ses conditions de production. Ainsi, le modèle politique change (cf. le principe de précaution) mais aussi, pour ce qui nous concerne ici, les politiques documentaires. Le numérique introduit de façon massive et inédite à cette échelle la réflexivité sur notre propre activité documentaire. La documentation générale de nos activités, y compris les plus triviales, constitue désormais une seconde nature, qui correspond bien à cette seconde modernisation²⁷. »

J'ai rappelé, en citant Paul Starr au chapitre précédent, que la notion d'objectivité était apparue comme une règle journalistique vers 1920 sous l'impulsion de Walter Lippmann. Reprenons les mots clés de la citation et comparons-les à ceux de Michael Cross notés plus haut, en ayant en tête que 1920 correspond à l'organisation systématique des systèmes documentaires qui a pour modèle la science positive, tandis que 2010 correspond à la transformation de notre rapport au document. Le tableau renvoie un écho du changement du rapport au document dans les pratiques journalistiques.

Tableau 2. Valeurs du journalisme d'investigation (1920) versus le journalisme de données (2010)

1920	2010
Objectivité ; les journalistes s'inspirent de la science en développant un « sens de la preuve »	Plus les données sont disponibles, plus il y a de chances pour que les gens en parlent
Véracité	Efficacité et fiabilité
Responsabilité	L'interprétation la plus fiable se retrouve mise en avant

Même si le propos de Michael Cross n'a évidemment pas le poids normatif de celui de Walter Lippmann, le contraste est net entre les valeurs documentaires des deux débuts de siècle. D'un côté, la recherche d'une véracité, c'est-à-dire une croyance partagée considérée comme objective, par la responsabilité des journalistes qui s'inspirent du modèle de la science ; de l'autre, la recherche de l'efficacité, c'est-à-dire une croyance qui obtient des résultats et non qui est vraie en soi, construite et obtenue par l'interaction des internautes à partir de données disponibles. D'un côté, on reconstruit le passé à partir de preuves pour en avoir une idée juste ; de l'autre, on négocie son interprétation à partir de l'agencement d'éléments existants dans le croisement des opinions pour qu'il serve au présent. D'un côté, on établit une véracité sur la base des faits ; de l'autre, on se rend utile sur la base d'un consensus. On ne dira plus « c'est vrai, je l'ai lu dans le journal », mais plutôt « j'y crois parce que les autres y croient sur le Web ».

De la neutralité

Dans ce nouveau régime de vérité documentaire, la neutralité devient un critère essentiel. En effet, pour pouvoir fonctionner idéalement, ce repérage et cette hiérarchisation *a posteriori* nécessitent que toutes les unités documentaires soient également disponibles sur le réseau. Ainsi, les plus engagés des internautes dans la construction du réseau se battent contre toutes les tentatives de censures politiques, de verrouillages économiques, juridiques ou techniques pour que celui-ci reste le plus ouvert et transparent. On retrouve aussi cet affichage de neutralité, par exemple, dans le fonctionnement de l'algorithme de Google. La

mission de Google est, selon les propres termes de la firme : « organiser les informations à l'échelle mondiale dans le but de les rendre accessibles et utiles à tous ». Google est soupçonné régulièrement de favoriser ses propres sites ou ceux de ses principaux annonceurs, mais la démonstration d'un parti pris manifeste réduirait la légitimité de son service et le placerait dans une position difficile vis-à-vis des lois antitrust, compte tenu de sa domination dans le monde occidental.

De plus, l'insistance sur la neutralité a peut-être aussi des causes plus cognitives. Il est probable que le choc des subjectivités ne serait pas supportable si nous ne disposions d'un socle commun de croyances, considérées comme « neutres », à partir desquelles nos désaccords ne sont plus que des opinions différentes qui peuvent développer quelques antagonismes sans pour autant remettre en cause les fondements du vivre ensemble. On retrouve cette exigence de neutralité de point de vue par exemple dans les règles de rédaction de Wikipédia : « La neutralité de point de vue est un des principes fondateurs (absolus et non négociables) de Wikipédia que tous les articles doivent respecter. Ainsi, d'après ce principe, les articles se doivent d'être écrits de façon à ne pas prendre parti pour un point de vue plutôt qu'un autre. Au contraire, il s'agit de présenter tous les points de vue pertinents, en les attribuant à leurs auteurs, mais sans en adopter aucun. Il ne faut *a priori* jamais affirmer, sous-entendre ou même laisser croire qu'un des points de vue est d'une quelconque façon meilleur, égal ou moins bon qu'un autre²⁸. » On retrouve aussi cette hypothèse de neutralité dans l'idée même d'un « Web de données ». Les données sont les plus petites unités documentaires représentant le monde réel. Ces unités documentaires sont « données », c'est-à-dire prises comme telles, non discutées, objectives en soi, neutres, même si l'on peut ensuite discuter leur agencement.

La neutralité est fondée sur un patrimoine documentaire partagé et ne serait pas envisageable sans celui-ci. La représentation commune du monde n'est plus construite à partir des objets analysés par la science, qui garantissait par ses règles objectives notre rapport au monde, comme le suggérait Paul Otlet, mais sur la base d'une multiplicité de documents existants auxquels maintenant chacun peut se référer directement. Les possibilités d'enregistrement immédiat des événements et de relation directe avec eux par des témoins, grâce aux outils mobiles, photographiques et textuels, tout comme celles de la consultation d'avis d'experts sur les sujets les plus pointus, ajoutent à cette sensation d'encyclopédie universelle. C'est un peu comme si tout avait été dit, et continuait à se dire, et qu'il suffisait de réagencer ce dit de différentes façons pour pouvoir éclairer notre représentation du monde sur n'importe quel sujet.

Comme l'objectivité dans le monde moderne, cette neutralité n'échappe évidemment pas aux pressions politiques et commerciales aux manipulations de toutes sortes, et doit s'apprécier selon les contextes culturels. Elle est aussi discutable que l'objectivité, mais ce n'est pas mon propos. Ce qui m'importe, c'est qu'elle soit bâtie sur un dispositif documentaire et passe, comme son ancêtre, par des règles pratiques que l'on peut décrire.

« Je » est un document

Le troisième changement spectaculaire de l'ordre documentaire nouveau, en phase avec ceux de la société, est de nous avoir transformés nous-mêmes en documents ou plutôt d'avoir radicalisé cette transformation, démarrée depuis fort longtemps. Ce changement est en quelque sorte l'envers et le complément du précédent. Alors que la neutralité de point de vue efface les subjectivités pour privilégier le contenu documentaire, la documentarisation des individus replace la subjectivité au centre du processus documentaire. La transformation de l'individu en document, ou peut-être faudrait-il dire en dossier documentaire, n'est pas née avec le Web. Elle vient de loin et commence par la célébration de personnages, saints, grands de ce monde ou criminels. L'étape suivante fut peut-être la construction progressive de l'auteur romantique et celle du chercheur entre le XVIII^e et le XIX^e siècle, parallèlement à la montée de l'éditeur. Avec l'arrivée du cinéma et de la radio, le phénomène s'est élargi par le *star-system* à Hollywood et dans les *majors* de la musique. La diffusion massive de l'audiovisuel par la télévision l'a considérablement amplifié, jusqu'à, *in fine*, faire « star » l'homme ou la femme ordinaires dans les télérealités. Chacun maintenant peut avoir l'espoir de devenir célèbre, d'avoir son quart d'heure de gloire comme l'annonçait Andy Warhol en 1968. Et c'est peu dire que l'Internet et le Web ont multiplié cette faculté ! Blogs et réseaux sociaux nous obligent à gérer aujourd'hui notre identité numérique, c'est-à-dire à contrôler plus ou moins laborieusement nos archives et nos traces, et font de tout un chacun un auteur romantique aux petits pieds, virtuellement un document, en déplaçant les frontières entre les traces de notre vie privée et celles de notre vie publique, brusquement élargie potentiellement à toute la planète. Les albums de famille étaient là pour nous rappeler comment nous et nos proches étions tout le long de notre vie, mais ces photos n'avaient pas vocation à sortir du cercle de nos intimes. Aujourd'hui, les cercles sont poreux. Il n'est pas surprenant que les adolescents soient les premiers touchés par cette nouvelle question existentielle, puisque l'adolescence

est une période cruciale pour construire son identité sociale. Mais les adultes doivent eux aussi gérer la complexité de relations souvent compartimentées et ici parfois brutalement mélangées. Ainsi, nous devons apprendre à gérer notre « je » document.

Celui-ci est aussi devenu un document utile pour l'organisation en société, potentiellement pour au moins trois champs de connaissances : le savoir scientifique, le commerce, le politique. Les sciences humaines et sociales, le marketing et l'action politique ont l'habitude d'observer et d'analyser les hommes en groupes, agrégés en catégories diverses. Tant qu'il s'agit d'agréger des données anonymes pour améliorer la connaissance, le service ou la régulation des sociétés, ces technologies ont des effets positifs. Mais, dès lors que la fonction de transmission et celle de preuve concernant les individus s'évaluent sur une grande échelle et sur une grande profondeur, les risques de manipulations au travers de leur double documentaire, petites ou grandes, ponctuelles ou planifiées, sont réels.

Les hommes ont externalisé leur mémoire pour pouvoir construire des connaissances en manipulant des objets porteurs de sens, les documents. Michel Serres, toujours optimiste, salue l'espace nouveau laissé à l'intelligence par l'allègement de notre mémoire²⁹. Le danger serait de retourner le raisonnement en se servant de ces mêmes objets pour orienter nos comportements. Michel Foucault y aurait sans doute trouvé confirmation de sa réflexion sur le pouvoir³⁰.

Économie de service

Un ordre documentaire nouveau se construit donc sous nos yeux. Il s'élabore dans la continuité du précédent et en même temps s'en éloigne, s'adaptant à la société du troisième millénaire, accusant sans doute aussi certains de ses traits. Paradoxalement, cet ordre documentaire s'est bâti à l'écart du développement des médias de masse, pourtant principal pourvoyeur de documents publics au ^{xx}e siècle. Le Web favorise l'activité des internautes, la recherche d'informations, une production de services en phase avec la bibliothèque et avec l'évolution générale de l'économie vers le service. Le document, lui-même produit à la demande, n'est plus un bien stable, mais un service labile. Le Web contredit la diffusion unilatérale des médias de masse traditionnels. Ainsi Yochai Benkler, chantre d'une économie positive du Web, fait dans son livre *La Richesse des réseaux*³¹ deux constats : 1) l'économie de l'information en réseau vient supplanter l'économie de l'information industrielle, qui a caractérisé la production de l'information depuis environ la seconde moitié du ^{xix}e siècle jusqu'à la fin du ^{xx}e ; 2) la production hors marché occupe une importance grandissante ; un individu peut désormais contacter, informer ou instruire des millions d'autres personnes à travers le monde. Ces constats sont largement repris par les acteurs et analystes du Web. Dans les chapitres suivants, nous verrons qu'il est nécessaire de les nuancer. Même si les contradictions et le bouillonnement sont bien là, les logiques industrielles marquent aussi très fortement le nouvel ordre documentaire.

Notes du chapitre 3

1. Françoise LÉVIE, *L'Homme qui voulait classer le monde : Paul Otlet et le Mundaneum*, Les Impressions nouvelles, Bruxelles, 2007.
2. Paul OTLET, *Traité de documentation*, op. cit., p. 41. Pour une meilleure lisibilité, les textes du tableau ont été recomposés.
3. *Ibid.*, p. 428.
4. British Broadcasting Corporation.
5. *Ibid.*, p. 430.
6. *Ibid.*, p. 430.
7. Respectivement MEDLARS et DIALOG.
8. Vannevar BUSH, « As we may think », *The Atlantic*, juillet 1945.
9. *Ibid.*, p. 6 (traduction personnelle).
10. L'inventeur de la classification à facettes est l'Indien S. R. Ranganathan.
11. Colin BURKE, « The other Memex : the tangled career of Vannevar Bush's information machine, the rapid selector », *Journal of the American Society for Information Science* 43, n° 10, 1992, p. 648-657. Cité par Guy TEASDALE, « L'hypertexte : historique et application en bibliothéconomie », *Cursus* (1) 1, 1995.
12. Tim BERNERS-LEE, « Information management : a proposal », Document de travail, CERN, mars 1989, mai 1990 (traduction personnelle).
13. HyperText Transfert Protocol.
14. Uniform Resource Identifier.
15. Hypertext Markup Language.
16. Tim BERNERS-LEE, James HENDLER et Ora LASSILA, « The semantic Web », *Scientific American Magazine*, 17 mai 2001.
17. *Ibid.* (traduction Élisabeth Lacombe et Jo Link-Pezet).
18. Tim BERNERS-LEE, « On the next Web », conférence TED février 2009. <www.ted.com>.
19. Steve BRATT, « Semantic Web, and other technologies to watch », diaporama W3C, janvier 2007, D24.
20. Pour une présentation simple de ces différents protocoles, on pourra consulter la version anglaise de Wikipédia (<<http://en.wikipedia.org>>) à la rubrique « Semantic Web ».
21. Pour une présentation infographique de l'évolution des navigateurs réalisée par les développeurs de Chrome : <<http://evolutionofWeb.appspot.com/>>.

- ²². Olivier ERTZSCHEID, « De quoi la page Web est-elle le nom ? Ou l'enluminure du code », *Affordance*, 20 mars 2011.
- ²³. « Les premiers wagons ressemblaient à des diligences et les premières automobiles, à des voitures à cheval. Les mentalités, habituées à des techniques désormais dépassées, utilisent les nouveaux outils avec des protocoles anciens, c'est ce que j'appelle l'effet diligence. » Jacques PERRIAULT, *Effet diligence, effet serendip et autres défis pour les sciences de l'information*. Non daté. <archive.limsi.fr>.
- ²⁴. James SUROWIECKI, *La Sagesse des foules*, Jean-Claude Lattès, Paris, 2008.
- ²⁵. « "C'est une période excitante pour l'open data" : entretien avec Michael Cross », *Regard sur le numérique*, 25 février 2011. <www.rslnmag.fr>.
- ²⁶. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, op. cit., p. 159-207.
- ²⁷. *Ibid.*, p. 161.
- ²⁸. Voir l'article « Wikipédia : neutralité de point de vue » sur <<http://fr.wikipedia.org>>.
- ²⁹. Michel SERRES, *Les Nouvelles Technologies : révolution culturelle et cognitive*, op. cit.
- ³⁰. Michel FOUCAULT, *Surveiller et punir. Naissance de la prison*, Gallimard, Paris, 1975.
- ³¹. Yochaï BENKLER, *La Richesse des réseaux : marchés et libertés à l'heure du partage social*, Presses universitaires de Lyon, Lyon, 2009.

