



HAL
open science

Bibliothèques numériques et Google-Print

Jean-Michel Salaün

► **To cite this version:**

Jean-Michel Salaün. Bibliothèques numériques et Google-Print. Regards sur l'actualité : mensuel de la vie publique en France , 2005. sic_00001576

HAL Id: sic_00001576

https://archivesic.ccsd.cnrs.fr/sic_00001576

Submitted on 21 Oct 2005

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.

Bibliothèques numériques et Google-Print¹

Jean-Michel Salaün
École de bibliothéconomie et des sciences de l'information
Université de Montréal

20-10-2005

Article pour la revue *Regard sur l'actualité*, décembre 2005 La Documentation française.
Texte non révisé par l'éditeur.

Chapeau :

Le projet de mise en accès de millions de livres numérisés lancé fin 2004 par la firme Google a brutalement réveillé les débats sur les relations entre le livre et le Web. Le président de la Bibliothèque nationale de France a vivement réagi, les « ayant-droits » ont protesté, les concurrents ont pris des initiatives.

En moins d'une année la réflexion a été renouvelée et l'accès au patrimoine livresque européen est devenu un objectif politique et culturel prioritaire.

Cette effervescence et la divergence des arguments et des initiatives illustrent un tournant dans notre relation au livre et dans sa place dans notre culture.

La courte vie de *Google-Print* et de ses alternatives, déjà fertile en rebondissements, mérite d'être contée car elle a révélé des dimensions, économiques, politiques et techniques, insoupçonnées dans la mise en ligne de livre. Elle n'est pas terminée, et pour bien en décrypter les épisodes à venir, il est sans doute préférable de replacer dans leur contexte les épisodes passés.

Le projet Google-Print dans son contexte

Google a été fondée, il y a sept ans, par deux informaticiens qui venaient de terminer leur thèse à l'université de Standford : Larry Page et Sergey Brin. Après avoir vainement contacté les principaux acteurs de la recherche d'information en place sur le Web pour exploiter leurs idées, ils ont ouvert leur propre entreprise et démarré une de ces *success stories* que l'on apprécie particulièrement en Californie. La croissance de la firme a été en effet explosive. Démarrée sur la base d'un moteur de recherche d'informations sur le Web, elle a très vite construit un modèle d'affaires original d'une impressionnante efficacité. Aujourd'hui, bien malin qui pourra dire où elle s'arrêtera, tant elle semble épouser les contours du développement du Web lui-même. Ainsi le projet *Google-Print* n'est qu'une des initiatives de *Google*², spectaculaire sûrement mais pas moins que d'autres lancées par la même firme dans différents secteurs.

¹ Pour réaliser cet article, l'auteur a bénéficié des échanges sur le réseau RTP-DOC (<http://rtp-doc.enssib.fr>). Il remercie tout particulièrement Elisabeth Noël, Olivier Ertzcheid, Nicolas Morin et Christian Vandendorpe dont les remarques ont permis une amélioration sensible du texte.

² On trouvera toutes les informations officielles de la firme sur son site : <http://www.google.com/>, notamment à l'onglet « about Google ». Pour des informations en français, on peut consulter : <http://www.google.fr>, moins complet.

Un moteur

Google est d'abord un moteur de recherche sur le Web, c'est-à-dire un outil de calcul analysant les phrases grâce à des méthodes statistiques et morphosyntaxiques pour repérer des pages Web correspondant aux requêtes des internautes. L'idée première et originale des fondateurs de *Google* était d'adjoindre aux techniques habituelles de recherche d'information une variable supplémentaire : la notoriété, mesurée par le nombre de liens hypertextuels qui faisaient référence à une page donnée, pour classer les réponses à une requête donnée.

Cette idée, combinée avec une simplicité de l'interface, une stricte utilisation de l'opérateur booléen « et » ainsi qu'une indexation plus large des pages, a très vite montré sa pertinence et a conduit au succès originel de la firme. Une des grandes difficultés de la recherche sur le Web, n'est pas tant, en effet, le repérage des pages que son classement afin que l'internaute trouve l'information qui satisfera sa demande. Aujourd'hui ce principe a été repris par les autres grands moteurs, et il ne saurait à lui seul expliquer la pérennité du succès de celui-là. Il faut y adjoindre bien d'autres modalités concernant, par exemple, la taille et la structure des index, les techniques d'indexation, de stockage, de filtrage, toutes relevant du secret commercial jalousement gardé et alimentant spéculations, enquêtes et analyses des experts³.

