HAL
open science

L'information sur l'information, enjeu stratégique de la construction des savoirs à l'échelle mondiale

Michel Arnaud

To cite this version:

Michel Arnaud. L'information sur l'information, enjeu stratégique de la construction des savoirs à l'échelle mondiale. Distances et savoirs, 2008, 6 (1), pp.143-158. halshs-00637072

HAL Id: halshs-00637072

https://shs.hal.science/halshs-00637072

Submitted on 29 Oct 2011

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.

L'information sur l'information, enjeu stratégique de la construction des savoirs à l'échelle mondiale

Michel Arnaud

Université Paris X Nanterre
Laboratoire CRIS
UFR LL Phi
200 avenue de la République
92001 Nanterre Cedex
michel.arnaud@u-paris10.fr

RÉSUMÉ. *Les opérateurs des réseaux en contrôlent l'accès, les fournisseurs de contenus verrouillent leurs offres et les propriétaires des moteurs de recherche orientent les résultats des requêtes des internautes. Dans le cadre de la promotion du bien public informationnel, spécialement dans le domaine de la formation à distance, les pouvoirs publics doivent donc se donner les moyens de compenser les inconvénients de cette situation et de réguler la chaîne de production, d'indexation et de mise en ligne des contenus avec des archives ouvertes, des métadonnées interopérables, des moteurs de recherche efficaces mais conviviaux, des plates-formes de formation ouverte et à distance permettant de réutiliser les contenus afin de garantir des procédures d'apprentissage optimisées pour faciliter la construction de connaissances de tout un chacun. Le but de ce texte est d'examiner comment et à quelles conditions.*

ABSTRACT. *Network operators control users access as content providers do the same with their offers as well as search engines owners monitor users search results. In order to promote informational public good, especially in distance education, public institutions have to regulate the entire content production and indexation chain as well as online access with open archives, interoperable metadata, efficient but user-friendly search engines, learning management systems allowing to reuse content so that learning process is optimized to facilitate knowledge for all. The aim of this text is to see how it could be possible and under which conditions.*

MOTS-CLÉS : *accès gratuit au savoir, marchandisation, archives ouvertes, métadonnées, compétences documentaires, bien public informationnel.*

KEYWORDS: *free access to knowledge, privatization, open archives, metadata, competences in information retrieval, informational public good.*

Introduction

Les projets visant à réduire la fracture numérique reposent sur l'idée selon laquelle la connexion aux réseaux est un moyen privilégié d'offrir à chaque individu la chance et l'occasion de compenser ses lacunes éducatives et culturelles par l'accès aux informations disponibles sur Internet. D'autres aspects, tels que la gratuité des contenus disponibles et la possibilité d'échanger et de collaborer avec d'autres internautes, contribuent à renforcer l'image de proximité, d'immédiateté et de liberté qui s'attache à Internet. Telle n'est pourtant pas la réalité : les opérateurs des réseaux contrôlent l'accès, les fournisseurs de contenus verrouillent leurs offres et les propriétaires des moteurs de recherche orientent dans le sens qui les intéresse les résultats aux requêtes des internautes.

L'hypothèse qui est à l'origine de ce texte est que, souvent oubliés, les dispositifs de médiation accompagnant et orientant la démarche de l'internaute en quête d'information constituent un maillon essentiel de la démocratisation ou de la commercialisation de l'accès aux connaissances. Il s'agit aussi — et la coïncidence n'est pas fortuite — d'un aspect essentiel du processus de construction des connaissances, et, à travers cet aspect, de la question capitale du rôle des pouvoirs publics en faveur de la liberté d'apprendre. C'est donc de ces dispositifs, de la manière dont ils sont conçus et des manières de les concevoir pour les dégager, au moins partiellement, des stratégies commerciales qu'il sera question ici.

Directement concernés, les enseignants se trouvent en première ligne, à cet égard, aux côtés des apprenants. Mais le rôle des documentalistes n'est pas moins essentiel, du fait de la mission qui leur est confiée de former les élèves à la recherche d'information sur Internet. Le B2i (Brevet informatique et internet) et le C2i (Certificat informatique et internet) témoignent, certes, de la volonté de l'institution éducative de donner aux internautes les capacités dont ils ont besoin pour être autonomes. Toutefois l'attention devrait également se porter sur l'offre d'information et les conditions de sa mise à disposition. Les modalités des archives ouvertes pour les chercheurs sont, par exemple, une réponse intéressante à la revendication en faveur de la liberté d'accès aux connaissances scientifiques. Elle est même si intéressante que beaucoup, au sein des communautés scientifiques et éducatives, pensent que cette ouverture devrait être étendue à tous les types de public, par-delà les seuls chercheurs.