4

L'économie du document

Les observateurs qui s'intéressent à l'économie de la culture ou à celle de l'information sous-estiment les caractéristiques économiques particulières du document. Les médias de masse se sont maintenant insérés dans tous les moments de notre vie, du moins tous ceux où nous restions disponibles pour eux, et cette position envahissante masque des mouvements plus profonds et de plus long terme. La prise en compte de la bibliothèque comme un média à part entière et la pleine conscience des trois dimensions du document autorisent une nouvelle interprétation des industries culturelles. Les trois dimensions du document font ressortir trois modèles de publication, anciens mais toujours d'actualité, parmi lesquels la bibliothèque a toute sa place. Le Web peut être alors compris comme une étape provisoirement ultime et radicale par son emprise documentaire sur notre quotidien. Il reprend certaines caractéristiques des anciens modèles, tout en imposant son économie propre, issue de la réingénierie documentaire. Son influence est forte sur les médias plus anciens, les transformant en « industries de la mémoire ».

Ce chapitre se conclut par un plaidoyer pour la prise en compte dans l'économie d'un secteur reconfiguré, celui du document, plus large que celui de l'économie de la culture et plus précis que celui de l'économie de l'information.

Les trois modèles de valorisation du document

La proposition du deuxième chapitre de représenter le document sous ses trois dimensions permet de relire l'histoire économique de sa publication en remarquant que, selon les époques et l'avancement des techniques d'enregistrement et de diffusion, différents modèles industriels sont apparus. Le tableau ci-dessous résume la proposition en classant les trois principaux modèles historiques de valorisation du document selon les trois dimensions de celui-ci. Un tableau n'a pas valeur démonstrative en soi, pourtant il est remarquable de constater à quel point le classement fonctionne. C'est un indice supplémentaire de la pertinence du découpage. Précisons avant de le lire ligne par ligne que chaque modèle intègre nécessairement les trois dimensions du document auxquelles il ne peut échapper, mais chacun en privilégie une.

Tableau 3. Les trois dimensions du document déclinées en modèles de publication

	Fonction	Nature	Échange	Modèle	Biface
Vu	Création	Matériel Rival	Bien Appropriation	Édition	Auteurs Lecteurs
Lu	Mémoire	Immatériel Non rival	Accès Commun	Bibliothèque	Collectivité Lecteurs
Su	Transmission	Immatériel Rival	Attention Espace-temps	Spectacle (Dialogue)	Annonces Spectateurs

Vu : édition

Le principe de l'édition est de considérer le document comme un objet unique que l'on peut reproduire : une œuvre et des exemplaires. La fonction première est de sélectionner ou susciter les propositions des auteurs pour construire des « œuvres », circonscrites et finies, des protodocuments, en fonction d'un lectorat potentiel. Il s'agit là d'un marché de la création avec tous ses aléas et spéculations. La diffusion est ensuite assurée grâce à la reproduction de ce protodocument en exemplaires par le truchement d'une imprimerie pour le livre (je m'en tiens ici à l'exemple du livre, mais le raisonnement peut s'étendre à tout produit édité) et réalisée par un réseau de distribution dont le dernier maillon est le libraire.

Mais un protodocument, comme objet singulier, est économiquement fragile. À partir d'un document diffusé, un exemplaire, il peut être facilement copié à son tour et rediffusé. Son exploitation échappe alors à son auteur et aussi à son éditeur, qui l'a sélectionné et lui a souvent permis d'arriver à maturité. La solution à ce problème a été imaginée et imposée progressivement à partir du XVIII^e siècle grâce au droit d'auteur (ou au copyright, selon les traditions nationales). Dès lors, l'impression sur un support matériel débouche sur des productions économiquement hybrides. Elles sont vendues comme des objets dont l'acheteur a la pleine propriété mais pourtant ne dispose que d'un droit d'accès aux modalités restrictives à l'œuvre (il ne peut la rediffuser librement). Ainsi, les produits de l'édition, les exemplaires, sont construits comme des biens matériels rivaux (la consommation par un agent empêche la consommation par un autre). Ils sont pleinement appropriés par l'acheteur, mais en pratique consommables par lui seul et un cercle restreint autour de lui. Il est important de noter que l'exemplaire, le document, est un bien rival par construction juridique. Un bien dit « rival » en économie est un bien dont la consommation par une personne prive une autre de le consommer. Lorsque vous lisez un document édité (un exemplaire que vous avez acquis), personne ne peut le lire en même temps que vous, sauf si vous en avez fait une copie, ce qui est prohibé par la propriété intellectuelle.

L'édition fonctionne bien sur un double marché, ou un marché biface, dont les éditeurs sont les agents, celui des œuvres en amont avec les auteurs, celui des exemplaires en aval avec les lecteurs. Œuvres et exemplaires (ou protodocuments et documents) ne sont pas de même nature économique, n'ont pas les mêmes modalités de production et de valorisation, mais sont évidemment articulés. Plus un éditeur est capable d'attirer d'auteurs à succès, et donc potentiellement de produire d'œuvres, plus il vendra de livres, c'est-à-dire d'exemplaires, et vice versa.

Contrairement à ce que pourrait laisser croire une réflexion trop rapide, dans les deux cas l'économie est fondée d'abord sur la forme et non sur le contenu. Le contenu est fixé sur l'objet, le « vu ». L'objectif de l'éditeur est avant tout de préserver la singularité du contenu en fixant sa forme. La relation économique entre les deux marchés se fait par les droits d'auteur patrimoniaux qui sont proportionnels au nombre d'exemplaires vendus. Le fonctionnement particulier de ce marché a fait l'objet d'une abondante littérature sur laquelle je ne reviens pas ici¹. L'édition est en effet souvent considérée comme le modèle fondateur de l'économie de la publication. En réalité, il s'agit du plus récent des trois modèles. Pour que le modèle éditorial s'installe, il a fallu d'abord que l'imprimerie s'invente et se perfectionne, puis que progressivement le commerce du livre se régule à un niveau international. La maison d'édition n'est apparue qu'au XIX^e siècle, quatre cents ans après la naissance de l'imprimerie². Mais il s'agit d'une économie de la création, très valorisée dans nos sociétés issues de la révolution industrielle.

Lu : bibliothèque

Comme nous l'avons vu au premier chapitre, le modèle de la bibliothèque est beaucoup plus ancien, il remonte à l'Antiquité. La bibliothèque utilise différemment les caractéristiques des documents. Plutôt que de tenter de brider la non-rivalité d'un document comme dans le modèle précédent, elle cherche au contraire à la développer pour construire un écosystème autonome. Un même document peut être consulté par différents lecteurs n'ayant pas de relation entre eux. Ainsi, c'est bien la non-rivalité du document qui fonde le modèle. Le cœur du métier est d'enrichir des entités humaines (individus ou collectivités) par l'accès à des documents ou plus précisément par l'accès au savoir contenu dans ces documents que la bibliothèque sélectionne, accumule et conserve. La valeur d'un ou de plusieurs documents, construite par leurs auteurs et producteurs dans le modèle précédent, est ici reconstruite par leurs lecteurs. Il s'agit d'une économie de service, immatérielle même si elle manipule des objets matériels ; contrairement à l'économie de biens précédente, ce qui est « vendu » n'est pas un objet mais un processus, la lecture.

Jusqu'à une période très récente, la bibliothèque ne pouvait pas financer son service par un marché : les coûts de collecte, traitement et gestion des documents étaient trop élevés pour les répercuter aux lecteurs, alors même que leur propension à payer était très faible du fait justement du caractère non rival de l'information mis en avant. Mais l'écosystème créé à l'écart du marché pouvait, par agrégation d'une collectivité à l'autre, élargir grandement l'offre grâce justement à cette non-rivalité. Un document absent

d'une bibliothèque pouvait être fourni par une autre, l'effet réseau étant tout à fait efficient dans ce modèle. Les bibliothèques ont alors constitué ce que l'on appelle aujourd'hui un « bien commun³ ». Les documents collectés ont le potentiel d'alimenter l'action collective. Ces objectifs participent à la configuration de la collectivité qui réunit les individus. La collectivité investit dans ce capital informationnel pour que les individus répondent mieux à ses objectifs. Ainsi, les bibliothèques sont financées par les collectivités qui les abritent comme des biens communs pour la réalisation de leurs objectifs et on pourrait sans doute relire leur histoire comme la constitution progressive des règles indispensables à la non-dégradation du bien commun par les individus qui l'exploitent.

Leur économie est aussi biface, les bibliothécaires devant justifier du service rendu à leurs usagers vis-à-vis des représentants de ceux-ci qui les financent. Il y a bien des externalités croisées, dans la mesure où un financement supérieur permet d'améliorer le service et potentiellement d'augmenter la fréquentation, et inversement une augmentation de la fréquentation fournit des arguments au bibliothécaire pour demander plus de financement.

Rappelons enfin que, contrairement au modèle de l'édition, celui de la bibliothèque n'est pas, ou plus depuis le développement de l'imprimerie, fondé sur la gestion de la création, mais privilégie la mémoire (immédiate ou à long terme). Il faut prendre le terme « mémoire » comme dynamique et non statique. Les documents sont accumulés, stockés certes, mais aussi traités, indexés, classés pour être retrouvés, consultés et confrontés. Pour ce modèle, les documents existent déjà donc. L'important est de les mettre à disposition du lecteur, dont l'activité est créatrice de richesse. Cette séparation entre création et mémoire n'a été effective qu'à partir du moment où l'imprimerie a fixé la forme des documents par leur reproduction mécanique et que l'activité de copie et de diffusion a alors quitté le giron des bibliothèques. Le Web, refermant la « parenthèse Gutenberg⁴ », donne aujourd'hui un nouvel écho à cette ancienne organisation.

Su : spectacle, dialogue

Comme nous l'avons vu au deuxième chapitre, le terme « *documens* » dans son orthographe ancienne désignait au Moyen Âge aussi bien une leçon présentée par un orateur, un spectacle ou un dialogue, qu'un écrit consigné sur un support. L'accent était mis alors principalement sur la transmission, l'enseigner de *docere*. L'acception moderne de *document* qui s'est imposée au XVIII^e siècle privilégie une inscription sur un support, un enregistrement. Mais l'arrivée des techniques d'enregistrement et de diffusion à distance de l'image et du son, à partir de la fin du XIX^e siècle, a autorisé une convergence nouvelle entre l'étymologie du mot et son utilisation moderne. Il est dès lors possible de diffuser à distance et d'enregistrer le spectacle *via* l'audiovisuel et même de tenir à distance des échanges *via* le téléphone. Cette convergence n'est pas pour nous anecdotique.

Le principe même du spectacle est de capter notre attention. L'ensemble du dispositif est conçu pour cela dès sa version antique : amphithéâtre, salle, scène, éclairage, découpage, rebondissements, etc. Le spectateur paye pour être captivé. La radio-télévision a poursuivi cette tradition en lui ouvrant un marché supplémentaire, déjà défriché par la presse populaire à la fin du XVIII^e siècle : celui des annonceurs. Puisque l'attention est captée, cette fois jusqu'au domicile des spectateurs, il est possible de la revendre à des annonceurs désireux de faire passer un message, et en premier lieu ceux qui ont quelque chose à vendre. Dans ce modèle, le document, ici le programme radio ou télé, est un bien intermédiaire, mais intermédiaire ne signifie pas secondaire. C'est lui qui attire le spectateur et capte son attention.

Exploiter au mieux la valeur et la fragilité de l'attention humaine est un équilibre délicat : il faut trouver le juste prix que le lecteur ou le spectateur sont susceptibles de dépenser en temps, argent ou énergie pour accéder à l'information qui servira à capter leur attention sans la saturer ni la décourager par des messages publicitaires trop ostensibles et parasites puisque non souhaités. Ainsi l'économie de la radio-télévision est, elle aussi, biface. Elle articule deux marchés, celui des spectateurs et celui des annonceurs. Chacun des marchés s'adresse à des clients différents, propose des marchandises différentes, mais influe fortement sur l'autre. Le marché des lecteurs permet de capter une attention et donc de valoriser le bien qui sera proposé à l'annonceur : l'espace publicitaire qui accueillera son message. Par les revenus générés, le marché des annonceurs permettra à son tour de proposer de meilleurs programmes ou d'abaisser le prix d'abonnement et ainsi d'augmenter l'audience, et par un effet de boucle de valoriser le prix de l'espace publicitaire. C'est le principe même d'un marché biface.

Le téléphone n'a pas repris la tradition documentaire moderne, mais plutôt celle de l'Antiquité, celle de la conversation, que préférait Socrate. Mais il s'agit d'un média de transmission qui opère aussi par le captage de l'attention. Lorsque le téléphone sonne, nous sommes dérangés dans notre activité du moment et « obligés » de répondre. On pourrait dire que le téléphone force notre attention en nous tapant sur l'épaule pour nous contraindre à engager un dialogue avec celui qui nous appelle, tandis que la radio-

télévision agit de même, mais en pointant du doigt vers le téléviseur pour nous inviter à visionner ses programmes. Leur économie est bien différente. Par certains côtés, celle du téléphone se rapproche plus de celle de la bibliothèque. C'est une économie de l'accès, où les effets de réseau sont très puissants et favorisent la constitution d'oligopoles.

Dans les deux cas, radio-télévision et téléphone, l'essentiel est la construction de la relation pour assurer la transmission au travers de la captation de l'attention. L'accent est bien mis sur la troisième dimension du document, le « su ». L'attention est un bien immatériel limité, donc rival (une attention captée ne sera plus accessible), limité par le temps disponible du spectateur ou de l'interlocuteur, mais limité aussi par la capacité de ces derniers à traiter les informations. Herbert Simon a montré que la rareté ne réside plus aujourd'hui dans le nombre d'informations disponibles, mais bien dans notre capacité à les traiter. Il a ainsi proposé la notion de « rationalité limitée » pour signifier que nos choix ne sont rationnels qu'à l'intérieur des limites de notre capacité à trouver et traiter les informations qui peuvent les éclairer⁵.

Ce troisième modèle, celui de la transmission, fondé sur une économie de l'attention, va tout comme le précédent se renouveler avec le Web, qui le fera basculer complètement dans une logique documentaire, car sur le Web, par construction, tout est enregistré, contrairement à la radio-télévision et au téléphone.

Articulation des modèles

En présentant ces trois modèles, j'ai forcé le trait pour en faire ressortir les principales caractéristiques. Dans la pratique, bien des situations sont hybrides, soit qu'elles superposent les modèles (ainsi, la presse vend à la fois des biens, les exemplaires en kiosque, des accès, les abonnements, et de l'attention, les espaces publicitaires ; la radio-télévision vend aussi de l'accès et aussi aujourd'hui des biens par la vidéo à la demande), soit qu'elles les articulent (les revues scientifiques articulent édition et bibliothèque ; le cinéma articule édition et spectacle).

La gestion de l'attention

La prolifération des documents a accru le poids de la troisième dimension, la relation, qui est de plus en plus convoitée du fait des limites de l'attention. Chacun des médias prend sa part d'attention et les modèles entrent en concurrence. Il est alors utile de mieux percevoir comment chacun gère l'espace-temps du lecteur. À y regarder de plus près, il semble que, de l'édition à la radio-télévision en passant par la bibliothèque, la gestion de l'espace-temps de la production-consommation des documents est régulée par l'organisation tarifaire.