Quoi qu'il en soit, le moteur est bien le métier premier de *Google* et son cœur. C'est dans ce domaine que les efforts les plus importants ont été réalisés. Jusqu'à présent le moteur de *Google* a su se maintenir dans le peloton de tête, tout au moins pour la pertinence perçue par les internautes. Mais, ce leadership pourrait devenir un talon d'Achille si cette activité supportait l'ensemble de la rentabilité. Aucune firme importante ne peut maintenir ses positions face à la concurrence à partir d'un seul créneau.

Ainsi la diversification est indispensable. Cette diversification va s'appuyer sur une conséquence annexe du fonctionnement du moteur. Pour atteindre sa pleine efficacité, celui-ci a archivé sur les disques durs de ses ordinateurs un nombre considérable de pages Web et ainsi il s'est construit un observatoire du fonctionnement de la toile par la connaissance des contenus croisée avec celle du comportement des internautes au travers de leurs requêtes. Cet effet d'échelle lui a ouvert d'autres horizons.

Mais pour comprendre vraiment les orientations de la diversification de *Google*, il faut d'abord repérer le modèle d'affaires de la firme. Que vend-elle ? Comment réalise-t-elle des bénéfices ?

Un modèle d'affaires

A première vue le modèle d'affaires de *Google* ne paraît pas très original. Il s'agit tout bonnement de la vente d'espaces publicitaires comme la réalise depuis longtemps les médias classiques : presse, radio ou télévision commerciales. Et pourtant la différence est radicale pour trois raisons au moins que les responsables de la firme ont exploitée avec un certain génie : le médium principal est le langage écrit soumis au calcul, la géographie n'a pas la même métrique, l'internaute est actif en temps réel.

Même si l'image devient de plus en plus présente sur le Web, ce dernier est d'abord construit sur la langue écrite. Le contenu des pages et leur indexation reposent sur l'écrit et un moteur analyse le langage au travers de sa manifestation écrite. Ainsi les mots permettent de repérer des pages

³ On trouvera les échos des débats de spécialistes par exemple dans l'un de ces blogs : <http://aixtal.blogspot.com/> ou <http://affordance.typepad.com/>.

aussi bien que des requêtes d'internautes. De là à s'en servir comme moyen pour placer des encarts publicitaires aux endroits les plus appropriés pour toucher des internautes intéressés il n'y a qu'un pas, vite franchi par *Google*. Ainsi, la firme vend des « mots-clés », correspondant aux pages Web indexées par la firme. Les annonceurs, insérant leur message face aux réponses aux requêtes des internautes (*AdWords* 51% du CA) ou sur une page Web particulière (*AdSense*, 48% du CA), font ainsi correspondre, du moins peut-on le supposer, leur proposition commerciale aux centres d'intérêts des clients potentiels. Cette vente est une mise aux enchères, donc, en théorie du moins, le prix est au plus haut du marché. Enfin, les coûts de transaction sont au plus bas puisque tout est automatisé grâce à l'ingénierie linguistique et statistique, métier premier de la firme.

Deuxième caractéristique originale du Web : les frontières géographiques et sociales sont transgressées, au moins potentiellement, par le réseau. Ainsi la firme, à condition d'avoir la force commerciale adéquate, s'adresse à un marché mondial d'annonceurs et, en même temps, peut segmenter les cibles publicitaires sur un territoire éclaté. Le vieux fantasme des professionnels du marketing est, en théorie là aussi, réalisé : la personnalisation de masse, personnalisation puisque le message commercial est chaque fois adapté à l'internaute, masse puisque la taille du réseau fait que chaque message touchera un nombre d'internautes statistiquement significatif.

Enfin, le caractère interactif du média fait que l'on peut vérifier, au moins partiellement, l'impact d'un message publicitaire. L'internaute, en cliquant sur le lien qu'on lui propose, manifeste un certain intérêt pour la marchandise ou le service qu'il croit trouver derrière.

Bien entendu en pratique, un tel modèle d'affaires n'est pas à l'abri des erreurs de cible, ni des comportements parasites et opportunistes, qui se multiplient au fur et à mesure du succès grandissant. Comme tous les acteurs disposent potentiellement des mêmes opportunités de calcul, le phénomène des « passagers clandestins » (Spam, Splogs et autres pourriels) prend des allures d'invasion. Mais la culture publicitaire est une culture de la négociation pragmatique. Et l'on trouve ici le même flou, entre les prix affichés, les remises, les retours etc., que sur n'importe quel support médiatique, mais évidemment à l'échelle du Web.