À propos de la diffusion et de la commercialisation des programmes et services éducatifs à l'échelle internationale, il est souvent question de la dérégulation, mise à l'ordre du jour par les négociations autour de l'Accord général sur le commerce des services de l'Organisation mondiale du commerce. Cette question est importante effectivement, car le poids des opérateurs privés les conduit à faire des propositions liées à leurs propres impératifs de rentabilité, au détriment des besoins d'accès des publics défavorisés. Non moins importante, pourtant, la négociation autour des outils techniques et du choix des dispositifs traite d'un aspect souvent négligé : la

normalisation des conditions de l'accès au savoir en ligne dont il est clair qu'elle aura une fonction structurante pour l'avenir.

La marchandisation des savoirs

Le rapport au savoir a changé : de ce changement, il faut dire quelques mots avant d'entrer dans le vif du sujet.

Dans la société dite "de l'information", les flux de production des savoirs reposent, en effet, sur des chaînes intégrant des travailleurs intellectuels membres d'équipes où l'information est traitée et transformée, dont la propriété est protégée par la prise de brevets (brevetabilité du vivant, biotechnologies, génie logiciel, etc..). Au sein de telles équipes, la notion d'auteur s'estompe, et le droit d'auteur n'est plus reconnu au salarié qui, pourtant, vend sa force de travail intellectuel à son employeur.

En outre, la nouvelle hiérarchie des savoirs fait que ceux qui débouchent sur des brevets sont placés au sommet tandis que les savoirs sociaux et culturels occupent les places inférieures. La valeur des savoirs, en effet, est fixée en fonction de leurs coûts de production, mais aussi et surtout selon les critères de l'intelligence économique, en fonction de leur importance stratégique pour les individus, les entreprises et les pays concernés[1]. La gestion des compétences et des connaissances au sein de l'entreprise établit les règles de promotion et de contrôle des salariés en fonction des objectifs fixés.

La question de l'employabilité des individus, c'est-à-dire du caractère opérationnel de leurs savoir faire en termes de production de plus-value, se concentre sur leurs capacités à recueillir l'information stratégique et à la traiter en vue de l'élaboration d'un savoir susceptible de donner naissance à un brevet. Le suivi fonctionnel des individus s'opère en fonction des besoins en compétences et connaissances des entreprises où ils sont ou bien de celles où ils souhaitent entrer. La mise à jour des compétences professionnelles relatives à la construction de savoirs sur commande devient donc une nécessité pour garder son emploi ou en obtenir un plus intéressant.

Coûts des savoirs à la carte

Du côté des organismes de formation, il s'agit, par conséquent, de répondre le plus précisément possible aux besoins de l'apprenant et de lui procurer les contenus demandés « juste à temps » : en fonction des compétences recherchées, les séquences de formation les mieux adaptées lui sont proposées. Aussi, par exemple,

[1] Sennett R., *La Culture du nouveau capitalisme*, Albin Michel, Paris, 2006, 160 p.

l'Association nationale pour la Formation Professionnelle des Adultes (AFPA) a-t-elle un référentiel emplois activités compétences indexé sur le Répertoire Opérationnel des Métiers et des Emplois (ROME) de l'Agence Nationale pour l'Emploi (ANPE), permettant de segmenter les contenus de formation en fonction des capacités à acquérir.

Une telle organisation ne va toutefois pas sans problèmes : les grains de contenus sont la plus petite unité correspondant à la séquence de formation ; agrégés, ils constituent un module, qui sera lui-même ajouté à d'autres pour constituer l'offre de formation qualifiante. Dès lors, la question de la cohérence des grains, les uns par rapport aux autres, se pose avec d'autant plus d'acuité au fur et à mesure de leur combinaison, en particulier lorsqu'il s'agit de les réutiliser dans un autre référentiel. Les coûts liés à la conception et à la mise en ligne de contenus peuvent également être importants, si ces contenus sont de qualité, ce qui conduit les coûts à être répercutés sur les frais de scolarité. Dès lors, et du fait de la concurrence entre les organismes de formation professionnelle, la maîtrise des coûts est devenue une nécessité. De là vient aussi que les impératifs de réutilisabilité et d'interopérabilité s'imposent plus que jamais aux concepteurs de formation.