Dans le modèle éditorial traditionnel, la production est saisonnière. Les livres, les disques sont présentés généralement sur le marché deux ou trois fois par an à la rentrée de l'automne, avant les fêtes de fin d'année ou avant un salon ou un festival marquants. Les nouveaux titres chassent progressivement les anciens, faute de place suffisante sur les étals et dans l'attention des consommateurs, bien que certains classiques bravent le temps. Comme lecteur ou amateur, je dois faire l'effort de me déplacer dans une boutique pour les acheter. Mais je lis mon livre, écoute mon disque, regarde mon DVD à ma guise, où je veux, quand je veux. Je l'ai payé intégralement, j'ai « gagné » la liberté de le consommer selon mon bon vouloir. Je peux même, si je le souhaite, le déposer sans l'ouvrir sur une étagère où il restera discrètement à la vue comme une promesse d'un temps de lecture ou d'écoute à venir.

A contrario, le temps de la radio-télévision est continu, on l'appelle parfois l'industrie du flot. La production des programmes est en effet industrielle et le personnage clé en est le programmeur, c'est lui le maître du temps. Je regarde les programmes de télévision gratuitement chez moi à partir d'un récepteur, mais je suis « contraint » par l'ordre des émissions qui s'adaptent à mon emploi du temps et me bombardent de publicités. Le téléviseur trône dans mon salon et, même s'il est de plus en plus plat, il est de plus en plus large et terne lorsqu'il est éteint. Tout se passe comme si les annonceurs avaient acheté la gestion du temps des téléspectateurs ou encore comme si les téléspectateurs avaient accepté d'aliéner leur temps pour pouvoir consommer des programmes gratuitement.

Le raisonnement est encore plus probant si l'on considère la situation intermédiaire de la presse, qui est payée deux fois : une fois par ses lecteurs, une autre fois par ses annonceurs. L'espace-temps de la presse est plus contraint que celui de l'édition, mais moins que celui de la radio-télévision. La presse est alimentée par l'« actualité », distribuée dans un réseau dense de kiosques ou même au domicile du lecteur, selon un rythme serré et régulier. Je paie moins cher mon journal, car il est vendu une seconde fois aux annonceurs,

mais je suis « obligé » de le lire le jour même et régulièrement. Les annonceurs ont acheté une part de cette gestion du temps ou inversement les lecteurs ont accepté de s'aliéner une part de leur liberté de consommation pour acquérir le journal à un tarif moins élevé.

La bibliothèque est dans un tout autre rapport à l'espace-temps et à la tarification. Je l'ai dit, la bibliothèque est le média du temps long et elle est financée par la collectivité. L'éventail de titres disponibles est bien plus conséquent et stable que celui de l'édition. Néanmoins, l'organisation en bien commun a aussi une limite spatio-temporelle. Malgré nos propos simplificateurs précédents, il faut reconnaître que les documents, avant l'ère numérique, sont des objets dont le caractère non rival est partiel. J'emprunte gratuitement quand je veux un document, n'importe quel document, quelle que soit sa date de publication, par exemple à la bibliothèque publique de mon quartier qui est financée par mes impôts. Mais je dois le lire dans un délai contraint et le ramener de façon à laisser à d'autres habitants du quartier la possibilité de le consulter. Je suis donc aussi contraint par l'espace et le temps limités de ma communauté. Le modèle de la bibliothèque comprend une seconde différence importante avec les deux autres dans la gestion de la transaction. Alors que, pour l'édition et la radio-télévision, la valeur informationnelle est créée en amont par la construction de protodocuments (les œuvres, les programmes, les articles) pour être vendue ou pour captiver dans un processus de diffusion, la valeur créée dans le modèle bibliothéconomique l'est en aval, par l'action du lecteur qui recherche les documents et les consulte. L'activité du lecteur est aussi pour lui une contrainte, une dépense d'énergie. En se déplaçant pour emprunter et rendre, pour sélectionner et prendre le document, il coproduit le service avec la bibliothèque.

Le Web entre flot et bibliothèque

Le Web est, pour le moment, une construction hybride. Nous retrouvons sur le Web les trois marchés présentés : la vente de biens avec, par exemple, la vente d'*e-books*, de morceaux de musique, de vidéos, d'articles de journaux ou de revues, ou même d'objets éditoriaux traditionnels, matériels, livres, albums, DVDs ; la vente d'accès avec les abonnements divers ou encore les licences proposées par les éditeurs scientifiques et tous les échanges et mises en commun qui l'accompagnent ; la vente d'attention avec la publicité en ligne, mais aussi les sollicitations continues des internautes entre eux et la surveillance de leur comportement. Malgré bien des discours sur le sujet, il ne semble pas que le Web ait été l'occasion d'ouvrir des marchés radicalement différents dans leurs modalités.

Néanmoins, les positions respectives des marchés et leur structure interne ont été profondément modifiées. Le marché des biens a détaché le document de son support, qui peut se décliner selon différents formats et sur différents terminaux et donc retrouve son caractère non rival bridé par l'exemplaire. L'accès est devenu un enjeu commercial majeur entre les différents opérateurs, et même, pour certains, un nouveau droit du citoyen. Le marché de l'attention a inversé le sens de l'affichage publicitaire, qui est passé de la diffusion à la navigation, et même principalement à la recherche. Le message publicitaire accompagne dorénavant plutôt le cheminement de l'internaute que celui du document. L'activité de l'internaute est devenue essentielle, investissant les trois dimensions par la publication de documents, le partage de contenu et la relation continue.

Alors que la presse avait construit son modèle industriel comme intermédiaire entre l'édition et spectacle, le média Web s'insère lui entre la bibliothèque et le spectacle. À la première, il emprunte l'accumulation des collections, le partage, les outils de recherche et l'activité de l'utilisateur. Il s'inspire aussi du modèle de la radio-télévision et même du téléphone, auxquels il emprunte la structure en réseau, la logique de la transmission, la captation constante de l'attention, l'abonnement au service, la nécessité du terminal et le financement par la publicité. Comme l'un et l'autre, il est gratuit une fois acquittés les coûts de l'accès. Il introduit enfin, comme les précédents médias, une forme de contrainte dans la relation entre le prix et la gestion de l'espace-temps : je navigue sur le Web à ma guise depuis même un terminal mobile et donc presque de n'importe où, mais « j'accepte » d'être surveillé, orienté, géolocalisé et exposé à une publicité ciblée grâce à mes traces. Le média Web est à la fois le plus libre dans sa consommation et celui qui contrôle de plus près les comportements. Le développement du Web concurrence alors doublement les autres modèles, puisqu'il s'appuie sur un accès libre aux documents et qu'il accapare à son profit l'attention captée de l'internaute.

Immédiateté et mémoire

« Année après année, j'ai renoncé à mes abonnements aux journaux papier, mais il m'était difficile de me séparer du *Times*. La qualité était élevée, le papier tombant sur le trottoir rendait un son agréable pour se réveiller le matin, j'aimais la sérendipité de la balade entre les rubriques, et me sentais obligé de payer pour

mon exemplaire papier au moment où les abonnés devenaient une espèce en voie de disparition. Mais, après des années de tergiversations, c'est fini. La préservation de l'environnement aurait pu être un argument déjà suffisant, mais tout simplement ma lecture se fait de plus en plus sur écran (seules résistent la fiction et la poésie). Et nombre de ces lectures viennent du *Times*. Ce qui m'a amené à l'inévitable est d'avoir réalisé un matin blême, lorsque je lisais le journal sur la table de la cuisine, que j'en avais déjà lu beaucoup (la plupart ?) en ligne. Même pour tout le plaisir du toucher et de l'impression, le *Times* sur papier arrive trop tard. Le journal du jour, ce sont les nouvelles de la veille⁶. » Cette nostalgie d'un blogueur pour son exemplaire quotidien du *New York Times* illustre le changement que le Web opère sur les médias plus anciens. Basculant sur un nouvel outil de diffusion, le média change son rapport à l'espace-temps. Les nouvelles sur le Web ont pris l'espace-temps de la radio-télévision, autant du côté de la production des articles, rédigés et diffusés en continu, que de sa consultation à partir d'un terminal à tout moment, tout en gardant le genre du journal : style écrit, articles, rubriques.

Mais la pression du flot n'est pas la seule à opérer, celle de la bibliothéconomie est tout aussi présente. Ainsi, les nouvelles du jour du *New York Times* ne sont plus les seules en ligne sur son site et ne viennent plus écraser celles de la veille. Bien au contraire, c'est maintenant l'ensemble de la production du journal qui est accessible et qu'il faut normaliser et indexer pour que les articles puissent être retrouvés grâce à des recherches simples. De ce point de vue, le journal est devenu une machine à construire en continu des archives précisément indexées. Enfin, le *New York Times* en ligne accueille beaucoup plus de collaborateurs extérieurs que sa version papier. Aux éditorialistes se sont ajoutés des blogueurs hébergés et le courrier des lecteurs est devenu omniprésent grâce à la fonction commentaire.

L'édition, la bibliothèque et la radio-télévision ne sont pas en reste non plus pour utiliser le Web comme un outil de diffusion de leurs produits et services. Et, très rapidement, elles ont été aussi confrontées à sa gestion particulière de l'espace et du temps. Elles ont intégré, *volens nolens*, des services nouveaux et ont rencontré la concurrence de nouveaux venus sur ces mêmes services : collections de livres numériques, archives, streaming, échanges *peer-to-peer*. Dans le même temps, leur fonction de sélection de protocuments ou de documents a été submergée par la facilité de publication ouverte à tous sur la Toile et la sélection documentaire effectuée *a posteriori*, aussi bien pour l'écrit que pour l'image, l'audiovisuel ou la musique.

On pourrait dire que le Web documentarise les médias traditionnels et qu'il commercialise la bibliothèque, les transformant tous en une vaste « industrie de la mémoire ». Les logiques commerciales et non commerciales s'y superposent, se contredisant et s'épaulant tout à la fois. Comme nous le verrons au chapitre suivant, d'un côté, les conseils donnés par Carl Shapiro et Hal Varian dès 1998⁷ sur le verrouillage, les externalités, la concentration ont porté leurs fruits, ouvrant la voie à la constitution de firmes ultra-rentables en situation quasi monopolistique... et menant aussi à quelques désillusions fracassantes au moment de l'éclosion de la « bulle Internet » au tout début du millénaire. De l'autre, les constats et les encouragements prodigués par Yochai Benkler⁸ sur l'apparition d'une économie du troc et de valorisations croisées illustrent la grande vitalité de l'activité des internautes, de sa création de richesse... et aussi de quelques naïvetés et de pas mal de désordre. Ainsi, le Web produit un modèle nouveau de média avec sa logique et son écriture propres, fondé sur l'accès et la transmission, que l'on peut comparer aux anciens modèles. Dans le même temps, il fait évoluer ces derniers en les transformant en industries de la mémoire.

Les cinq industries de la mémoire

Il est possible de reprendre les constatations précédentes sur un tableau⁹ qui positionne les différents modèles industriels de l'économie des documents, illustrant leur continuité, isolant leur identité et soulignant l'impact du Web.

En passant d'une colonne à l'autre, les modèles industriels forment bien une continuité par la relation économique, temporelle et spatiale qu'ils entretiennent avec le lecteur. Mais les modèles se distinguent aussi les uns des autres par des ruptures représentées par les lignes verticales qui séparent les modèles, soit, de gauche à droite : ouverture d'un nouveau marché par la presse, inconnu dans l'édition (annonceurs) ; passage au signal pour la radio-télévision alors que l'on gérait encore des objets matériels dans la presse ; inversion du flux avec le Web où l'internaute est aux commandes ; retour aux objets dans la bibliothèque ; et enfin, pour boucler le tableau entre la dernière colonne de droite et la première de gauche, réinversion du flux entre la bibliothèque et l'édition. Ces ruptures impliquent que chaque modèle a aussi développé des savoir-faire originaux, différents des autres, et expliquent leur autonomie et leur stabilité. Chacun a développé une économie propre.

Tableau 4. Les cinq modèles de l'industrie de la mémoire

	Édition (livres, disques...)	Presse (journaux)	Flot/Broad- casting (radios, télévisions)	Toile/ Webmédia (moteurs, plate- formes)	Biblio- thèque
Temps	Saisonnier	Régulier	Continu	Discontinu	Ponctuel
Produc- tion	Protodocu- ments reproduits : livres, disques, dvd...	Série de pro- todocuments reproduits : quotidiens, hebdoma- daires, men- suels...	Séries de protodocu- ments dif- fusés en continu : variétés, jeux, fic- tions, infor- mation	Documents partagés : texte, vidéos, PodCast, bases de données	Documents mis en commun : livres, jour- naux, disques, dvd...
Accès	Distribu- tion en magasins	Distributions en kiosques, magasins et portage à domicile	Diffusion à domicile : radios, télé- viseurs	Navigation indivi- duelle : ordinateur et mobiles	Consulta- tion et prêt en biblio- thèque
Finance- ment	Paiement par le consom- mateur	Annonces, consomma- teur (pièce, abonnement)	Annonces, consomma- teur (abon- nement redevance)	Annon- ces, insti- tution, donations et consom- mateur	Institution Donations
Web méta- média	E- commerce Plate- formes e-books et musique	Sites journaux avec archives Revue de presse Blogs spécia- lisés	Streaming, vidéos à la carte Gestion numérique	Tout	Biblio- thèques numériques Archives ouvertes

Mais le numérique et le Web ont aussi ouvert des opportunités nouvelles à chacun des modèles par la réingénierie documentaire. Celles-ci sont notées sur la dernière ligne. Ces opportunités sont différentes et spécifiques à chaque modèle, car le Web rencontre chaque fois les logiques particulières qu'il devra adapter. Ainsi, le Web est à la fois l'occasion de bâtir un modèle nouveau qui s'insère entre les anciens et celle de renouveler ces derniers. Ce double mouvement alimente son dynamisme et est aussi la source de bien des malentendus, puisque le Web est à la fois une plateforme ouverte au développement de chacun et l'occasion de l'arrivée d'un nouveau modèle concurrent, un méta média et un média.

L'économie du document

L'économie du document n'est pas cantonnée à celle des documents publiés que nous venons de décrire. La définition du protodocument donnée au deuxième chapitre, « le protodocument est une trace permettant d'interpréter un événement passé à partir d'un contrat de lecture », sous-entend un rôle beaucoup plus large dont les opportunités numériques ont facilité la manipulation. Aujourd'hui, les documents ont envahi et régulent l'ensemble de nos vies qui se construisent beaucoup plus qu'autrefois à partir d'une mémoire inscrite sur les objets. C'est pourquoi il paraît utile de bien circonscrire le secteur de l'économie du document, mais les économistes n'y ont pas encore prêté grande attention. Cette tâche ne peut être celle d'un seul chercheur, aussi je me contenterai ici, pour conclure ce chapitre, de quelques suggestions.

L'année 2009 a vu la publication ou l'actualisation par de grands organismes internationaux de plusieurs guides pour mesurer l'économie de la culture et l'économie de l'information¹⁰. La Commission européenne a publié en 2010 un Livre vert intitulé *Libérer le potentiel des industries culturelles et créatives*. Ces travaux témoignent de la reconnaissance du rôle de la culture et de celui de l'information pour la croissance économique. Je vais m'en inspirer pour présenter le secteur de l'économie du document, en modifiant sensiblement les catégories proposées.