Ce modèle a fait et continue de faire la fortune de *Google*. Certains⁴ l'ont appelé « l'économie de la recherche ». Contrairement à bien des firmes du secteur, celle-là a été très vite rentable et s'appuie sur un chiffre d'affaires réel. Mieux, la firme a fait, l'année dernière, une entrée en bourse fracassante et, profitant d'une cotation intéressante, a levé tout récemment de nouveaux capitaux. On considère aujourd'hui que le cash disponible est de 7 milliards de dollars, son bénéfice net avoisinerait les 4 milliards (12,25 pour *Microsoft*) et sa capitalisation boursière 90 milliards. Beaucoup s'interrogent sur la destination de ce pactole et pensent que nous sommes à la veille d'annonces stratégiques d'envergure.

Néanmoins, le modèle n'est pas sans faiblesse. Que les internautes se détournent du moteur et 51% des revenus s'évaporent (*Adwords*). Ainsi la firme a intérêt à développer son activité de régie publicitaire ou d'agence-média (*Adsense*).

Une écologie de l'information ou une nouvelle forme de mass-média ?

⁴ Batelle John, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, 2005.

Voir aussi son blog : <http://battellemedia.com/>

Pour pallier ces difficultés, la firme a développé une activité multiforme, ouvrant un très large éventail de services, souvent très performants et innovants, comme sur les messageries, sur la cartographie, sur le traitement de l'actualité, sur le repérage des images, sur l'indexation de la littérature scientifique, sur la réalisation de blogs, sur la téléphonie, sur le Wi-fi et, bien entendu, sur la numérisation de livres qui nous intéresse ici. L'accord stratégique conclu entre Google et Sun le 4 octobre 2005⁵ devrait encore élargir cet éventail de services en ligne, qui ne nécessitent qu'un navigateur pour offrir toutes les fonctionnalités d'un service d'exploitation classique.

Il y a deux façons d'analyser cette stratégie multidimensionnelle, débouchant sur deux orientations différentes, mais à ce stade de développement plus alternatives pour le futur que contradictoires dans le présent : soit on considère qu'il s'agit d'une stratégie monopolistique, comparable à celle de *Microsoft* dans la bureautique, soit on pense qu'il s'agit de la naissance d'un nouveau modèle de média, comme en leur temps la radio ou la télévision.

Dans le premier cas, on souligne le savoir faire technique de l'entreprise. *Google* est d'abord une firme de services informatiques qui a compris mieux que d'autres que la domination du marché passait moins dans l'avenir par la maîtrise de l'ordinateur personnel que par celle du Web. Les différents développements visent en effet à garantir une position dominante, inquiétante par son contrôle, en construisant ce que certains⁶ ont appelé une « écologie de l'information » de façon que l'internaute se sente chez lui parmi les services proposés par *Google*, comme il l'est ou l'était sous Windows, c'est-à-dire chez *Microsoft*.

Dans la seconde interprétation, on met l'accent sur son modèle d'affaires. L'objectif premier alors est de capter l'attention de l'Internaute en balisant l'ensemble de son cheminement informationnel. Les revenus provenant de la publicité s'annoncent considérables. Déjà pour 2005, *The Economist* prévoit que les combinés de *Google* et *Yahoo !* devraient être équivalents à ceux des trois grands réseaux de télévision américains⁷. La croissance des revenus publicitaires devrait continuer dans les années à venir si l'on considère que les grandes firmes réservent actuellement moins de 4% de leur budget publicitaire à l'Internet alors que ce média capte déjà environ 15% du temps moyen dédié à la consommation médiatique aux États-Unis.

Dans cette perspective, la firme s'intéresse plus à l'accès aux contenus qu'aux outils eux-mêmes. L'initiative *Google-Print*, comme celle plus ancienne de *Google-News* (accès aux actualités), relève plutôt de cette seconde logique. La personnalité des acteurs qui ont réagi à ces initiatives est révélatrice : bibliothèques et éditeurs, ou agence de presse, journaux et journalistes.

Google-Print

Le projet *Google-Print*⁸ comprend deux volets articulés.

L'un, annoncé en décembre 2004, s'adresse aux bibliothèques, plus précisément à cinq bibliothèques partenaires⁹. Il s'agit de numériser à grande échelle leur collection. Chaque

⁵ <http://www.sun.com/smi/Press/sunflash/2005-10/sunflash.20051004.1.html>

⁶ Voir par exemple : Guillaud Hubert, De la force de l'écosystème au péril du monopole, *Internet Actu*, Fing, 25-08-2005.