À cet égard, le monde de la formation à distance se partage *grosso modo* en deux ensembles distincts : d'un côté, les organismes disposant des moyens nécessaires, adoptant une structure industrielle et faisant payer aux usagers les sommes indispensables à la rémunération de l'ensemble des contributeurs à la chaîne de production et de diffusion de contenus multimédias : constructeur de terminaux, éditeurs de systèmes d'exploitation, opérateurs télécoms, intégrateurs d'applications et de services (portail), fournisseurs de contenus, régie publicitaire. La sélection par le pouvoir d'achat joue évidemment un rôle déterminant dans ce scénario.

De l'autre côté, on trouve des organismes ne disposant pas des moyens techniques et financiers nécessaires pour élaborer une offre riche et diversifiée. *A fortiori* les techniques de production et de granularisation des contenus échappent-elles à l'univers de l'enseignant traditionnel, habitué à utiliser des manuels scolaires ou à réaliser lui-même ses supports de formation en fonction d'objectifs liés aux usages dans la classe. Cet enseignant n'aura à sa disposition aucun outil de conception facile à utiliser pour mettre ses cours en ligne, et le coût de la simple mise en ligne a déjà un coût d'autant plus important que ce sont des moyens industriels qu'il faudrait mobiliser pour créer et diffuser ses contenus.

Or, les offres de formation en ligne réalisées à moindre coût connaissent des taux d'abandon élevés, faute de tutorat suffisant et en l'absence d'outils de suivi performants. En revanche, le modèle haut de gamme de formation à distance correspond à une stratégie gagnante, mais il consiste à mettre en ligne des offres à forte valeur ajoutée, dont les contenus sont protégés par des *copyrights* imposant aux utilisateurs des frais de scolarité élevés. Une telle dichotomie dans le champ de la formation à distance n'est évidemment pas satisfaisante ; elle l'est encore moins

quand elle s'étend à l'international, justifiant la revendication en faveur de l'accès libre.

Contrôle de la diffusion des savoirs en ligne

Le combat des partisans de l'accès libre aux contenus en ligne a été engagé par Lessig, le créateur de *Creative Commons*, ainsi que par Stallman, à l'origine de la licence d'exploitation libre GNU. Ces spécialistes se réfèrent aux valeurs de partage des savoirs qui ont caractérisé l'Internet originel et ils s'élèvent contre les tendances actuelles du commerce électronique, de la propriété intellectuelle et de la brevetabilité de l'immatériel[2].

La question est toutefois de savoir si le débat sur le téléchargement pirate n'a pas occulté les vrais enjeux autour de l'interopérabilité. Les DRM (*Digital Rights Management* : gestion des droits numériques) ont, en effet, une autre fonction, en plus de celle de faire payer les consommateurs : l'usage des DRM permet de contrôler les clients et d'établir des monopoles dans la distribution des œuvres en refusant toute interopérabilité avec des systèmes de distribution concurrents. La mise en œuvre des DRM correspond donc à une stratégie de préemption du marché, et elle risque de conforter la position dominante de quelques firmes, issues des industries de contenus, des télécommunications ou de l'informatique, au détriment des auteurs et des petites structures n'ayant pas les moyens de négocier des catalogues de droits. À plus forte raison les consommateurs sont-ils lésés, cantonnés comme ils le sont à une offre spécifique.

Les développements proposés par les entreprises informatiques pour contrôler les ordinateurs des consommateurs afin de vérifier le paiement des droits d'accès sont donc préoccupants. De fait, ils restreignent la liberté d'accès de l'usager grâce à l'informatique dite « de confiance » qui consiste, en réalité, à contrôler les usages privés de l'ordinateur par logiciel interposé (*Next Generation Computing Base*) de Microsoft et par des puces électroniques incluses dans l'équipement informatique (*Trusted Computing Group*).

Gratuité des savoirs assurée par la publicité

L'offre de contenus gratuits sur Internet est intéressante, car elle semble remettre en question le principe de la sélection par l'argent pour l'accès aux savoirs. Coexistent, en réalité, sur Internet deux types de savoirs gratuits : ceux générés par les communautés d'internautes bénévoles tels que *Wikipedia* et ceux dont le coût de

---

2 Lessig L., *L'Avenir des idées. Le sort des biens communs à l'heure des réseaux numériques,* PUL, Lyon, 2005, 414 p.

production est pris en charge par la publicité. Les deux options ne sont naturellement pas à mettre sur le même plan.

Dans ce cas du financement publicitaire, les savoirs offerts gratuitement le sont en vue d'inciter le consommateur à acheter un produit ou un service annexe. Toutefois, derrière cette incitation, apparemment innocente, les techniques et réseaux de communication permettent aux opérateurs de suivre à la trace les agissements des consommateurs dans leurs usages des téléphones portables, des cartes de crédit, d'Internet. Des profils sont établis en fonction du pouvoir d'achat et des goûts et préférences de chacun, afin de mieux cibler les offres de produits et de services.