Le secteur culturel et créatif

Des experts européens¹¹, après avoir fait la synthèse des propositions des différents pays et organismes, ont suggéré de regrouper les activités culturelles en un « secteur culturel et créatif », dont les frontières sont précisées dans le tableau ci-dessous.

On peut remarquer tout d'abord le découpage en un « cœur » et trois « cercles ». Le cœur représente les activités fondatrices du secteur : les arts visuels, les arts du spectacle, le patrimoine. Le cercle 1 reprend les industries culturelles dans leur acception classique. Les cercles 2 et 3 présentent ce que la Commission européenne appelle désormais les industries créatives (le design, l'architecture et la publicité) d'une part et les industries connexes d'autre part. La délimitation de ces dernières constituant le cercle 3 est indiquée comme plus floue, car elles appartiennent aussi à d'autres sous-secteurs, les fabricants de matériels d'enregistrement et de lecture et les opérateurs de réseaux, reliés au secteur des technologies de l'information et de la communication.

Le secteur du document

Tableau 5. *Contours du secteur culturel et créatif*⁴²

Cercles	Secteurs	Sous-secteurs	Caractéristiques
Cœur : champs des arts	Arts visuels	Métiers d'art Peinture – Sculpture – Photographie	Activités non industrielles Le produit est un prototype « potentiellement protégé par la propriété intellectuelle » (c'est-à-dire que ces ouvrages comprennent une grande part de création qui pourrait relever du droit d'auteur mais qu'ils ne sont pas systématiquement protégés, comme c'est le cas de la plupart des produits artisanaux, de quelques productions des arts vivants ou visuels, etc.)
	Arts vivants	Théâtre – Danse – Cirque – Festival	
	Patrimoine	Musées – Bibliothèques – Sites archéologiques – Archives	
Cercle 1 : Industries culturelles	Film et vidéo		Activités industrielles fondées sur une reproduction de masse
	Télévision et radio		Produits soumis au droit d'auteur
	Jeux vidéos		
	Musique	Marché de la musique enregistrée – Spectacles musicaux – Revenus des sociétés de collectes de droits dans le secteur musical	
	Livres et journaux	Publication de livres – Publication de journaux et des magazines	

Cercles	Secteurs	Sous-secteurs	Caractéristiques
Cercle 2 : Activités et industries créatives	Design	Design de mode, graphique, d'intérieur, de produit	Les activités ne sont pas toujours industrielles et peuvent être des prototypes
	Architecture		Bien que les produits soient fondés sur le droit d'auteur, ils peuvent inclure d'autres éléments de propriété intellectuelle (droit des marques par exemple)
	Publicité		La créativité (les compétences et les personnes créatives venant du champ des arts ou de celui des industries culturelles) est essentielle aux performances de ces secteurs non culturels
Cercle 3 : Industries associées	Fabricants d'ordinateurs, de lecteurs MP3, industrie des mobiles, etc.		Cette catégorie est floue et impossible à délimiter à partir de critères précis. Elle touche de nombreux autres secteurs économiques qui sont dépendants des « cercles » précédents, comme le secteur des technologies de l'information et des télécommunications

Je vais modifier le tableau des experts européens (tableau 5) pour construire un tableau délimitant le secteur du document (tableau 6). Le cœur représente bien aussi les origines anciennes de l'économie du document, on pourrait dire sa préhistoire. Ces activités restent artisanales. Leurs productions ont les caractéristiques d'un document, mais elles restent limitées au stade prototypaire, protodocumentaire : l'œuvre d'art, le spectacle, le monument ou le document d'archives. Néanmoins, nous avons vu que les bibliothèques ont quitté ces prémices artisanales et aléatoires pour se multiplier en constituant progressivement la « bibliothéconomie », c'est-à-dire une organisation rationnelle, industrielle, d'un service d'accès aux documents. Je les retire donc du cœur, au moins pour celles dont la mission première n'est pas la conservation, pour les placer dans la rubrique suivante, le premier cercle qui réunit sur le tableau 5 les industries culturelles.

Plusieurs corrections ou ajustements doivent de plus être effectués dans ce premier cercle. La musique, tout d'abord, est mise comme une catégorie générale dans les industries culturelles. Cela se justifie en économie de la culture par la cohérence, la complémentarité des différentes activités musicales et leur articulation. Mais, d'un point de vue documentaire, les concerts, la musique vivante relèvent du spectacle, c'est-à-dire de la catégorie précédente, tandis que la musique enregistrée est un objet documentaire à part entière et participe clairement à la catégorie industrielle. Celle-ci comprend aussi les bases de données, qui n'apparaissent pas dans le tableau de la Commission. Enfin et surtout, à part les jeux vidéos, les nouveaux industriels du Web documentaire sont absents de ce premier cercle : ni les portails, ni les moteurs, ni les réseaux sociaux ne sont mentionnés. Nous verrons pourtant au chapitre suivant qu'ils jouent un rôle crucial aujourd'hui pour l'économie du document. Je les ajoute donc dans cette rubrique sous l'appellation de Webmédia. L'ensemble a été nommé par les experts de la Commission européenne « industries

culturelles ». J'en ai modifié les contours, il faut donc en changer le nom. Conformément à mon propos, je les ai nommées « industries de la mémoire ». On y retrouve naturellement les trois modèles génériques et leurs déclinaisons présentés dans la première partie de ce chapitre.

Dans le deuxième cercle, seule parmi les éléments initiaux des experts de la CEE, la publicité concerne l'économie du document. Mais son importance relève moins de sa créativité que de l'ouverture d'un second marché, celui des annonceurs, ou plus précisément la construction d'un marché biface, comme nous l'avons vu. La publicité participe donc aussi aux industries de la mémoire, mais d'une façon décalée. Son activité est articulée à celle de la presse, de la radio-télévision et, depuis peu, du Webmédia. Le reste du deuxième cercle, le secteur créatif, ne fait pas vraiment sens pour l'économie du document, je le supprime.

En revanche, d'autres activités ne relevant pas directement des industries de la mémoire reposent sur une création, manipulation, distribution de documents. L'éducation transmet des savoirs par l'intermédiaire de professeurs, mais aussi d'une intense activité documentaire. Les administrations, privées et publiques, fonctionnent avec des documents. Toute une série de professions, notaires, avocats, agents divers, sont là pour rédiger et certifier l'authenticité de documents, agissant comme tiers validateurs. Ces trois domaines se sont transformés progressivement avec l'arrivée des moyens légers de production et reproduction documentaires : machines à écrire, reprographie, photocopieuses. Ils sont bousculés aujourd'hui par la bureautique et les facilités offertes par le traitement, les mémoires et les réseaux numériques, sans pour autant être remis en cause. Ils forment un secteur que l'on appelle de plus en plus la « gestion des connaissances », plus connue sous sa dénomination anglaise : le « *knowledge management* ». L'économie de ce secteur est difficile à circonscrire et mesurer, mais elle n'en est pas moins essentielle. C'est pourquoi j'ai remplacé le deuxième cercle précédent des « industries créatives » du tableau 5 par celui de la « gestion des connaissances » dans le tableau 6.

Le troisième et dernier cercle enfin est succinctement évoqué comme « industries associées ». Il agrège les fabricants de matériels, de logiciels et les opérateurs de télécommunication et comprend un « etc. » qui laisse la porte grande ouverte. Il est indiqué en remarque qu'il s'agit notamment du secteur des technologies de l'information et de la communication (TIC). Pour l'économie des documents, ces acteurs sont essentiels, car ce sont eux qui supportent, mettent en forme les documents et les rendent accessibles, ce sont les industries du contenant. Mais leur dynamique dépasse très largement ce rattachement, car ils concernent aussi d'autres activités et s'adressent à d'autres marchés comme le e-business, le e-commerce, ou simplement le téléphone. Le secteur des TIC comprend aussi bien l'infrastructure, c'est-à-dire essentiellement les réseaux informatiques et de télécommunications, que les produits, biens et services, depuis le matériel informatique et électronique, les composants, les logiciels et les services reliés, y compris les services de télécommunications.

Tableau 6. *Le secteur documentaire*

		Branches
Protodocuments		Art visuel Spectacle Patrimoine
Cercle 1 : Industries de la mémoire	Biens	Livre, imprimé Musique enregistrée Film, vidéo
	Attention	Presse, Radio-télévision Web-média
	Accès	Bases de données Bibliothèque
Cercle 2 : Gestion des connaissances		Éducation Administration Tiers
Cercle 3 : Industrie du contenant		Équipements électroniques Télécommunications Informatique

Publicité

Toutes ces remarques permettent maintenant de présenter le secteur de l'économie du document sur un tableau qui s'inspire du tableau de l'Union européenne précédent, en l'amendant et le simplifiant. Ce nouveau tableau mériterait discussion et validation, mais, tel quel, il permet déjà de percevoir l'importance du secteur dans l'ensemble de l'économie et suggère sa mesure.

Caractéristiques économiques du document

Les économistes de la culture et ceux de l'information s'accordent pour souligner la particularité des caractéristiques des biens de ces secteurs. Ils parlent de non-rivalité (la consommation par un individu ne prive pas un autre individu), de non-excluabilité (impossibilité de réserver la consommation à certains), de biens d'expérience (biens non connus avant d'être consommés), d'externalités (influences indirectes d'une action), mais ces caractéristiques sont souvent présentées au fil de la plume et convoquées de façon opportuniste pour conforter une partie du raisonnement. Il me semble que l'analyse tridimensionnelle du document autorise un classement plus précis et plus explicatif. Sans régler définitivement la question, on trouvera ci-dessous quelques suggestions rapides à partir des éléments développés précédemment :

- Selon sa première dimension, un document est une promesse. Sans connaître son contenu, nous savons déjà que c'est un document et nous pressentons qu'il sera pour nous utile, intéressant, distrayant ou au contraire sans attrait. C'est la définition même du bien d'expérience en économie. Cette dimension qui privilégie l'objet permet de souligner une autre caractéristique très importante pour son économie, même si elle est souvent négligée ou mal comprise par les économistes : sa plasticité. À les lire, on a parfois l'impression qu'ils considèrent l'information comme insaisissable, un gaz, un liquide, un fluide. Cette présentation est trompeuse. Un document est un objet fini, constitué d'éléments finis, mais il est portable, transférable. Il peut être facilement découpé, réagencé pour devenir autre.

- De la deuxième dimension, celle du contenu ou du texte, découle d'abord la variabilité de la valeur d'un document, qui dépend de son interprétation, par définition changeante selon les contextes. Lire un document en épuise la valeur pour son consommateur. Autre facteur important pour l'économie, le contenu d'un document est singulier. Deux contenus identiques n'en forment qu'un seul. Le contenu a la potentialité d'un bien non rival et non excluable.

– Enfin, la troisième dimension, celle du médium, qui insiste sur sa portabilité dans le temps et dans l'espace, suggère de fortes externalités. Les documents sont liés les uns aux autres par des citations, des références ou de simples allusions, chacun est une sorte de tête de réseau. Les individus ont leurs collègues, partenaires, amis, connaissances, avec qui ils échangent. Les individus sont aussi des têtes de réseau. Les documents concernent ou intéressent les individus qui prennent connaissance des documents. Des effets de résonance se mettent alors en place, concentrant l'attention, dont nous avons vu les limites, sur un petit nombre de documents.

Au travers des propositions de ce chapitre, j'espère avoir convaincu le lecteur de l'intérêt de poser les fondations d'une économie du document. L'étude de celle-ci serait d'autant plus souhaitable que l'appétit des nouveaux industriels du Web fait bouger très rapidement l'ensemble de notre système documentaire, non sans conséquences sur notre rapport aux savoirs.

Notes du chapitre 4

1. Pour une première entrée, on peut commencer par ce petit livre : Michel PERONA et Jérôme POUYET, *Le Prix unique du livre à l'heure du numérique*, Centre pour la recherche économique et ses applications, Éditions rue d'Ulm, Paris, 2010. Consultable sur : <www.cepremap.ens.fr>.
2. Jean-Yves MOLLIER, « Naissance, développement et mutations de l'édition de l'*Encyclopédie* de Diderot à Internet », *loc. cit.*
3. ASSOCIATION VECAM, *Libres Savoirs. Les biens communs de la connaissance*, C&F Éditions, Caen, 2011.
4. Thomas PETITT, *Before the Gutenberg Parenthesis*, *op. cit.*
5. Herbert SIMON, *Les Sciences de l'artificiel*, Gallimard, Paris, 2004.
6. Jimmy GUTERMAN, « Goodbye, *New York Times* », *O'Reilly Radar*, 24 mars 2008. <radar.oreilly.com> (traduction personnelle).
7. Carl SHAPIRO et Hal R. VARIAN, *Information Rules, A Strategic Guide to the Network Economy*, Harvard Business School Press, Harvard, 1998. Site compagnon <www.inforules.com>. Hal Varian est aujourd'hui économiste en chef chez Google.
8. Yochai BENKLER, *La Richesse des réseaux*, *op. cit.*
9. Ce tableau est une adaptation et un élargissement de celui présenté bien avant l'apparition du Web dans : Bernard MIÈGE, Patrick PAJON et Jean-Michel SALAÜN, *L'Industrialisation de l'audiovisuel, des programmes pour les nouveaux médias*, Aubier, Paris, 1986.
10. INSTITUT STATISTIQUE DE L'UNESCO, *Cadre de l'Unesco pour les statistiques culturelles 2009*, Unesco, 2009 ; UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT, *Manual for the Production of Statistics on the Information Economy*, UNCTAD, 2009 ; ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES, *Guide to Measuring the Information Society*, 2009, OCDE, 2009.
11. KEA EUROPEAN AFFAIRS, *The Economy of Culture in Europe*, DGEC – Commission des Communautés européennes, octobre 2006. Consultable sur <<http://ec.europa.eu>>.
12. *Ibid.*, p. 56 (traduction personnelle).

À la recherche du néodocument

L'Internet et le Web, le réseau des réseaux et la Toile, désignent une infrastructure interconnectée. De nombreux observateurs du Web mettent alors en avant le réseau et l'organisation égalitaire de la liaison des ordinateurs. Cette insistance minimise une autre caractéristique aussi importante : l'augmentation considérable des capacités de mémoire et de traitement des traces par le calcul. L'Internet et le Web sont aussi et surtout des machines à produire et diffuser des documents, et leur succès accompagne, relaie et entretient le renouvellement documentaire. Les changements sont tels que l'on peut s'interroger sur l'objet lui-même. Quel est ce néodocument qui naît et s'échange sur le Web ? Est-il comparable au document du xx^e siècle ? Ou, en reprenant notre définition du document : de quelles traces s'agit-il ? De quels événements ? Comment les interprétons-nous ? Avec quel contrat de lecture renouvelé ?

Pour répondre pertinemment aux questions, il faudrait les lumières d'un ensemble coordonné d'experts de différentes disciplines de sciences humaines et sociales et de plusieurs spécialités de l'informatique et des télécommunications. Une telle ambition dépasse de très loin celle de ce livre. Pour ce dernier chapitre, je suggérerai simplement quelques pistes suivant toujours mon fil d'Ariane, l'histoire longue du document et son analyse tridimensionnelle.