⁷ <http://www.internetactu.com/index.php?p=6095>

⁸ « The online ad attack », *The Economist*, 27 avril 2005.

⁹ <http://print.google.com>

⁹ Les bibliothèques des universités de Harvard, de Stanford, du Michigan aux USA, de l'Université d'Oxford en Grande Bretagne et la bibliothèque publique de New York.

bibliothèque dispose d'une version numérique des ouvrages, tandis que *Google* met en ligne, de son côté, une autre copie.

L'autre, lancé trois mois auparavant, vise les éditeurs. *Google* propose de mettre en ligne des extraits d'ouvrage de façon à en assurer la promotion en vue de leur vente via des services de librairies en ligne.

A y regarder de plus près, il semble bien que les deux volets ne soient que les deux faces d'une même médaille : les fonds des bibliothèques sont dans leur grande majorité encore couverts par le copyright ; la numérisation, nécessairement industrielle à cette échelle, supporte mal les traitements individuels générateurs de surcoûts ; et donc d'une part, il n'y a qu'un seul programme de numérisation, et, d'autre part, seule une minorité d'ouvrages dans le domaine public sont accessibles en totalité, tous les autres sont présentés sous forme d'extraits de quelques pages correspondant aux mots de la requête.

Le projet présente un certain nombre de différences avec les programmes de numérisation engagés jusqu'alors qui, additionnés, l'apparente à une rupture et explique sans doute pour une bonne part les réactions qu'il a suscité :

- la première différence est un changement radical d'échelle. L'ensemble des fonds des bibliothèques concernées représente 15 millions d'ouvrages, soit 4,5 milliards de pages Web. *Google*, malgré ces chiffres qui ont largement circulés, n'a jamais confirmé ni l'ampleur des investissements réalisés, ni celui des collections touchées. On peut être sceptique sur ces chiffres et reconnaître néanmoins que la firme dispose d'une capacité d'investissement considérable qui rend crédible une opération qui restera spectaculaire, même si son ampleur n'atteint pas les sommets décrits dans la presse.
- Par ailleurs, ici comme ailleurs, la firme montre un savoir-faire exceptionnel par sa précision et sa rapidité sur le découpage et le traitement des documents, autorisant des recherches sur le texte intégral d'une simplicité, fluidité et une pertinence inégalées à cette échelle. Quiconque peut s'en convaincre en navigant sur le site de test ouvert en avril 2005.
- Enfin, la question des droits était écartée par une interprétation de la philosophie nord-américaine de la propriété intellectuelle dont on connaîtra bientôt le degré de pertinence puisqu'une plainte est en cours d'instruction. L'idée primitive était de renverser l'ordre traditionnel. Un livre était considéré comme autorisé à être mis en ligne, sous forme d'extraits, sauf avis contraire des ayant-droits. Normalement une autorisation préalable est nécessaire, mais *Google*, comme il l'a toujours fait, joue le fait accompli en ajustant si nécessaire dans un second temps. Le pari ici était que les ayant droits verraient dans l'opération une opportunité de promotion lucrative, puisqu'ils étaient intéressés aux ventes, des ouvrages. En cas de refus, la firme s'appuyait sur la notion de *fair use* pour une consultation d'extraits. Le *fair use* autorise aux États-Unis la consultation de documents sans autorisation préalable dans un certain nombre de situations précises. Cette position n'est pas complètement originale. D'autres acteurs du Web l'appliquent aussi.

Les réactions

Passé l'effet de surprise et une fascination admirative, le projet a suscité réactions et oppositions vives. La première et la plus médiatisée fut celle de Jean-Noël Jeanneney, président de la

Bibliothèque nationale de France. Puis un certain nombre d'ayant droits ont contesté l'interprétation juridique de la firme. Enfin, des projets alternatifs sont annoncés, soit chez des concurrents, soit au niveau de la Communauté européenne.

Première charge : le Président de la Bibliothèque nationale de France

Lorsque François Mitterrand lance, le 14 juillet 1988, le chantier de l'établissement public national qui porte aujourd'hui son nom, il appelle de ses vœux : « une bibliothèque d'un type entièrement nouveau ». De cette utopie primitive la *Bibliothèque Nationale de France* ne semble pas avoir gardé grand chose, si ce n'est un remarquable service d'accès à des documents numérisés qui n'avait jusqu'à peu pratiquement pas d'équivalent au monde par l'ampleur des fonds mis en ligne sur un seul site : *Gallica*¹⁰.