Ce qui se passe avec la vidéo à la demande est significatif : une infime partie de l'offre est payante, tout le reste est gratuit parce que financé par la publicité. Il est donc parfaitement envisageable de faire payer par la publicité la mise en ligne de contenus de formation si des achats de produits et de services supplémentaires viennent la compléter. Cependant, une telle option serait lourde de conséquences. Dans ce contexte, en effet, la question de la qualité des savoirs ainsi mis en ligne se pose, celle de la liberté du consommateur également. Quant au risque d'assister à une segmentation de l'offre de savoirs en fonction des profils de consommateurs visés, il est énorme.

Significatif est, à cet égard, le projet *Google Print*, qui concerne la mise en ligne intégrale de 15 millions de livres libres de droit, en provenance des bibliothèques des plus prestigieuses universités américaines (Harvard, Stanford et Michigan), de la *Bodleian Library* d'Oxford et de la *New York Public Library* , soit 4,5 milliards de pages en six ans. En outre, devenant le service Google Recherche de Livres (http://print.google.com/), le projet s'est diversifié et enrichi : il propose à l'internaute de lire des extraits, de parcourir certains paragraphes et d'acheter en ligne en partenariat avec Amazon et d'autres distributeurs en ligne comme Yahoo et son projet de bibliothèque numérique OCA (*Open Content Alliance*). Le résultat est double : non seulement les sites américains de livres ou de bibliothèques sont privilégiés, au détriment de ceux d'autres cultures, moins référencés, donc moins cités dans les requêtes, mais surtout le service Google Recherche de Livres contribue fortement à la privatisation de l'accès au savoir, avec ses partenaires privilégiés Amazon et Yahoo, autour desquels se rassemble la quasi-totalité des éditeurs de livres de langue anglaise. La *Google Teacher Academy*, qui regroupe des enseignants dont les vidéos sont mises en ligne, et *Google Scholar* (http://scholar.google.fr/), qui a fait alliance avec Elsevier pour distribuer ScienceDirect, répondent à la même logique.

La publicité peut-elle donc remplacer le financement public pour l'offre de contenus en ligne ? Si la publicité crée un régime artificiel de gratuité, il y a de fortes chances pour qu'elle renforce aussi les disparités sociales, privilégiant la segmentation des consommateurs selon leur pouvoir d'achat.

Inégalité en matière d'accès aux savoirs en ligne

La question se pose également sur un autre plan : celui, cognitif, des conditions d'accès au savoir. Si l'on considère que l'apprentissage marque le passage de la compétence, comme savoir identifié, à la capacité, opération mentale stabilisée et reproductible dans des champs divers de connaissances, l'on voit alors que cet apprentissage est provoqué par un décalage entre l'information nouvelle et la représentation existante[3] ; ce décalage force, en effet, l'apprenant à réorganiser son système de représentations, et l'oblige à la reconstruction de ses connaissances[4]. Meirieu, après bien d'autres, note le rôle essentiel que l'enseignant joue dans le triangle pédagogique (apprenant, savoir, enseignant) pour accompagner le mouvement de décrochage et de recentration de l'apprenant[5]. Or, la présence des outils d'information et de communication vient bouleverser ce triangle, en introduisant un quatrième pôle, intercalé de manière non transparente, entre l'apprenant et le savoir ainsi qu'entre l'apprenant et l'enseignant quand il s'agit d'un échange à distance. Ce phénomène n'est pas sans répercussion sur les conditions de l'accès au savoir.

Procédures d'accès aux savoirs en ligne

De fait, l'accès aux données, et aux documents en ligne en particulier, modifie profondément la manière d'apprendre. Trouver des savoirs sur Internet implique trois compétences : prendre conscience du besoin d'information, rechercher les informations pertinentes, les exploiter.

La notion de besoin d'information est paradoxale : pour qu'un sujet prenne conscience qu'il manque de connaissance, il a besoin de connaissances sur qui lui fait défaut, sous forme de méta-connaissances. Le sujet en question peut ne pas savoir de quelles informations il a besoin ou penser que ses questions n'ont pas de réponse. La certitude tout comme l'incertitude, liée à l'absence de connaissances, supprime le besoin d'information. Les méta-connaissances servent donc à aiguiller l'apprenant vers les champs de connaissances à explorer. L'apprenant peut également ne pas savoir où se trouve l'information une fois qu'il a réalisé qu'il en a besoin ou ne pas avoir connaissance de l'existence des sources d'information ou de ses propres difficultés en recherche d'informations. La recherche nécessite, en effet, des compétences documentaires. Quant à l'exploitation des résultats, elle est la dernière étape : il s'agit d'évaluer les résultats, de les trier et de raffiner éventuellement les

---

[3] Piaget J., J., *La Psychologie de l'intelligence,* Armand Colin, Paris, 1967, 234 p.