Les paradoxes du néodocument

Pour pouvoir se définir et s'affirmer, le néodocument doit résoudre plusieurs paradoxes anciens, mais exacerbés par les nouvelles opportunités. Les solutions se cherchent parfois fiévreusement et le compromis entre les éléments contraires n'est pas abouti. Il en résulte un déséquilibre, prétexte de blocages ou au contraire occasion d'inventions, que l'on peut exprimer sous forme de paradoxes.

Propriété et partage

Le premier paradoxe est celui de la propriété et du partage. Un document numérique a la vertu ou le défaut d'être infiniment échangeable sans dégradation, il a les potentialités d'un « bien public », un bien auquel tout le monde a accès librement sans priver personne, des potentialités que le Web peut rendre effectives. Cette caractéristique contredit la propriété intellectuelle, introduite de façon réglementaire pour construire une rivalité du document et permettre l'économie de sa création et de sa publication. Un dialogue de sourds s'est alors installé, attisé par les intérêts commerciaux des éditeurs et producteurs d'un côté et ceux des opérateurs de réseaux et fournisseurs de services Web de l'autre. On comprend que la proposition des promoteurs du Web de rendre l'ensemble des documents librement accessibles, souvent présentée comme un destin et non comme un choix, puisse être considérée par les ayants droit comme une régression et non un progrès. Mais ces derniers ont, à leur tour, le défaut de croire à la naturalité de ce qui n'est qu'une construction juridique à la portée limitée. Au-delà d'une période qui a varié selon les époques et les pays, les droits patrimoniaux tombent et l'œuvre passe dans le domaine public. Par ailleurs, bien des exceptions au droit d'auteur sont prévues pour que son application ne bride pas l'accès à l'information. Droit d'auteur et droit à l'information sont tous deux des droits fondamentaux de l'homme, mais leur application pratique est contingente aux situations et l'un ne saurait occulter l'autre.

Lawrence Lessig, en inventant les *creatives commons*¹, suggère aux auteurs d'ouvrir volontairement les portes de leur propriété intellectuelle en autorisant, sous certaines conditions, la consultation et la diffusion gratuites de leur œuvre. Ce faisant, sans bouleverser le principe du droit d'auteur, il subvertit la pièce essentielle du modèle de l'édition, les droits patrimoniaux, qui créait la non-rivalité des exemplaires, et

illustre par le partage la filiation du Web et de la bibliothèque. Il contredit aussi la prétention des éditeurs à faire, de leur modèle, le modèle fondateur de la publication du document et conteste leur tentation de se réserver des rentes de situation par l'extension quasi infinie des droits patrimoniaux. « Aujourd'hui, les lois sur le droit d'auteur régulent beaucoup trop la culture. Leur champ d'action est trop large. [...] Avant, si vous n'aviez jamais touché à une photocopieuse, vous n'aviez jamais enfreint le droit d'auteur. Aujourd'hui, nous vivons dans un monde où chaque geste soulève le problème du droit d'auteur² », remarque Lessig.

Même sans référence à la propriété intellectuelle, l'auteur romantique décrit par Jean-Claude Guédon³ perd de son aura sur le Web. Le personnage principal y est moins celui qui crée que celui qui partage, le passeur qui dépose ou signale des documents si possible en les commentant. On pourrait écrire avec un clin d'œil que, de l'édition au Web, le héros, l'auteur inséparable de son œuvre, laisse la première place au héraut, celui qui repère, relaie, commente pour une communauté de lecteurs les documents d'une bibliothèque numérique trop vaste et changeante pour être accessible à tous. Pour autant, ce héraut-là n'est pas moins chargé de romantisme, valorisé, applaudi ou contesté, parfois même aussi ténébreux, solitaire et désargenté que le précédent. Le plus sulfureux est celui qui met en ligne des protodocuments que l'ancien ordre tenait pour confidentiels ou indécents. Le plus aventureux, celui qui témoigne directement d'une situation occultée dans l'ancien ordre. Il y a du journaliste dans ce héraut-là et Dan Gillmor a même un temps revendiqué cette filiation en proclamant : « *We the media*⁴ ». Même quand il appartient à une organisation qui le mandate, l'individu prime sur l'institution. « Je » est un document assumé, pourrait-on ajouter.

Son influence est mesurée à la notoriété de son nom ou pseudonyme et au nombre de personnes qui le suivent personnellement sur son blog, sur Twitter, sur les réseaux sociaux et éventuellement le relaient à leur tour. Son autorité n'est plus transcendante comme dans l'ordre ancien, mais réside dans la fiabilité de son travail, dans l'intérêt des documents qu'il relaie et commente pour sa communauté conformément aux valeurs de l'ordre nouveau. Sa rémunération est, sinon plus incertaine, du moins pas (encore) réglementée comme celle de l'auteur par la propriété intellectuelle. Le héraut qui voudrait vivre de son travail n'a que deux solutions, non exclusives l'une de l'autre : soit il réussit à monnayer sa notoriété auprès d'annonceurs ou d'autres clients ; soit il est rémunéré directement ou indirectement par sa communauté ou par un média traditionnel. Cette fragilité économique rend incertaine la pérennité de son rôle. Et nombreux sont les hérauts qui s'essouffent, remplacés par d'autres attirés par l'audience et le dynamisme du Web, le renouvellement continu des outils, l'efficacité de la veille et la possibilité d'une notoriété.

Sans remonter au héraut du Moyen Âge, l'importance des relais dans la circulation de l'information n'est pas vraiment nouvelle. Ils ont été étudiés par la sociologie américaine des médias au lendemain de la Seconde Guerre mondiale⁵, qui a mis en évidence le processus du *two-step flow of communication*, une circulation de l'information à deux étages, pour analyser l'influence des médias sur le vote. Les médias touchent d'abord une série de personnes plus concernées, des leaders d'opinion, qui ensuite influent les choix de leur communauté. Mais, aujourd'hui, le relais est intégré au média, il passe directement sur le réseau et laisse sa trace. Dès lors, le leader d'opinion sur le Web peut être pour sa communauté autant un héraut, un haut-parleur dont chaque nouvelle alerte efface la précédente, qu'un bibliothécaire ou un documentaliste qui accumule un patrimoine documentaire. Le Web, en réduisant l'emprise de la propriété intellectuelle sur le document, fait émerger de nouvelles figures de médiateurs plus proches de la postmodernité évoquée au chapitre 3. Ils reflètent, à leur échelle individuelle, le positionnement du Web entre flot et bibliothèque constaté au chapitre 4.

Lecture et calcul

Le deuxième paradoxe est très différent, plus abstrait. Il pourrait se formuler ainsi : le Web facilite conjointement la lecture linéaire de documents par le plus grand nombre de personnes et leur traitement sous forme de données discrètes par le plus grand nombre de machines. Un tel paradoxe peut se traduire en termes anthropologiques et philosophiques. Ainsi Jack Goody a montré combien l'arrivée de l'écriture avait transformé nos sociétés et nos façons de concevoir le monde par la « raison graphique⁶ ». Bruno Bachimont a proposé le terme de « raison computationnelle⁷ » pour signifier une bascule comparable, induite par les modalités nouvelles de représentation calculées par ordinateur. Clarisse Herrenschmidt, de son côté, parle de trois écritures (langue, nombre, code)⁸. Stephen Wolfram prétend qu'une science nouvelle serait possible où les processus naturels et artificiels pourraient être représentés par des programmes simples⁹. Il ne manque pas de penseurs pour relever et tenter de dépasser ce paradoxe, non sans susciter de multiples controverses. Sans aller aussi loin, je noterai simplement quelques conséquences des capacités de calcul à grande échelle sur le Réseau.

La réponse immédiate à nos questions posées sur le Web depuis n'importe quel point connecté est un des changements les plus spectaculaires de notre ordre documentaire. Derrière le projet de « Web des données » du W3C se cache, nous l'avons vu, une conception de la représentation du monde, un rapport différent à la vérité. Les documents peuvent être retrouvés ou produits à la demande à partir d'un calcul : des unités documentaires élémentaires normalisées réunies dans des bases de données interconnectées (les traces, vu) sont agencées grâce à des algorithmes (l'interprétation amont, lu), sous l'impulsion des internautes (l'interprétation aval, su). Déjà, l'invention des bien nommés « moteurs de recherche » relevait du même type de raisonnement en l'appliquant à l'analyse de textes.

Christian Vandendorpe voit dans ces changements de représentation du document une des causes du développement d'une lecture fragmentée : « Le véritable supplément à la mémoire n'est plus le document ni même la bibliothèque, mais Google. C'est vers lui et des outils comparables que l'on se tourne désormais non seulement pour interroger les milliards de pages d'informations disponibles sur le Web, mais aussi pour retrouver un élément précis dans nos archives personnelles¹⁰. » Ainsi, le contrat de lecture se transforme avec la facilité d'accès aux documents grâce aux calculs du moteur et peut-être plus radicalement encore par la reconstruction en temps réel de documents à la demande envisagée par le W3C.

Mais la lecture n'est pas la seule à être transformée par le calcul sur les documents. Les modes de représentation sont aussi directement touchés, comme l'avait déjà noté Pédaque : « Les représentations graphiques (tableaux, schémas, courbes, cartes, maquettes, etc.), les représentations analogiques (images, photographies) ou encore les représentations symboliques (formules mathématiques, chimiques ou génétiques, etc.) deviennent les supports de calculs immédiats ouvrant la porte à des pans entiers de connaissance ou d'expression inaccessibles jusqu'ici faute d'outil formel adéquat pour les exprimer¹¹. » Bien entendu, ces nouvelles représentations peuvent suggérer des idées nouvelles tout comme de fausses interprétations, ou simplement dessiner des formes plaisantes sans réelle signification. Ainsi, par le calcul, l'ordre documentaire nouveau non seulement fournit des réponses immédiates à toutes nos questions courantes, mais autorise des représentations inédites de phénomènes.

Conversation et traces

Les néodocuments sont pris encore dans un troisième paradoxe, issu de deux facultés du Web : l'immédiateté et l'enregistrement. Le Web favorise conjointement le développement d'échanges spontanés (conversations, commentaires, échanges de messages, photos, vidéos) entre individus et leur fixation automatique sur un support public, relativement pérenne et documenté. Le Web rend doublement caduque l'adage « les paroles s'envolent, les écrits restent ». Les paroles ne s'envolent plus, elles sont continuellement enregistrées ; les enregistrements, les « écrits », se noient dans le flot continu de leur renouvellement sans fin.

La plus spectaculaire expression de ce paradoxe est le développement des services du Web 2.0 à partir du milieu des années 2000. Grâce à ses outils conviviaux de publication sur des plateformes de partage, le Web 2.0 a multiplié les occasions d'échange de protocuments (photos, vidéos amateurs, lettres, collages, carnets, signets, etc.). Ces protocuments, autrefois réservés à une diffusion restreinte dans un cercle d'intimes, sont devenus de fait des documents, des traces interprétables par des internautes auxquels ils n'étaient souvent pas *a priori* destinés. On n'écrit plus sur un ordinateur connecté au Web 2.0 comme on écrivait autrefois sur du papier. On écrit « dans le ciel à la vitesse de la pensée », selon l'expression de Stevan Harnad¹². Non seulement les protocuments passent les frontières du cercle des intimes, mais ils sont enregistrés et répertoriés automatiquement, au point qu'il est parfois difficile de les détruire, car il reste une copie dans une mémoire cache d'un centre de données. Le Web 2.0 fonctionne comme une machine à fabriquer des archives publiques en continu, de façon sauvage et désordonnée.

Pédaque avait pointé en 2006 toute une série de décadrages induits par l'opposition entre immédiateté et pérennité, aussi bien dans les sphères domestique, professionnelle ou publique, depuis les journaux intimes transformés en blogs, les photos amateurs devenues témoignages, les activités professionnelles surdocumentées jusqu'à un darwinisme documentaire dans l'espace public par la sélection *a posteriori* des documents sous l'effet de l'interaction des internautes. Ces constats sont toujours valables cinq ans plus tard malgré la rapidité des changements. Loin d'une mise en ordre, chaque innovation, chaque initiative, a tendance à disperser les échanges plutôt qu'à rationaliser les traces. À l'évidence, le média n'est pas encore arrivé à maturité.

Néanmoins, les internautes s'habituent peu à peu à la nouveauté et prennent la mesure de ses difficultés et richesses. Progressivement, de nouvelles normes sociales se mettent en place. Les autorités tentent aussi, non sans tâtonnements, de réguler les flux. On peut le repérer à une meilleure maîtrise par les individus des réseaux selon leur fonction, à la spécialisation des outils selon les types de communication, à la gestion

plus circonspecte des identités numériques, ou encore à l'évolution des styles et des propos dans les discussions sur le Web. Il me semble même que l'on voit poindre ici et là une sorte d'éloquence écrite du spontané, précieuse, savante ou familière, qui rappelle l'improvisation théâtrale ou les duels oratoires.

Reste que le Web, malgré sa fonction d'enregistrement continu, est loin de permettre de décrypter simplement les mouvements de la société. Et je crois qu'il faut prendre au sérieux la remarque d'Éric Schmidt, ancien responsable de Google, qui suggérait sur ce paradoxe un débat sociétal : « Je ne crois pas que la société comprenne ce qui arrive quand tout est accessible, potentiellement connu et enregistré par tout le monde tout le temps¹³. »

Mémoire et oubli

Le dernier paradoxe sur lequel je voudrais insister est celui de la mémoire et de l'oubli. Le Web est une formidable machine mémorielle, une prothèse de notre mémoire et, en même temps, il a une forte capacité à perdre l'information qu'il enregistre et transporte. Cette perte a plusieurs causes. La première est mécanique, car les supports d'enregistrement numérique n'ont pas encore fait leurs preuves sur le long terme. La deuxième est plus systémique et est bien représentée sur le schéma de Bruno Bachimont reproduit au deuxième chapitre (figure 2). Un document numérique est toujours réinventé à partir de données, d'un logiciel de traitement et de terminaux. Il a donc plusieurs variantes possibles, ce qui complique sérieusement sa conservation. Ensuite, le Web est un média de flux. Les sites s'actualisent en continu et leur enregistrement est une gageure puisque, par principe, un enregistrement est statique alors que l'objet est dynamique. Il est même impossible de faire une « photographie » du Web à un moment donné puisque les sites se seront modifiés sans synchronisme durant le temps de l'enregistrement. Dans le même temps, les enregistrements de fichiers s'accumulent sur les serveurs et les terminaux des internautes. Et des documents réapparaissent de façon parfois aléatoire et intempestive. Chacun peut mesurer sur sa machine l'ampleur de ce paradoxe.

Les archivistes se trouvent face à des défis inédits. Il leur faudrait intervenir en amont et non plus en aval, puisque la collecte de documents n'existe plus, sans pour autant perturber le flux documentaire. De nombreux travaux sont menés sur la préservation des documents numériques et du Web pour retrouver les facultés de conservation des anciens supports. Mais, du fait de l'apparition d'un néodocument, il y aurait un chantier à ouvrir sur ce que signifie la fonction de mémoire externe pour un individu, une collectivité, une société. Certains, poussant au bout la logique de l'homme-document, utilisent déjà des capteurs pour enregistrer et analyser l'ensemble de leur vie quotidienne dans ses moindres détails¹⁴.