Quinze années plus tard, on est en droit de se demander si *Google-Print* ne réalise pas le rêve du promoteur de la *BnF*. Entre *Gallica* et *Google-Print*, il y a, en effet, un saut radical dans le nombre de livres et dans le service (l'interrogation se fait sur l'intégralité des textes et non sur une notice). Mais l'histoire est ironique : la première contestation, vigoureuse, vient du président de la *BnF*, rapidement soutenu par le président de la République française. Tout d'abord dans une tribune du journal *Le Monde*, puis dans divers médias et même dans un petit livre¹¹, J.-N. Jeanneney a souligné le danger qu'il y aurait, selon lui, à laisser *Google* développer son projet sans alternative de ce côté de l'Atlantique.

L'argumentation est très gaullienne, dans le ton comme le fond : le projet *Google-Print*, mené par une firme californienne à partir de collections de bibliothèques anglo-saxonnes, renforcerait encore la domination de la culture et donc le point de vue américain sur le Web. Le raisonnement rappelle d'anciennes batailles et, dans son livre, l'auteur, lui-même historien des médias, l'illustre par des exemples pris dans l'histoire du cinéma ou encore dans la naissance de la télévision.

Quelle que soit la pertinence du raisonnement, l'important à noter est que l'objectif de l'auteur n'est pas d'arrêter le projet *Google-Print*, mais bien de susciter une alternative européenne. Il ne s'agit pas de se plaindre que *Google* numérise des millions de livres, mais de l'absence de projet équivalent en Europe, porté par des institutions publiques. Cet appel à un sursaut européen, même s'il suscite ici et là railleries ou contestations, n'est pas resté sans écho chez les décideurs.

Au cours du printemps 2005 successivement : le président de la République française lui a apporté son soutien en portant le débat devant les instances européennes ; dix-neuf bibliothèques nationales européennes ont signé une motion dans le même sens et présenté un service de recherche croisées sur leur catalogue¹², six États européens ont envoyé un courrier aux présidents de la Commission européenne et du Conseil, suggérant à l'Union européenne de prendre une initiative.

Le 13 juillet 2005, le ministre français de la Culture a mis en place un « comité de pilotage en vue de la création d'une bibliothèque numérique européenne » très français¹³. Et, dernière initiative

¹⁰ <http://gallica.bnf.fr/>

¹¹ Jeanneney Jean-Noël, *Quand Google défie l'Europe, Plaidoyer pour un sursaut*, Mille et une nuit, Paris 2005.

¹² 43 bibliothèques nationales participent au catalogue croisé ; <http://www.theeuropeanlibrary.org>.

¹³ Le Comité comprend : Olivier Cazenave (dir de la Documentation française), Franck Dangeard (pdg de Thomson), Olivier Duhamel (Prof IEP Paris), Serge Eyrolles (prés du Syndicat National de l'Édition), Jean-Michel Hubert (ancien prés de l'ART), Alain Kouck (pdg d'Éditis). Il est coordonné par A. Moatti.

au jour où nous écrivons ces lignes, le 30 septembre, la Commission européenne a lancé une consultation sur les bibliothèques numériques en prévision du lancement d'un programme¹⁴.

Seconde charge : les ayant-droits

Google a toujours présenté son service aux ayant-droits comme une opportunité d'augmenter leur vente d'exemplaires, et propose même une participation aux recettes de la publicité contextuelle pour les éditeurs. Ceux-ci, après avoir hésité quelques mois, vont contester à la fois la possibilité de numériser des livres sans autorisation préalable et l'interprétation du *fair use*.

Plusieurs courriers de protestation ont été envoyés par des représentants d'éditeurs américains à la firme au printemps 2005. *Google* a alors annoncé dans le courant de l'été la suspension partielle de son programme de numérisation pour la partie concernant les livres encore sous copyright, demandant aux éditeurs la liste des livres qu'ils voudraient exclure du programme. Dans le même temps, elle s'est adressée aux éditeurs européens pour leur proposer de numériser leurs publications, certains, peu nombreux jusqu'à présent, ont répondu favorablement. Des interfaces adaptées ont été ouvertes dans chaque pays en octobre¹⁵. Ces épisodes ont amené la firme à prendre en compte les différences de droits d'auteur selon le pays de commercialisation du livre et à demander un enregistrement des internautes pour la consultation.

Mais ce sont les auteurs qui vont attaquer en justice les premiers. En septembre 2005 une plainte est déposée à la cour fédérale conjointement par trois auteurs et surtout la puissante *Authors Guild* qui comprend 8.000 membres. Cette première salve a été rapidement suivie d'une seconde de l'association des éditeurs américains, *AAP*, représentant 300 maisons d'édition, qui a déposé plainte à son tour le 19 octobre contre *Google* pour avoir numérisé des livres sans autorisation des ayant-droits¹⁶.