[4] Perret-Clermont A.-N., *La Construction de l'intelligence dans l'interaction sociale*, Peter Lang, Berne, 1979, 256 p.

[5] Meirieu P., *Apprendre...oui, mais comment*, Collection Pédagogies, ESF Editeur, Paris, 1999, 192 p.,p 93-95

termes de la requête pour la relancer en réintroduisant de nouveaux mots-clés ou en utilisant des opérateurs booléens par exemple.

Or, l'expérience montre qu'il est illusoire d'imaginer que la pratique d'Internet puisse, sans expérience antérieure, développer la maîtrise de schèmes complexes tels que la déduction, l'induction, l'abduction et la sérendipité chez la majorité des internautes. La complexité des processus cognitifs impliqués démontre, si besoin en est, l'erreur consistant à croire qu'il est facile pour l'internaute de mobiliser ce type de schèmes mentaux. En réalité, la majorité des apprenants se contente d'utiliser les schèmes les plus simples lorsqu'elle surfe sur Internet, pour des résultats décevants concernant la construction des connaissances en autonomie et en équipe collaborative.

Google comme outil de recherche tronquée

Que se passe-t-il donc dans le cas où un jeune apprenant se trouve confronté sans médiation pédagogique à des savoirs et ressources sur Internet ? Tricot note que « les enfants et adolescents évaluent moins et moins bien les sources, utilisent moins de critères d'évaluation de la pertinence, utilisent les images pour évaluer la précision et la validité, recherchent des résultats exploitables (que l'on peut copier-coller par exemple), sont trop sensibles aux marqueurs paralinguistiques[6] ». Les enseignants constatent effectivement la récurrence du couper-coller, consistant à reprendre, par exemple, des paragraphes entiers de *Wikipedia* et à les insérer tels quels dans les devoirs à rendre. Si cette pratique est combattue par les enseignants capables d'aller retrouver les pages copiées en insérant dans un moteur de recherche quelques phrases clés, il n'en reste pas moins qu'elle dénote un décalage entre l'effort nécessaire pour apprendre et le climat de facilité propice à la consommation proposée par la publicité.

L'usage des moteurs de recherche grand public repose sur un compromis imposé par leurs concepteurs en fonction d'impératifs commerciaux. Une majorité écrasante d'internautes utilise Google et ne lit que la première page des résultats d'une requête à partir d'un ou de deux mots clés. D'après Jean Véronis, l'index de Google serait constitué de 8 milliards de pages Web référencées automatiquement soit 60% de l'ensemble des premières pages de sites Web reconnues et classées par les robots de recherche à partir du calcul du nombre de visites, la prime étant donnée aux sites les plus visités selon la technique du *Page Rank* [7]. Les sites bénéficiant du plus grand nombre de visites par jour paraissent en haut de la liste, c'est-à-dire sur la première

---

[6] Tricot A., « L'utilisation d'Internet permet-elle de nouveaux apprentissages documentaires ? », *Colloque Politique documentaire des EPLA et ressources numériques*, CRDP Amiens, novembre 2006

[7] Véronis J., *Moteurs : Folles duplications (1,)* Technologies du langage, Actualités, Commentaires, Réflexions, juillet 2006, http://aixtal.blogspot.com/

page de références lue par l'internaute en priorité. L'ordre peut être modifié à la demande expresse de clients payants pour que soit effectuée cette manipulation au profit de leur site par les ingénieurs de Google.

Le raisonnement des créateurs de l'algorithme de recherche de Google se place dans la logique des fournisseurs d'accès qui est d'accroître les flux de données se répercutant sur le chiffre d'affaires. Dans ce contexte où la priorité est à la gestion des flux et non pas à la précision des retours d'information, le bruit généré par les requêtes dans Google est phénoménal. De fait, les sites ramenés en référence se chiffrent en milliers. Les pages à lire par l'internaute consciencieux s'empilent les unes après les autres sans apporter plus de documents pertinents par rapport à la requête initiale.

Malgré tous ses défauts, force est de constater la résilience de cette démarche basée sur les pages les plus consultées et qui renvoie à une approche monocorde de la culture dominante de base. Le prix à payer en termes d'autonomie dans la construction de connaissances est élevé, cependant, car les résultats de la requête sont biaisés par des considérations d'ordre commercial. Seuls, des moteurs de recherche spécialisés sont capables de répondre à des demandes d'information formulées par des experts. La recherche d'information pointue (*data mining*) correspond à une expertise avancée et à un savoir haut placé dans l'échelle des compétences monnayables dans la société de l'information.