L'élaboration d'un modèle commercial

Ces paradoxes s'illustrent dans un foisonnement d'initiatives dispersées repérables à tout moment sur le Web. Mais, lorsqu'on se tourne vers les entreprises commerciales, ce qui frappe c'est au contraire l'extrême concentration des acteurs sur le réseau. Carl Shapiro et Hal Varian avaient pronostiqué que le gagnant prendrait tout sur le Web du fait des externalités de réseau¹⁵. La suite leur a donné raison. En général, les analystes des stratégies des entreprises sur le Web se contentent d'une approche classique d'économie industrielle, opposant par exemple la stratégie d'intégration verticale d'Apple avec celle d'intégration horizontale de Google. Sans contredire ces analyses, la théorie du document amène un éclairage différent, complémentaire, permettant de mieux comprendre les motivations de ces stratégies et leurs conséquences sur notre rapport aux connaissances.

Je ne saurais ici proposer une analyse de toutes les stratégies commerciales, ni de leurs différents étages, ni de la diversité des secteurs concernés, ni des contextes nationaux contrastés, ni du jeu croisé de l'ensemble des firmes. Je réserverai mes remarques à trois des plus importantes d'entre elles : Apple, Google et Facebook, dont les initiatives concernent au premier chef l'émergence d'un néodocument. Chacune à sa manière tente de résoudre un ou plusieurs des paradoxes cités en s'appuyant chaque fois sur une des dimensions du document. Ainsi, une curieuse concurrence les oppose, à la fois brutale et frontale, par ses appétits sur le même objet, et décalée, par sa manifestation sur une seule des dimensions. Chacune a une ambition monopolistique sur le néodocument, qui ne s'exerce réellement que sur une seule dimension... pour le moment.

La stratégie de la forme : Apple

Plusieurs annonces récentes ont confirmé l'insolente santé financière de la firme Apple. Satisfait des résultats du deuxième trimestre 2011, son président-directeur général, Steve Jobs, déclarait dans une de ses dernières interventions publiques¹⁶ : « Nous sommes ravis de présenter notre meilleur trimestre depuis la création de la firme, avec des revenus en hausse de 82 % et des bénéfices en hausse de 125 %¹⁷. » À l'origine fabricant d'ordinateurs, concurrent en difficulté de Microsoft sur la bureautique au début des années 1990 malgré de fortes compétences en design, Apple est aujourd'hui la deuxième capitalisation boursière mondiale (passée devant Microsoft en janvier 2011). Elle dispose, au moment où j'écris ces lignes, de 75 milliards de dollars de liquidités.

Son succès peut s'analyser à partir de la grille tridimensionnelle du document. Utilisant son savoir-faire sur le design, Apple a construit son avantage concurrentiel sur la première dimension du document, la forme. Pour le document numérique, la forme se dessine par l'articulation de plusieurs éléments : le design du terminal, l'affichage du fichier et, le cas échéant, le rendu sonore. Le génie d'Apple est d'avoir perçu la valeur de la forme pour les documents numériques et d'avoir su répondre à l'attente des usagers dans ce domaine. La répartition de son chiffre d'affaires depuis l'engagement de la firme sur le Web introduira utilement mon propos.

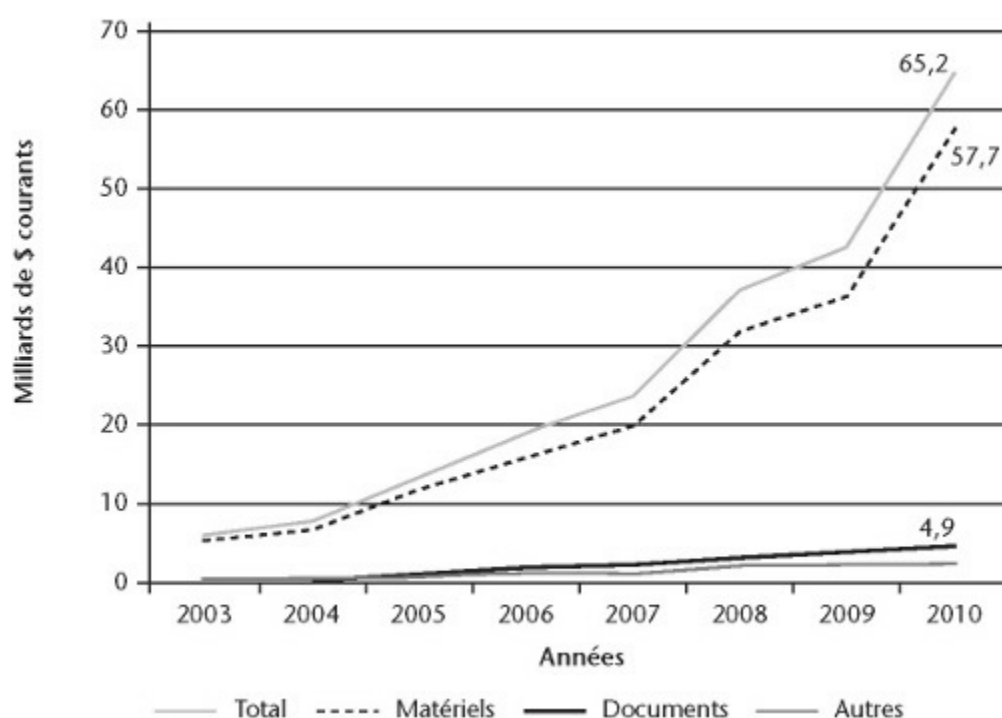


Figure 7. Évolution du chiffre d'affaires annuel d'Apple (2003-2010)¹⁸

On peut tirer deux constats de ce graphique. Tout d'abord, la vente de documents, quasi nulle en 2003 (36 millions de dollars, début de iTunes), avoisine aujourd'hui les 5 milliards et concerne aussi bien la musique que l'écrit ou l'image. Pour important que soit ce chiffre, il ne pèse pourtant pas lourd devant celui de la vente de machines, qui explique à lui seul l'orientation spectaculaire de la courbe. Parti de 5,5 milliards en 2003, le chiffre d'affaires du matériel est multiplié par dix en 2010 avec les succès de l'iPod, puis de l'iPhone et enfin de l'iPad. Poursuivant sa tradition de fabricant, l'intérêt d'Apple pour les documents du Web est motivé d'abord par la vente des terminaux. La firme s'assure un monopole sur l'accès aux documents numériques, soit en les distribuant lui-même, soit en gardant l'exclusivité des applications de navigation sur le Web proposées par des développeurs extérieurs, pour favoriser l'achat de son matériel. Ainsi, un cercle économique vertueux s'engage entre documents et terminaux sous la marque Apple.

L'iPod, l'iPhone et maintenant l'iPad sont, comme le livre, des promesses. Tout comme le livre, ils sont maniables, transportables, appropriables par l'individu. Ces caractéristiques, qui ont fait la supériorité du *codex* sur le *volumen* au début de notre ère, soulignent aujourd'hui celle de ces nouveaux objets nomades (*smartphones*, tablettes) sur les micro-ordinateurs pour la consommation courante de documents

numériques. Les documents numériques ont conquis le quotidien des populations, il faut des outils simples et intuitifs pour y accéder. Les promesses sont d'autant plus importantes que le Web est une caverne d'Ali Baba dont l'internaute a entr'ouvert la porte et dont la profondeur semble infinie. Dès lors, celui qui détient le sésame du contrat de lecture de cette première dimension, les règles plus ou moins explicites du repérage des documents, se trouve dans une position privilégiée, comme l'était l'imprimeur-libraire au XVIII^e siècle quand le livre imprimé a trouvé un marché de masse.

Il est naturel que les internautes les plus avertis et les plus cultivés aient été les premiers séduits. Souvent, ils étaient déjà familiers de l'univers bureautique d'Apple. Ils mesurent la valeur du contenu du Web et sont sensibles par leur éducation à l'esthétique proposée. Ces premiers acheteurs ont entraîné les autres. Le soin mis au graphisme, au tactile et à l'image animée captive aussi les plus jeunes générations. L'ergonomie, la maniabilité des récents outils d'Apple pourraient conduire au premier ébranlement sérieux des fondations de la citadelle livre s'ils sont adoptés à l'école, car le meilleur outil de marketing du livre est d'abord l'école obligatoire où l'on apprend à lire, à écrire, à compter et à accéder au savoir sur des codex.

Au cours de l'histoire du livre imprimé, la valeur économique est passée progressivement de l'imprimeur-libraire à l'éditeur, c'est-à-dire de la dimension forme à la dimension texte, au fur et à mesure que les savoir-faire d'impression et de mise en page se sont stabilisés et que leurs coûts ont été internalisés dans l'ensemble de la filière. Pour suivre l'analogie, on pourrait dire qu'Apple a un temps d'avance sur l'impression et la reliure (le terminal et son administration) et a réussi à affermer une armée de typographes indépendants (les producteurs d'applications). Mais l'évolution n'est pas terminée. Sur le moyen terme, il n'est pas sûr que le nouveau contrat de lecture formel qui s'invente aujourd'hui sur le numérique, et qui progressivement s'internalisera, puisse être approprié par une seule firme, même protégée par toutes sortes de brevets. La forme des documents est une dimension trop essentielle de notre relation au savoir pour être confisquée par un seul joueur. La stratégie d'Apple n'est gagnante que tant que l'innovation interdit à la concurrence de prendre place.

Pour la suite, tout dépendra donc de la position de la firme sur le marché des documents numériques qu'Apple tente de conquérir et verrouiller. Cette stratégie poursuit, en l'adaptant, l'ordre documentaire antérieur dominé par l'édition qui vend des objets et protège le contenu par la propriété intellectuelle. Les discussions avec les éditeurs sont délicates. Ceux-ci craignent à juste titre les conséquences sur leur marge d'un monopole de la distribution. Mais ces négociations difficiles ne remettent pas, pour le moment, en cause leur modèle. À sa façon, Apple s'applique à résoudre le premier des paradoxes notés dans la partie précédente sur la propriété et le partage. En devenant distributeur de documents, il préserve la propriété intellectuelle. Mais, en ouvrant ses terminaux largement sur le Web, il autorise le partage. Tout cela dans l'intérêt bien compris de la firme.

Il existe pourtant une différence fondamentale entre le modèle de l'édition et celui que met en place Apple pour la consommation de biens culturels. Robert Stallman l'a bien souligné¹⁹, même s'il s'exprimait sur son concurrent Amazon. Aucune de ces caractéristiques du produit de l'édition, le livre, n'est maintenue :

- « On peut l'acheter de façon anonyme en payant comptant.
- Il vous appartient.
- On n'est pas obligé de signer une licence qui en restreint l'usage.
- Son format est connu, aucune technologie propriétaire n'est nécessaire pour le lire.
- On peut le donner, le prêter ou le vendre à quelqu'un d'autre.
- On peut concrètement numériser et copier le livre, et c'est même parfois autorisé malgré la propriété intellectuelle.
- Personne n'a le pouvoir de détruire votre livre. »

Nous ne sommes plus dans le modèle de l'édition, mais bien dans celui du Web média fondé sur le service, dont les caractéristiques sont un accès large et immédiat... sous contrôle.

La stratégie du texte : Google

Google pour sa part a construit sa fortune en privilégiant la deuxième dimension du document, le texte. Recherche et extraction d'informations ainsi qu'ingénierie linguistique sont ses compétences de base mises au service de la recherche de documents. Il les a appliquées directement sur le Web en le considérant comme un vaste texte, organisé par les liens entre les documents et la demande des internautes. Ici, la forme du document est secondaire, à la limite même gênante si elle entrave l'intense « lecture industrielle²⁰ » effectuée en continu par des robots au service du moteur de recherche. Bien d'autres éléments soulignent l'accent mis sur cette deuxième dimension : la mise en mémoire cache de l'ensemble du Web dans d'immenses centres de données ; la mise en avant du *fair-use*²¹ pour échapper au

copyright aux États-Unis ; la présentation d'extraits des documents sur les pages de résultats ; la « recherche instantanée » qui suggère aujourd'hui le texte de la requête au fur et à mesure de la frappe de l'internaute ; et j'en passe beaucoup.

En repérant à notre place les documents par leur contenu, Google applique un contrat de lecture d'un autre type que celui d'Apple, un contrat sur l'interprétation par le déchiffrement du document. Ses règles sont fixées par un algorithme qui dépasse maintenant largement le simple calcul du *Pagerank*. L'algorithme est continuellement affiné au sein de la firme dans une alchimie secrète à partir de l'observation statistique des requêtes des internautes et de leur satisfaction. Il fait l'objet de nombreuses spéculations de la part de tous ceux qui recherchent la visibilité sur le Web et donne lieu à des stratégies variées pour faire remonter tel ou tel site dans le classement. Ainsi, le contrat de lecture de Google est une sorte de négociation continue sur l'interprétation des textes, pour affiner l'algorithme positivement quand il s'agit de répondre au mieux aux requêtes, négativement quand il s'agit de se défendre des ruses. Et, en cas d'échec, Google ne s'interdit pas quelques interventions manuelles pour modifier les classements.

Il est possible que, sur le moyen terme, cette négociation ait une influence décisive sur les genres des documents numériques, puisque ici contrat de lecture et genres sont corrélés. Sans doute Google ne se substitue pas au lecteur, mais, privilégiant tel ou tel document selon la requête, sélectionnant quelques extraits sur une page de résultat, il oriente sa façon de lire. Le passage à une recherche universelle multimodale (transversale à toutes sortes de textes, d'images, de vidéos), l'intégration des livres par leur numérisation à grande échelle, l'intégration de la presse par Google-Actualité ou la présentation des pages résultats ne sont pas des choix sans conséquence pour les genres numériques. La capacité de chercher transversalement dans l'ensemble des textes réduit l'importance de leur classement antérieur, bouscule les anciennes hiérarchies intellectuelles, favorise les collages et participe au déplacement d'anciennes autorités intellectuelles. Enfin, la tendance à la personnalisation du service pourrait bien conduire progressivement à des genres davantage organisés par communauté que par fonction, non sans risque d'enfermement dans des bulles informationnelles plus ou moins étanches, comme le dénonce Eli Pariser²².

La pression des annonceurs est un autre risque pour l'indépendance d'un moteur, risque très souvent souligné, y compris par Larry Page et Sergey Brin, fondateurs de la firme, dès leur première présentation du *Pagerank* en 1998²³. Mais, si on ne peut exclure que la firme favorise un annonceur dans les classements, le risque paraît moins élevé que pour tout autre média construit sur l'économie de l'attention, ou plutôt l'influence des annonceurs est plus indirecte. Dès les premières années de la firme, entre les années 2000 et 2002, Google a construit en effet un système tout à fait original pour organiser à son profit et sous son contrôle sa relation avec les annonceurs. Il s'est pour cela appuyé sur son savoir-faire en vendant en ligne aux annonceurs des mots clés, les mots clés définissant la place de l'annonce sur une page Web correspondante. Les mots clés sont mis aux enchères selon un système particulier²⁴. De plus, l'enchère n'est pas le seul paramètre pour déterminer le gagnant, Google y ajoute un indicateur de qualité de l'annonce dont il détermine lui-même les paramètres.