Vraisemblablement, un feuilleton judiciaire démarre. Il sera suivi attentivement par tous les acteurs, aux États-Unis d'abord, mais aussi partout dans le monde car il fera jurisprudence. Les sujets de clarification ne manquent pas en effet. Au-delà des réglementations particulières, le droit de la propriété littéraire et artistique vise à résoudre un paradoxe des produits culturels entre leur caractéristique économique de « biens publics » et la nécessité de financer leur création et fabrication. Or le Web redistribue les cartes en accusant fortement les termes du paradoxe¹⁷. Quel que soit l'aboutissement judiciaire, il est peu probable que les pratiques réglementaires en vigueur dans le monde du papier puissent se décliner dans celui du numérique sans d'importants changements.

Illustration de ces difficultés de principes juridiques, la position française face aux droits d'auteur dans l'Internet contredit son interpellation de la firme *Google*. Le gouvernement français, en effet suivant une tradition ouverte depuis l'ouverture du Ministère de la culture par A. Malraux, a proposé la position la plus restrictive d'Europe sur la mise en ligne de documents dans

¹⁴ Commission des Communautés européennes, *i2010 : Bibliothèques numériques*, 30 sept 2005
http://europa.eu.int/information_society/activities/digital_libraries/doc/fr_comm_digital_libraries.pdf

La consultation est ouverte jusqu'au 20 janvier 2006 :

http://europa.eu.int/information_society/activities/digital_libraries/doc/annex2_fr.pdf

¹⁵ En réalité, la France, l'Italie, l'Allemagne, les Pays-Bas, l'Autriche, la Suisse, la Belgique et l'Espagne. Pour la France : <http://print-google.fr>.

¹⁶ <http://www.publishers.org/press/releases.cfm?PressReleaseArticleID=291>

¹⁷ Pour une présentation et discussion de ce paradoxe :

Curien Nicolas & Muet Pierre-Alain, *La société de l'information*, La Documentation française, 2004.
<http://www.cae.gouv.fr/rapports/dl/47.pdf>

l'application de la directive européenne sur les droits d'auteur en discussion au Parlement français¹⁸. Le Ministre de la culture s'est aussi réjoui publiquement de la réaction des ayants-droits américains à *Google-Print* qui, pense-t-il, conforte ses positions. La Commission européenne, pourtant, dans son document de réflexion sur les bibliothèques numériques, issu lui-même de l'interpellation française face à *Google-Print*, souligne qu'il est indispensable pour les projets de numérisation à grande échelle d'assouplir les règles de propriété intellectuelle, faute de quoi il sera plus coûteux de rechercher les ayants-droits que de réaliser le travail, un point sur lequel le consensus est très large.

Troisième charge les concurrents

La troisième réaction est celle des concurrents de *Google*. Nous ne citerons ici que les entreprises commerciales impliquées dans des projets comparables par l'ambition et par le type de services. Le projet de *Google* est une réelle menace pour eux, car il a un caractère monopoliste en interdisant, de fait, la création d'index concurrents sur le texte intégral des livres, sauf à réaliser un effort parallèle et redondant de numérisation.

On aurait pu s'attendre à une réaction du libraire en ligne *Amazon*. En réalité, celui-ci a ouvert à peu près au même moment mais beaucoup plus discrètement, un service de consultation en ligne des livres proposés à la vente¹⁹. Il s'agit d'extraits, l'accord des éditeurs est préalable et le lien avec la vente est manifeste. *Amazon* ne sort pas de son métier d'origine et l'ensemble du processus est sous le contrôle de la chaîne de production et distribution du livre.

Une autre réponse commerciale au projet *Google-Print* pourrait être une alliance d'acteurs, à but lucratif ou non, baptisée *Open Content Alliance*²⁰, menée par *Yahoo !*, le concurrent direct de *Google* sur la recherche d'information sur le Web, *Internet Archives*²¹, les bibliothèques des universités de Californie et de Toronto et bien d'autres encore.

À vrai dire même si le sérieux des acteurs lui donne une crédibilité certaine, le projet est encore trop flou pour qu'on puisse le juger. Il semble vouloir articuler des initiatives déjà existantes ou à venir de numérisation, une compétence en archivage et une compétence en recherche. Les arguments paraissent surtout construits en réaction à *Google-Print*. Ils insistent sur le respect de la propriété intellectuelle et sur l'importance de l'ouverture des outils.