Le hiatus se creuse donc entre les pratiques du grand public et celles des spécialistes, entre les savoirs communs gratuits et les savoirs savants chèrement monnayés. L'Organisation de coopération et de développement économiques (OCDE) elle-même prédit l'avènement d'un système éducatif à deux vitesses comprenant : une partie payante destinée aux classes aisées, faisant accéder, de l'école et du domicile, à des savoirs savants, et, d'autre part, une partie publique et gratuite, destinée aux apprenants des classes défavorisées, dans laquelle seuls les savoirs communs seront disponibles.

Les compétences documentaires à développer

Dans le cas de la recherche avancée d'information sur Internet, les compétences mobilisées sont pointues et appartiennent au champ professionnel de la documentation : il s'agit de maîtriser la combinaison des mots clés, l'usage des thésaurus et des listes d'autorité, avec le plus souvent l'apprentissage d'un moteur de recherche spécialisé.

Les cinq profils d'apprentis chercheurs identifiés par Halttunen correspondent à ses observations d'élèves faisant une recherche sur Internet : *l'identificateur du processus* qui centre sa démarche sur les premières étapes de la recherche d'information à savoir la définition du besoin d'information, *l'identificateur des sources* capable d'évaluer leur pertinence par rapport au besoin, *le chercheur* apte à

manipuler les moteurs de recherche, *le formalisateur de problème* sachant formuler la requête, *l'évaluateur* pouvant valider les résultats[8]. Le problème est qu'aucun de ces profils ne couvre l'ensemble des étapes nécessaires pour mener à bien une recherche d'information sur Internet. La solution consiste donc soit à former des équipes composées d'élèves ayant ces différents profils soit à les accompagner un par un, en les encourageant à changer de profil au fur et à mesure de leur progression.

L'enseignant, bien sûr, n'a pas vocation à devenir un documentaliste, pas plus que l'apprenant moyen. Pourtant la présence de l'enseignant est quasi indispensable auprès d'apprenants non autonomes. Les savoir faire qui se construisent par l'intermédiaire des techniques documentaires impliquent donc un changement de statut de l'enseignant, devenu personne-ressource face à une information plus abondante et moins sécurisée que dans la salle de classe. C'est aussi à l'enseignant qu'il revient d'apprendre à transformer l'information en une connaissance appropriée. Dans ce contexte, les savoirs procéduraux deviennent aussi importants que les savoirs déclaratifs car les premiers permettent d'accéder aux seconds.

Les compétences documentaires reposent d'ailleurs sur des méthodes dont l'aspect formateur, bien que souvent mal perçu, favorise l'acquisition de savoir faire que l'on retrouve tout au long de la vie, pour rechercher l'information, l'analyser, la comprendre, la classer, la réutiliser. Si le B2i et le C2i permettent de délivrer des certificats concernant l'acquisition des connaissances procédurales nécessaires, le saut qualitatif entre moteurs grand public et moteurs spécialisés reste très important. Or, force est de constater qu'il y a un manque dans la palette des moteurs de recherche offerts aux enseignants et aux apprenants : font défaut des dispositifs qui leur permettraient de se familiariser avec les procédures de recherche avancée et leur ouvriraient l'accès aux savoirs de manière précise, efficace et libre.

Les bases d'une politique publique d'accès égalitaire au savoir

Le bien public est par définition non compétitif et offert à tous[9], le rôle de la puissance publique étant d'en garantir l'accès égalitaire. L'enjeu de l'accès au savoir, de sa production et de son mode de diffusion, est donc central car l'accès en question détermine la capacité d'innovation en matière de recherche et développement de pans entiers de l'humanité. De là vient l'importance d'une politique publique adaptée.

---

[8] Halttunen K., "Students' conceptions of information retrieval: Implications for the design of learning environments". *Library & Information Science Research*, 25, 2003, pp. 307-322

[9] Stiglitz J., *La grande désillusion*, Poche, Paris, 2003, 407 p

Mise en place d'archives numériques ouvertes

La mise en place de bibliothèques et d'archives numériques ouvertes, offrant un accès libre aux informations, bénéficie de l'argent public et repose sur un compromis avec les producteurs et éditeurs privés. Les initiatives telles que l'*Open Archive* sont intéressantes, à cet égard, et les portails *Persée*, *revues.org* en France et *Erudit* au Canada ont leur correspondant dans le portail des éditeurs privés, *Cairn*.