Sous l'impulsion de Hal Varian, qui a rejoint la firme comme économiste en chef en 2002, Google a fait basculer la totalité de son marché sous ce système. Celui-ci répondait parfaitement aux principes de la théorie des jeux, où un acteur ne prend de décision qu'en fonction de l'anticipation qu'il fait des décisions de ses concurrents. Dès lors, Google s'est mis à recruter des économètres. La firme utilise des dizaines de tableaux qui s'affichent en temps réel tout comme la Bourse. Sur un tableau, on peut surveiller les recherches, le montant de l'argent récolté, le nombre d'annonceurs, sur combien de mots clés ils surenchérissent et le niveau de retour pour chaque annonceur. L'évolution est comparable à celle qui a suivi l'informatisation des places boursières avec l'arrivée massive de mathématiciens-statisticiens à la recherche de martingales pour optimiser les gains dans le maelström de cours d'actions, avec les conséquences que l'on connaît. On peut dire que Google a construit son propre système économique interne en contrôlant tous les paramètres du marché des annonceurs.

Cette organisation du marché publicitaire doit être mise en parallèle avec le développement de services ou le rachat d'entreprises visant à couvrir tout l'espace documentaire numérique et son suivi statistique selon tous les paramètres possibles. Le principal effet de la pression du marché publicitaire est sans doute d'avoir incité la firme à se développer vers les services documentaires. L'année 2005 marque une nette accélération. Cette année-là ont été ouverts coup sur coup : Google Earth, Google Maps, Google Talk, Google Video, Google Desktop, Google Book Search. Si ces services ont connu depuis des fortunes diverses, le mouvement général d'investir l'ensemble du système documentaire ne s'est pas ralenti, bien au contraire, avec, entre autres, Android, Youtube, Google Chrome, Google +.

Pour bien comprendre le succès de la firme et ses conséquences sur le néodocument, il est utile d'observer l'évolution de son chiffre d'affaires. En 2010, Google a engrangé 29,3 milliards de dollars, un chiffre impressionnant pour une firme aussi jeune (moins de la moitié tout de même du chiffre d'Apple. Le

contenu rapporte moins que la forme). Le graphique ci-dessus présente l'évolution de la répartition du chiffre d'affaires, mesurée cette fois par trimestre depuis que la firme a clairement affirmé sa stratégie en 2005.

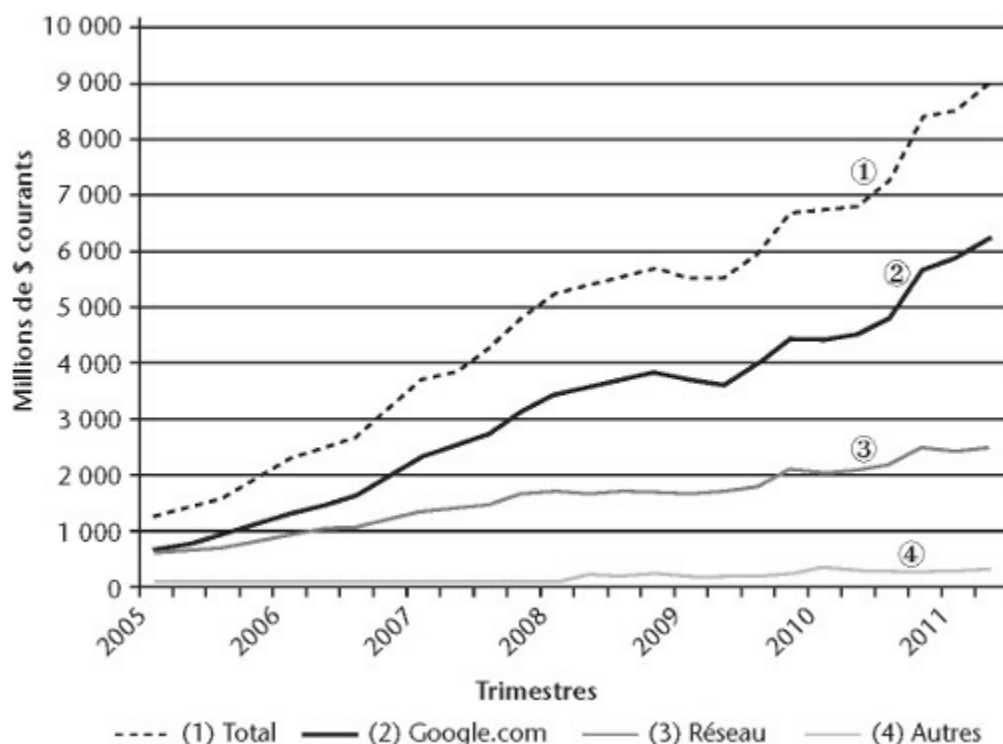


Figure 8. Évolution du chiffre d'affaires trimestriel de Google (2005-2011)²⁵

La courbe 4 présente les revenus autres que la publicité. On constate qu'ils ne décollent pas. Google doit toute sa fortune à la publicité (97 % du chiffre d'affaires du deuxième trimestre de 2011). Mais le plus intéressant est l'évolution contrastée des courbes 3 et 2. La courbe 3 présente l'activité de régie publicitaire de Google, c'est-à-dire le placement de publicités dans des sites tiers (*Adsense*). La courbe 2 représente la publicité sur les sites propres de Google. En 2005, les deux rentrées publicitaires faisaient jeu égal. Puis, tandis que la régie avait une croissance modeste, les activités propres de la firme décollaient, expliquant à elles seules la croissance générale du chiffre d'affaires. Le développement de Google sur l'ensemble des services documentaires porte ses fruits et peut se lire très directement dans son chiffre d'affaires. On peut aussi l'interpréter comme le contrôle à des fins publicitaires de l'ensemble de notre système documentaire personnel par la firme.

Google s'appuie sur le deuxième paradoxe relevé plus haut, celui de la lecture et du calcul, en cherchant à le dépasser par une couverture globale de notre système documentaire. Pour ce faire, il a besoin d'énormes centres de données qui enregistrent en permanence l'ensemble du Web, du moins celui qui est accessible à ses robots. La croissance des services propres de Google assure aussi des économies en réduisant les coûts de bande passante, notamment pour l'audiovisuel (YouTube). À l'intérieur, ces centres de données fonctionnent comme des indexeurs, réalisant le rêve de Paul Otlet d'un répertoire universel mais sans intervention humaine, sinon celle des programmeurs affinant les algorithmes. Autre différence, le répertoire reste confidentiel, propriété de la firme. Nerf de la guerre, cette capacité de lecture industrielle est jalousement gardée, parfois même protégée par des contrats léonins, comme dans le cas de la numérisation des livres des bibliothèques.

Contrairement à Apple, l'exploitation de la deuxième dimension du document par le calcul sur le texte entre en contradiction frontale avec les médias plus anciens, car il privilégie une construction de services à partir de la lecture et non une diffusion d'œuvres inviolables. Les polémiques et procès n'ont pas manqué. Mais, même si Google a réduit plusieurs fois ses ambitions face au principe de la propriété intellectuelle, il est peu probable qu'il modifie radicalement une stratégie à l'origine de son succès.

Nous avons vu que les moteurs étaient apparus pour pallier une difficulté du Web traditionnel, qui n'était pas construit comme une bibliothèque : retrouver les documents pertinents parmi tous ceux publiés. Même si Google reprend à son compte certains raisonnements du Web des données, par exemple dans l'utilisation des cartes et la géolocalisation, il semble que ce soit dans une autre direction qu'il porte maintenant principalement ses efforts : l'étude des comportements des internautes pour cibler au mieux les informations. Il s'agit d'une évolution récente²⁶. La firme détient par son moteur la plus grosse base de données sur les intentions des internautes, mais le savoir-faire de Google est d'abord contextuel, une expertise de traitement des textes, aussi bien les requêtes des internautes que les documents publiés sur le Web, et non comportemental, une connaissance des réactions des personnes. Cette dernière expertise serait plutôt celle de Facebook ou d'autres jeunes firmes intéressées par la revente de données. D'un côté, une expertise linguistique ; de l'autre, une expertise sociologique. C'est aussi dans ce contexte que l'on peut relire le slogan « *don't be evil* » (ne soyez pas méchant) ; on pouvait chez Google manipuler les mots, pas les personnes. Les débats internes sur l'utilisation des *cookies*²⁷ par exemple ont été très vifs. Suite à l'achat de la régie publicitaire Double-Click en 2007, à la montée de la concurrence de Facebook, la position de la firme a évolué. Eric Schmidt, dans l'entretien déjà cité, indiquait : « Ce qui rend les journaux si fondamentalement fascinants – les trouvailles fortuites (*serendipity*) – peut aujourd'hui être calculé électroniquement. [...] La technologie sera si efficace qu'il sera très difficile pour les gens de regarder ou consommer quelque chose qui n'aura pas été d'une certaine façon taillé sur mesure pour eux²⁸. » Mettons de côté l'arrogance de la formule, et constatons que, ce faisant, Google s'éloigne en effet de son premier métier et qu'il s'agit ni plus ni moins du raisonnement d'un média traditionnel de diffusion cherchant à personnaliser sa formule.

La stratégie du médium : Facebook

Facebook est allé encore plus loin dans la prise en compte de l'internaute, en privilégiant non pas sa relation au document, mais le rôle du document dans les relations humaines. Ce n'est pas l'ordre documentaire ancien, ni même le contenu d'un document qui priment ici, mais le fait qu'un document relie entre eux des lecteurs. Les capacités documentaires du Web sont alors renversées. Les documents ne sont plus des éléments recherchés pour eux-mêmes, mais des vecteurs qui relient entre eux des humains. En même temps, les humains représentés par leur page personnelle deviennent eux-mêmes des documents qui peuvent être repérés, non seulement dans leur « forme », c'est-à-dire leur identité et leurs caractéristiques, mais aussi dans leur « texte », leurs comportements et intentions, ce qu'ils lisent et écrivent ou ce que les autres lisent et écrivent sur eux ou communiquent avec eux. On pourrait dire comme Olivier Ertzscheid que « l'homme est un document comme les autres²⁹ » ; les techniques du Web documentaire servent à indexer les individus. Facebook, en voulant utiliser sa maîtrise du réseau de relation comme un avantage concurrentiel décisif pour valoriser la vente d'attention, polarise l'ordre documentaire sur sa troisième dimension, le médium, la transmission.

La radicalisation de l'homologie individu-document modifie encore une fois l'ordre des documents, qui est alors soumis aux réactions des individus. La réflexivité est immédiate et la versatilité des relations sociales s'introduit dans les logiques documentaires, qui ne les ignoraient évidemment pas mais avaient auparavant plus de recul et d'inertie. La page Facebook d'un individu s'apparente à un écran de télévision dont tous ses amis et lui-même se partageraient la télécommande. Le flot des nouvelles et des commentaires s'affiche en continu, les dernières faisant disparaître les plus anciennes. L'ordre documentaire est celui d'une mémoire sociale vive, avec son efficacité pour recommander tel ou tel document avec confiance, sa rapidité, ses fulgurances, ses emballements, ses rires, ses futilités, ses agacements, ses lourdeurs et ses oublis. Il peut être compris comme une tentative pour dépasser le troisième paradoxe entre conversation éphémère et traces pérennes.

La vraie innovation de Facebook est peut-être de considérer qu'il n'existe qu'un et un seul « graphe social », une possibilité de représenter l'ensemble des interconnexions des internautes qu'il cherche à couvrir dans sa globalité, comme un seul programme de télévision auquel tout le monde serait connecté et sur lequel tout le monde zapperait. Comme l'explique Pierre Bélanger³⁰, fondateur de SkyRock, après les premiers réseaux « entre-soi », Facebook propose un « hybride mutant entre la micro-socialisation et la globalisation planétaire » en devenant le premier réseau social universel sous identité réelle. « L'expérimentation et les évolutions d'identité de l'adolescent n'y ont pas leur place. La complexité des relations et l'hétérogénéité des contextes de socialisation d'un adulte – lui-même à facettes – n'y sont pas prises en compte », ajoute-t-il.

Le succès planétaire de cette innovation illustre la différence entre les échanges sociaux et leur traduction dans le monde documentaire ou, si l'on veut, la différence entre sociologie et sciences de l'information. Les

relations sociales, en se documentarisant, s'ouvrent à un ordre différent : celui de la publication, que je m'efforce d'éclairer un peu dans ce livre. Le premier usage du réseau est alors à forte valeur ajoutée par les retrouvailles et les découvertes qu'il autorise, montrant l'efficacité de l'outil documentaire.

Mesuré à la croissance du trafic ou à son nombre d'abonnés, Facebook est une réussite impressionnante, peut-être la plus spectaculaire de la petite histoire du Web. Le réseau compterait six cents millions de membres début 2011, répartis sur toute la planète. Malgré cette réussite, la valorisation économique de ce positionnement paraît se chercher encore aujourd'hui sur deux marchés : la publicité ciblée et la revente des données comportementales. « Vous pouvez cibler votre publicité par géographie, sexe, âge, mots clés, situation amoureuse, emploi, lieu de travail ou établissement scolaire. Lorsque vous choisissez vos critères de ciblage, Facebook affiche le nombre approximatif d'utilisateurs concernés », peut-on lire sur les pages réservées aux annonceurs. Ce ciblage idéal suppose que les membres du réseau aient renseigné de façon sincère ces critères. Nombreuses ont été les tentatives de la firme pour « obliger » les membres à diffuser leurs profils et nombreuses ont aussi été les réticences et polémiques.

Les données comportementales sont récoltées par les traces de navigation des internautes laissées implicitement ou par les avis déposés explicitement (le bouton « j'aime »). Il s'agit cette fois d'un repérage automatique. Un marché souterrain de ces données, transversal aux réseaux, s'est mis en place³¹. Aujourd'hui, le débat éthique est très vif. Jusqu'à présent, seules les organisations policières confondaient l'ordre documentaire et l'ordre social pour contrôler ce dernier. Cela n'était acceptable qu'avec des normes sociales explicites et acceptées sous le contrôle strict d'un État démocratique, sans pour autant éviter de nombreux dérapages. Même en faisant abstraction de ce débat, la valeur économique de cette stratégie n'est pas encore vraiment démontrée. On en perçoit les potentialités, qui 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.

Il est délicat de mesurer l'économie de Facebook, car la firme n'est pas cotée en Bourse et ne publie pas ses résultats. L'économie de Facebook est souvent confondue avec l'économie générée par ceux qui utilisent le réseau comme plateforme pour développer leurs propres activités rémunératrices, autour principalement de la publicité. Les chiffres récoltés ici ou là sont des spéculations destinées le plus souvent à faire monter les enchères des investisseurs intéressés. Il semble que la firme a juste équilibré ses comptes en 2010. Elle est encore sans doute très loin des résultats d'Apple ou de Google, même en tenant compte de sa jeunesse. Si les revenus étaient à la hauteur du succès populaire, ils seraient clairs.