Le renouvellement de la réflexion

Il y aura, à l'évidence, un avant et un après *Google-Print*. L'initiative et les réactions qu'elle a engendrées ont obligé les différents acteurs à se repositionner et la réflexion s'est brutalement accélérée.

¹⁸ Le vote doit se tenir en décembre 2005.

¹⁹ Le programme s'appelle « Chercher au cœur ». Présentation :

http://www.amazon.fr/exec/obidos/tg/browse/-/14210421/ref=br_lr_2/171-9177844-6142615

²⁰ <http://www.opencontentalliance.org/>

²¹ *Internet Archives* est un service qui archive en continu des pages Web depuis 1996. Accessible en ligne, il est devenu la mémoire du Web. Il lui aurait été impossible de demander l'autorisation préalable des ayants-droits. Il applique donc la même politique *a posteriori* que *Google*.

<http://www.archives.org>

Jusque là les projets et réalisations n'ont pas manqué et bien malin celui qui pourrait broser un panorama exhaustif des bibliothèques en ligne, tant l'activité est éclatée, foisonnante et hétérogène. On peut espérer que le débat actuel permettra de clarifier les situations.

La diversité des bibliothèques numériques

Selon les critères choisis, la notion de bibliothèque numérique désigne, en effet, des situations bien différentes :

- si on met en avant la notion patrimoniale traditionnelle, alors on considérera les documents numérisés. Toutes les grandes bibliothèques partout dans le monde ou presque ont mis en ligne un nombre plus ou moins conséquent de leur « trésor ». Bien d'autres acteurs, riches ou modestes, proposent aussi des documents de toutes sortes. Journaux, enluminures, photographies, livres anciens, collections historiques, de quelques volumes à des milliers, les documents patrimoniaux, accessibles sur les sites les plus divers, sont innombrables²².
- L'impératif patrimonial concerne aussi les documents numériques, et là encore d'immenses collections sont en ligne, comme celles de l'archivage du Web ou encore les photographies de la *NASA*²³.
- Si on prend le critère volumétrique (p ex en nombre de pages accessibles), alors incontestablement, à l'exception des précédents, les plus grosses collections se trouvent dans le monde des revues scientifiques ou chez les académiques, chez les éditeurs commerciaux ou non commerciaux, avec sans doute en première ligne *Elsevier*²⁴ et *Oclc*²⁵ parmi de très nombreux autres.
- Si on considère que la caractéristique principale est l'accès à des livres en ligne, alors on soulignera les librairies en ligne, comme *Amazon*, ou encore les initiatives collaboratives de mise en ligne de livres du domaine public.

Ces critères ne sont pas exhaustifs, ainsi chacun peut présenter des listes qui ne se recouvrent pas et il serait salutaire de construire un classement sur des bases consensuelles afin de pouvoir comparer et coordonner. Il est vraisemblable que beaucoup de moyens sont dépensés de façon redondante ou sans garantie de pérennité. Le débat qui se développe pourrait aider à clarifier les situations.

Un débat utile

Il faut donc remercier à la fois les dirigeants de *Google* et Jean-Noël Jeanneney d'avoir attiré l'attention des hommes d'affaires et celle des pouvoirs publics sur les enjeux culturels et commerciaux de la numérisation des livres. Il ne faut pas leur faire grief des hésitations de leur argumentation. Ils écrivent ensemble une histoire inédite dont ils ne sauraient maîtriser les conséquences et l'aboutissement. Nous n'avons vraisemblablement assisté qu'aux premiers épisodes d'un long feuilleton à venir.

²² Voir, par exemple, le catalogue réalisé par le Ministère de la culture français :

http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_02.htm

²³ <http://www.nasa.gov>

²⁴ <http://elsevier.com>

²⁵ <http://www.oclc.org>

Déjà, en moins d'une année, les avancées dans le débat et les questions portées au plus haut niveau des décideurs sont d'importance. En voici quelques-unes dans le désordre qui ne manqueront pas de rebondir prochainement :