Au lieu du paiement à la demande, les modalités de la licence globale représentent une option de rechange : une redevance mensuelle est prélevée sur les postes d'accès et répartie ensuite en fonction des principales sources d'information utilisées. Des mesures de promotion du *copyleft*, par exemple, peuvent concilier la volonté de l'auteur de partager les contenus qu'il a créés et le principe d'une redevance modique. Les contenus jugés les plus utiles sont préalablement achetés par les pouvoirs publics pour être mis en ligne gratuitement. Les modalités d'achat par des organisations telles que l'UNESCO de bases de données utiles à des populations défavorisées sont intéressantes en ce qu'elles concilient dédommagement des auteurs et consultation gratuite.

Le rôle stratégique des métadonnées

De nombreuses ressources numériques publiques destinées à des usages pédagogiques sont produites par des enseignants, des chercheurs, des animateurs TICE (Technologies de l'information et de la communication pour l'éducation), des inspecteurs de l'Éducation Nationale, les services culture éditions ressources pour l'Education Nationale (SCEREN) du réseau du Centre national de documentation pédagogique (CNDP). Ces ressources sont disséminées sur des sites Web institutionnels à tous les niveaux (national, académique, départemental, local), associatifs et personnels. Elles constituent un corpus riche mais foisonnant et hétérogène de ressources en ligne qui restent méconnues si un système commun d'identification et de repérage n'est pas mis en place.

Il n'est pas suffisant, en effet, de créer et de déposer des contenus en ligne : il s'agit aussi de permettre de les retrouver pour qu'ils soient utilisés à volonté. Afin que les moteurs de recherche identifient rapidement et avec la plus grande efficacité possible les ressources disponibles sur un sujet donné, il faut donc les indexer avec des métadonnées qui contiennent un schéma de description commun. Un compromis particulièrement délicat est alors à trouver entre précision et rapidité.

Google, pour sa part, a fait le choix de la rapidité, faisant indexer automatiquement les pages scannées en texte intégral par ses robots. Les conséquences en sont une pléthore de résultats à toute requête et, ainsi que cela a été indiqué plus haut, l'importance démesurée accordée à la première page. À l'opposé, les instances publiques enrôlent des indexeurs manuels qui remplissent méticuleusement les champs des métadonnées, mais au prix d'un temps excessif de

travail. Quant à l'utilisation des *tags*, pour répondre à la nécessité d'identifier immédiatement des ressources, elle repose sur la définition des descripteurs par les communautés de concepteurs et d'utilisateurs, mais elle n'est pas exempte du risque de manque d'uniformité sémantique d'un ensemble à l'autre.

Par ailleurs, un non documentaliste a du mal à remplir les champs des métadonnées, les bases de ressources pédagogiques telles *qu'ARIADNE, ELENA, EDUTELLA* étant basées sur le format LOM et demandant beaucoup d'investissements en temps d'indexation. En outre, la compatibilité des profils d'application nationaux des métadonnées basés sur le format LOM n'est pas acquise : les différences sont importantes entre les définitions des champs et surtout s'agissant des contenus des thésaurus et des listes de termes de référence permettant de les remplir. Si, en effet, chaque profil national d'application du LOM compte entre 15 et 25 champs obligatoires, ce ne sont souvent pas les mêmes si on se réfère à la grille de référence globale. Cette difficulté a conduit le consortium européen EUN (*European Schoolnet*) à définir 13 champs obligatoires pour le profil d'application de métadonnées européennes CELEBRATE. Les éditeurs de métadonnées utilisés le plus communément (*Lompad, Formist, Ariadne, Reload* par exemple) se contentent de respecter les différents formats des profils d'application du LOM et proposent des passerelles entre eux pour les conversions.

Dans le cadre du développement du format français LOMFR, l'effort porte sur l'aide à l'édition des métadonnées et sur l'indexation des champs correspondants. Un premier travail a été réalisé par le SCEREN CNDP en partenariat avec le laboratoire CRIS (Centre de recherche en information spécialisée et médiation des savoirs) de l'Université Paris X Nanterre, sous forme d'un éditeur permettant à un enseignant, par exemple, d'indexer lui-même des ressources pédagogiques et de les rendre compatibles avec la base de ressources éducatives du CNDP : *Educasources*. Ce développement ne se limite pas à un convertisseur de formats de métadonnées (LOM, LOMFR, Dublin Core, *Educasources)* mais il porte aussi sur les procédures permettant de remplir les champs des métadonnées de manière semi-automatique. En effet, du point de vue d'un enseignant qui n'a pas le temps matériel de remplir ces champs pour tous les documents à caractère pédagogique qu'il met en ligne, un éditeur de ce type, basé sur des procédures de reconnaissance sémantique, est d'une grande utilité.