Intégration

Même si chacune des firmes privilégie une dimension pour en faire un avantage concurrentiel, délaisser les autres c'est prendre le risque de voir un concurrent verrouiller le marché définitivement. Ainsi, un modèle commun s'esquisse, intégrant les trois dimensions. Il comprend de gigantesques plateformes d'échange où l'on pourra retrouver l'ensemble de la production des industries de la mémoire, depuis la musique, les films, les séries TV, les livres numériques, les journaux, les jeux vidéos, jusqu'aux productions des internautes eux-mêmes, blogs, photos, vidéos, archives personnelles, en passant par des milliers de petites applications utiles ou divertissantes réalisées par des développeurs de la firme elle-même ou indépendants entrepreneurs. L'accès à ces documents et services passe par le contrôle des terminaux mobiles et leur sélection se réalise soit par un moteur de recherche, soit par une recommandation d'un internaute. Ainsi, l'ensemble des documents et services associés est déporté « dans les nuages », c'est-à-dire sur les énormes centres de gestion de données de ces firmes. Sa rentabilité s'appuie sur la connaissance fine du consommateur afin d'adapter les services à sa demande ou de revendre son attention à des annonceurs. Ce modèle peut s'appliquer aussi bien dans un cadre de loisir que dans un cadre professionnel, mais dans ce second cas le recours à un financement publicitaire paraît exclu et il sera remplacé par une location de services.

Au moment où j'écris ces lignes, la bataille pour le contrôle global du modèle est lancée. Google cherche ainsi à prendre pied sur les deux dimensions qu'il avait un peu négligées jusque-là avec des initiatives majeures à l'été 2011 : le lancement du réseau social de Google +, d'un côté, et le rachat du fabricant de téléphones Motorola, de l'autre. Ce modèle de Web média commercial ne concerne évidemment pas que les firmes que nous venons de citer, d'autres tentent d'y prendre place. Amazon est de celles-là pour l'Europe et l'Amérique du Nord, Baidu pour la Chine. Quels que soient les contextes régionaux, le modèle ne laisse néanmoins pas beaucoup de place à la concurrence. Seuls quelques champions dominant, laissant se développer autour d'eux des firmes innovantes sur tel ou tel créneau, qu'ils rachèteront en cas de succès grâce aux énormes liquidités qu'ils amassent. Pour le moment, le plus gros danger qui les guette est leur puissance même et la menace de procédures antitrust déjà parfois engagées aux États-Unis et en Europe.

Notes du chapitre 5

1. Lawrence LESSIG, *L'Avenir des idées. Le sort des biens communs à l'heure des réseaux numériques*, op. cit.
2. « Entretien avec Lawrence Lessig, professeur de droit à Harvard », *Bibliobsession*, 27 mai 2011. <<http://www.bibliobsession.net>>.
3. Jean-Claude GUÉDON, « Le Créateur, le génie et l'éditeur », loc. cit.
4. Dan GILLMOR, *We the Media. Grassroots Journalism by the People, for the People*, op. cit.
5. Paul Felix LAZARSFELD, Bernard BERELSON, Hazel GAUDET, *The People's Choice : how the voter makes up his mind in a presidential campaign*, Columbia University Press, New York, 1948 ; Elihu KATZ & Paul Felix LAZARSFELD, *Personal Influence : the Part Played by People in the Flow of Mass Communications*, Transaction Publishers, Piscataway (NJ), 2006 (1955).
6. Jack GOODY, *La Raison graphique. La domestication de la pensée sauvage*, Éditions de Minuit, Paris, 1979.
7. Bruno BACHIMONT, *Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*, mémoire d'habilitation à diriger les recherches, Université de Technologie de Compiègne, 12 janvier 2004. Accessible sur <www.utc.fr/~bachimon>.
8. Clarisse HERRENSCHMIDT, *Les Trois Écritures. Langue, nombre, code*, Gallimard, Paris, 2007.
9. Stephen WOLFRAM, *A New Kind of Science*, Wolfram Media, Inc., Champaign (IL), 2002. Accessible sur <www.wolframscience.com>.
10. Christian VANDENDORPE, « La lecture en éclats », *Arguments*, vol. 11, n° 1, automne 2008-hiver 2009, p. 30-39.
11. Roger T. PÉDAUQUE, *Le Document à la lumière du numérique*, op. cit., p. 187.
12. Stevan HARNAD, « Retour à la tradition orale : écrire dans le ciel à la vitesse de la pensée », in Jean-Michel SALAÜN & Christian VANDENDORPE (dir.), *Les Défis de la publication sur le Web : hyperlectures, cybertextes et méta-éditions, Actes des Quinzièmes Entretiens du Centre Jacques Cartier*, Presses de l'Esssib, Villeurbanne, 2002.
13. Eric SCHMIDT, cité in Holman W. JENKINS Jr, « Google and the Search for the Future », *The Wall Street Journal*, n° 14, août 2010 (traduction personnelle).
14. Par exemple : Deb ROY, « The birth of a word », conférence TED, mars 2011. <www.ted.com>. Pour une présentation rapide de ces problématiques : Marie MARTEL, « L'hyperdocumentation et la mémoire qui fabrique le futur », *Argus*, 2 septembre 2011.
15. Carl SHAPIRO et Hal R. VARIAN, *Information Rules*, op. cit.
16. Steve Jobs est décédé d'un cancer le 5 octobre 2011.
17. Présentation des résultats financiers du deuxième trimestre 2011 (traduction personnelle).
18. Source : bilans financiers de la firme.
19. Richard STALLMAN, « The danger of e-book », non daté, disponible sur : <<http://stallman.org>>.
20. Alain GIFFARD, « Lectures industrielles », *Ars Industrialis*, 5 décembre 2007. <<http://arsindustrialis.org>>.
21. Le *fair-use* ou « usage équitable » est à la fois une doctrine et un ensemble de règles qui tentent d'équilibrer intérêts des ayants droit et intérêt public aux États-Unis, en autorisant certains usages d'une œuvre malgré les droits exclusifs de l'auteur. Il s'appuie sur des critères (*factors*) que les tribunaux doivent apprécier. Le *fair-use* a donné leurs bases juridiques à plusieurs firmes du Web et la Computer and Communications Industry Association (CCIA) publie régulièrement un rapport sur les « industries du *fair-use* ». <www.ccianet.org>.
22. Eli PARISER, *The Filter Bubble : What the Internet is Hiding from You*, The Penguin Press, Kindle Edition, New York, USA, 2011.
23. Sergey BRIN et Lawrence PAGE, « The anatomy of a large-scale hypertextual web search engine », Seventh International World-Wide Web Conference, Brisbane, Australia, 14-18 avril 1998. Voir : « Appendix A : advertising and mixed motives ».
24. Steven LEVY, « Secret of Googlenomics : data-fueled recipe brews profitability », *Wired Magazine*, 17 juin 2009.
25. Source : bilans financiers de la firme.
26. Jessica E. VASCELLARO, « Google agonizes on privacy as ad world vaults ahead », *Wall Street Journal*, 10 août 2010.
27. Les *cookies* sont des données envoyées par le serveur d'un site visité une première fois au navigateur de l'internaute et stockées par celui-ci. Par la suite, le navigateur les renverra au serveur à chaque nouvelle visite, permettant de pister leur cheminement.
28. Eric SCHMIDT, cité in Holman W. JENKINS Jr, « Google and the Search for the Future », loc. cit.
29. Olivier ERTZSCHEILD, « L'homme est un document comme les autres : du World Wide Web au World Life Web », *Hermès*, n° 53, 2009, p. 33-40.
30. « Pierre Bellanger : l'avenir des réseaux sociaux », *Henri Verdier Blog*, 23 mars 2011. <<http://www.henriverdier.com>>.
31. Julia ANGWIN, « The Web's new gold mine : your secrets », *Wall Street Journal*, 30 juillet 2010.

Conclusion

Architectes et archithèques

Au terme de ce voyage dans le monde du document, nous constatons que le modèle le plus ancien, la bibliothèque, est celui qui a inspiré le modèle le plus contemporain, que nous avons appelé Web média. Même si les appétits commerciaux ont débouché sur la constitution d'énormes firmes en situation d'oligopole, nous avons vu aussi que d'autres constructions sont en cours, en particulier sous l'impulsion du consortium W3C avec le Web des données. Dans le même temps, sous l'effet du numérique, l'ensemble des industries culturelles, bibliothèques comprises, se transforment en « industries de la mémoire ».

L'ampleur des mouvements devrait nous obliger à penser plus globalement les transformations de l'ordre documentaire et l'émergence d'un néodocument. La question n'est pas oiseuse. Il s'agit de la reconsidération de notre rapport à la représentation des événements passés pour envisager l'avenir. Premières concernées, les infrastructures épistémiques d'autrefois, portées par les bibliothèques, les centres documentaires ou encore les musées, se sont déplacées aujourd'hui vers les réseaux et les collections numériques et irriguent tous les échelons de la société. La science elle-même, construite pour une bonne part sur la confrontation des documents, bascule dans de nombreuses disciplines vers une *e-science*, c'est-à-dire une science où les outils numériques deviennent dominants, ouvrant des possibilités inédites d'investigation, de calcul et d'analyse. On parle parfois de bibliothèques numériques ou encore de cyberinfrastructures pour signifier ces nouvelles institutions au service d'une collectivité particulière.

Les institutions traditionnelles étaient portées par des professions reconnues, aux compétences codifiées et à l'éthique forte, comme nous l'avons vu. Les nouvelles infrastructures épistémiques ne disposent pas de cette tradition et souvent même leur rattachement institutionnel est flottant. Il serait donc utile de former de nouveaux professionnels de l'information capables de comprendre les mouvements en cours pour construire et gérer les nouvelles infrastructures documentaires qui émergent. Ceux-ci devront connaître tout autant les outils documentaires traditionnels que les développements informatiques du Web soutenus par le W3C et ceux impulsés par les stratégies industrielles. Ils devront être capables de gérer en amont l'organisation des documents sous toutes leurs formes et en aval de comprendre, maîtriser et faciliter l'action des internautes par des services au design adapté pour une rencontre optimale des deux mondes.

En Amérique du Nord, un mouvement académique récent a déjà pris la mesure des nécessités nouvelles de formation. Les *iSchools* (pour *information schools*) issues des meilleures écoles de bibliothéconomie forment aujourd'hui des professionnels aussi compétents en informatique qu'en sciences de l'information. On les appelle parfois des « *information architects*¹ », des architectes de l'information, pour signifier leur homologie avec les architectes qui dessinent les plans d'un bâtiment. Un fort accent est mis sur le design et la conception d'outils intuitifs pour l'internaute. Les architectes de l'information ont leur association professionnelle, l'Information Architecture Institute², qui propose cette définition pour l'architecture de l'information :

- la conception structurelle des espaces d'information partagée ;
- l'art et la science d'organiser et de cataloguer des sites Web, des intranets, des communautés virtuelles et des logiciels pour en faciliter l'utilisation et le repérage ;
- une communauté de pratique émergente fondée sur l'apport de principes du design et d'architecture spécifiques à un environnement numérique.

La profession est jeune et bien des débats ont eu lieu sur sa définition. L'important pour moi est surtout de constater la nécessité de faire converger les pratiques issues des professions documentaires, y compris dans leurs traditions éthiques, et celles du Web, les sciences de l'information et les sciences informatiques. L'enjeu pour ces nouveaux professionnels sera demain de concilier les pratiques des différents mondes du document : le monde traditionnel où, par l'inscription d'un contenu sur un support, la logique de la première dimension dominait et domine encore ; celui des inventeurs du Web, qui mettent plutôt l'accent

sur la deuxième dimension en reconstituant les documents par la sémantique ; et enfin celui des industriels, dont la logique commerciale privilégie de plus en plus la troisième dimension par la traçabilité des navigations. Il leur faudra, de façon très pragmatique, construire à la fois des prestations et des institutions qui soient réellement dédiées à la communauté qu'ils servent, reprenant à leur compte la longue tradition des infrastructures épistémiques, sans l'inféoder aux stratégies industrielles qui visent à verrouiller le Web ni la réduire à la logique des ingénieurs, en conciliant les trois dimensions du document.

Je suggérerais le néologisme « archithèque³ » pour nommer les nouvelles infrastructures épistémiques héritières des bibliothèques et du Web qui s'inventent aujourd'hui. La parenté de consonance avec *architecte* et *archives* me semble bienvenue, car les compétences de leurs responsables puisent aussi largement dans ces traditions-là. La relation à l'architecture vient d'être soulignée, il s'agit bien du design de l'infrastructure documentaire. Concernant celle avec l'archivistique, Seamus Ross, actuel doyen de la faculté d'information de l'Université de Toronto, remarquait en 2007 : « En réfléchissant aux bibliothèques numériques, on s'aperçoit rapidement qu'elles sont peut-être bibliothèques de nom, mais archives de nature. Le contenu qu'elles accueillent est, pour l'essentiel, unique et n'a pas vraiment besoin d'être détenu ailleurs, car les services en réseau font qu'il peut être offert n'importe où et n'importe quand à partir d'une source unique. Quand les usagers accèdent à ce contenu, ils s'attendent à pouvoir lui faire confiance, vérifier son authenticité (mais pas nécessairement sa fiabilité), ils demandent la connaissance de son contexte de création et des preuves de sa provenance⁴. » Pour le dire avec les mots de la théorie du document que nous avons déployée dans ce livre, sur le Web il n'y aurait pas vraiment de différence entre un protodocument et un néodocument.

Enfin, le préfixe *archi-* signifiant aussi « au-delà », et *-thèque*, « étagère », on peut comprendre aussi l'archithèque comme le dépassement d'une bibliothèque. Et donc, pour éviter les querelles de dénomination avec une profession protégée par un ordre, les architectes, on pourrait nommer les responsables d'une archithèque des « archithécaires ».

Notes de la conclusion

1. Louis ROSENFELD et Peter MORVILLE, *Information Architecture for the World Wide Web*, O'Reilly, Sebastopol (CA), 2006 (3^e éd.).

2. <<http://iainstitute.org/fr/>>.

3. Il a pu m'être reproché par quelques puristes de faire un barbarisme en construisant un néologisme à partir d'un préfixe latin et d'un suffixe grec, mais ce serait oublier qu'« archives » vient tout autant d'*archivum* que d'*ἀρχαία*.

4. Seamus ROSS, « Digital preservation, archival science and methodological foundations for digital libraries », conférence d'ouverture 11^e European Conference on Digital Libraries (ECDL), Budapest, 17 septembre 2007 (traduction personnelle).

Table des Matières

Page de titre	4
Présentation	5
L'Auteur	5
Copyright	6
Table	7
Dédicace	8
Introduction. Une approche documentaire du Web	9
1. Au commencement, la bibliothèque	12
2. Les mutations du document	20
3. Réingénieries documentaires	33
4. L'économie du document	47
5. À la recherche du néodocument	60
Conclusion. Architectes et archithèques	71