- Jusque là réservée à quelques passionnés ou professionnels éclairés, la mise en ligne de la culture est devenue au travers de ce débat, pourtant circonscrit aux livres, un problème autant industriel que politique. Dans cette perspective quelle stratégie construire ? Avec quels acteurs ? Quelles priorités définir ?
- La numérisation des documents, autre que des livres homogènes, demandera des dispositifs autrement plus coûteux que ceux mis en place par *Google*. Déjà la Commission européenne insiste sur l'urgence plus pressante de la préservation du patrimoine audiovisuel dont les supports se dégradent très rapidement.
- La prise en charge de la numérisation par une société privée pose la question de la pérennité à long terme des copies numériques.
- La relation entre propriété intellectuelle et domaine public, dans ses dimensions spatiales, temporelles et dans son épaisseur est et sera rediscutée. Comment redonner aux éditeurs leur vocation première de soutien à la création en échappant à un simple raisonnement de rente où tout nouveau support est synonyme de rentrée supplémentaire pour des œuvres existants par ailleurs ? Mais aussi peut-on privatiser les index qui ouvrent l'accès à notre patrimoine culturel ? Ou encore ne doit-on pas, au minimum, disposer d'un contrôle démocratique sur les outils d'accès à la culture, comme il en existe sur les autres médias ?
- La relation entre les notions d'accès et celles de publication se brouillent au point que le président de la Bibliothèque nationale, qui a vocation à tout rendre accessible propose de sélectionner, tandis que l'opérateur privé, lui, collecte sans exclusive²⁶. Cet aspect se décline aussi sur le traitement linguistique : le Web est-il un seul grand livre ouvert dans lequel on navigue ou une série de documents autonomes ? Ou encore, n'assiste-t-on pas tout simplement à la première exploitation commerciale du modèle de la bibliothèque ?
- La réalisation d'une bibliothèque à l'échelle mondiale favorise-t-elle une culture dominante qui sera plus représentée et citée, ou au contraire des cultures minoritaires qui n'ont jamais disposé d'un tel outil de promotion ? Est-ce un instrument de conformisme ou une incitation à la curiosité ?

Ces questions, chacune essentielle à l'avenir de notre culture, appellent des réponses concrètes. Mais leur orientation fait référence à deux conceptions différentes sous-jacentes du rapport aux documents et à la bibliothèque, ou, plus pompeusement, de notre relation à l'universel se profilent.

L'une plus latine privilégie l'unité, la construction des contenus, la transmission des savoirs.

L'universel trouve sa source en amont dans ce qui est produit. La souveraineté, pour cette conception, signifie la préservation de ce patrimoine commun, symbolisé par le dépôt légal de tout document.

L'autre, plus anglo-saxonne, insiste sur la diversité, la priorité est l'accès. Les États-Unis n'ont pas de Bibliothèque nationale ; l'institution comparable est la Bibliothèque du Congrès, c'est-à-dire celle des représentants du peuple. L'Universel est en aval dans l'accès pour tous au patrimoine commun.

²⁶ « Google a pour mission d'organiser à l'échelle mondiale les informations dans le but de les rendre accessibles et utiles à tous » ainsi commence la présentation de la société sur son site. N'est-ce pas là précisément la mission d'une bibliothèque mondiale dédiée à l'intérêt général ?

Bien entendu, les deux conceptions sont complémentaires et l'important est de construire un équilibre satisfaisant entre l'une et l'autre, néanmoins chacune des sensibilités insiste plutôt sur l'un ou l'autre des deux pôles. On peut relire les débats précédents à la lumière de cette dialectique.

Dans le contexte de la mondialisation des législations et des pratiques juridiques, la rencontre entre les conceptions française et américaine du droit d'auteur pourrait avoir pour effet pervers de poser des barrières infranchissables à la réalisation d'une véritable bibliothèque universelle en ligne.

Aujourd'hui le numérique redistribue les cartes aussi bien du côté de la construction des contenus (numérisation dans notre exemple) que dans l'accès. Le Web souligne et déplace les différences de sensibilité, parfois jusqu'à la caricature : d'un côté, le pragmatisme et l'efficacité, dans une symbiose californienne, candide, décontractée et juvénile, entre Universités et Silicon Valley, surfant sur la multiplicité des comportements individuels ; de l'autre, la conscience et la planification prudente, dans un mélange d'esprit républicain et d'espoir européen, sérieux jusqu'au solennel, commission et compromis entre Bibliothèques nationales, acteurs traditionnels et institutions gouvernementales.

Ainsi, la différence de sensibilité entre nouveau et ancien monde débouche sur une divergence entre post-modernisme et modernisme. Il n'est pas sûr qu'il faille choisir entre l'un ou l'autre, mais il est sûr qu'il faudra réviser nos jugements et les équilibres.

Il est intéressant de constater que c'est au moment où un des acteurs principaux de la nouvelle culture, *Google*, s'est attaqué à un des principaux vecteurs de la culture traditionnelle, le livre, que le débat a, enfin, pu démarrer.