La création semi-automatique des métadonnées évite de passer par l'étape fastidieuse du remplissage des champs des LOMFR à la main, ou, du moins, elle allège considérablement la procédure en proposant des solutions à valider. Pour ce faire, il s'agit d'identifier les meilleurs algorithmes d'analyse de texte pour extraire du document étudié les mots clés les plus significatifs. Un autre aspect concerne l'utilisation des références bibliographiques et des thésaurus permettant de choisir les termes les plus significatifs à placer comme métadonnées. Enfin, les profils d'utilisateurs concernés dans la chaîne de production des métadonnées (concepteur, enseignant, indexeur, administrateur) sont mis à profit pour observer les pratiques

usuelles et définir des règles d'inférence[10]. Le prototype d'outil de création semi-automatique mis au point par le SCEREN préfigure les développements à venir tendant à faciliter l'indexation des ressources éducatives, dans un premier temps, par les documentalistes et, dans un deuxième temps, par les concepteurs de ressources eux-mêmes.

L'enjeu de l'efficacité des moteurs de recherche

Un bon outil de recherche permet d'écrire des requêtes de recherche variées et puissantes et de présenter les résultats de façon claire et utile pour les utilisateurs. L'outil de recherche doit alors tenir compte non seulement du texte des articles mais aussi de leurs métadonnées ainsi que de la structure des documents et des nuages de tags qui leur sont associés. La manière d'utiliser les termes de la requête permet de restreindre ou d'élargir les critères de recherche, par exemple de chercher des mots seulement dans les sujets ou les auteurs ou les titres de sections, mais aussi d'influencer le tri par ordre de pertinence des résultats.

Pour la présentation des résultats, la situation est plus complexe. Le mode le plus simple consiste dans le tri des documents par ordre de pertinence, une fonction intéressante des outils de recherche consistant à mettre en évidence les mots recherchés dans les documents pour faciliter la reformulation de la requête afin de restreindre ou élargir le champ envisagé. Le compromis à trouver entre précision de la requête et complexité de sa formulation nécessite le développement d'interfaces conviviales visant à aider l'utilisateur débutant. Cet objectif est loin d'être anodin puisqu'il contribue, en complément des autres aspects relatifs aux contenus gratuits mis en ligne et aux formats ouverts de métadonnées, à la facilitation de l'accès aux savoirs en ligne.

Mise en place de dispositifs publics de formation en ligne

Pour réduire la fracture cognitive, les solutions proposées vont de l'offre de formations en ligne individualisées, avec des procédures de validation des acquis de l'expérience et des outils de suivi et d'évaluation, à l'accès facilité à des bases de connaissances, à des efforts de normalisation visant une interopérabilité accrue entre plates-formes, avec les environnements numériques de travail (ENT) en France par exemple. Des lieux de formation adaptés à ces nouvelles configurations d'apprentissage sont proposés, comme les points d'accès à la téléformation (P@T). Des projets de formations en ligne sur téléphone mobile sont développés comme *IDF Virtual* du pôle de compétitivité Cap Digital en Ile-de-France. De nouvelles formes de partage des connaissances et d'apprentissage collectif sont rendues

[10] Gomez de Regil R.-M., Cahier des charges de l'outil de création semi-automatique de métadonnées, Projet PPRE, CRIS UPX-SCEREN CNDP, 2007, 40p.

possibles par les technologies de syndication de production d'information : flux RSS, d'agrégation coopérative de contenus comme *Wikipedia*.

Le statut de la formation comme bien public informationnel justifie la volonté de contrecarrer les tendances marchandes qui font que, seules, les tranches supérieures d'utilisateurs bénéficient de l'accès à un savoir en ligne de qualité. *A contrario*, les normes ouvertes garantissent l'interopérabilité et la multiplicité des logiciels d'accompagnement des démarches d'accès. Les plates-formes d'apprentissage en ligne en logiciels libres se sont développées ; le respect de standards *de facto*, tels que SCORM 2004, renforce l'interopérabilité et la fiabilité de ce type de plates-formes gratuites. En ce qui concerne les interfaces normalisées, la notion de bien public informationnel s'articule avec la mise à disposition du code à titre gratuit aussi bien que de la documentation qui le décrit, de telle sorte que les développeurs de tous les pays du Nord et du Sud puissent tirer profit de cette base commune logicielle. Précisons d'ailleurs que ce type de démarche n'est pas incompatible avec l'économie de marché ; elle favorise au contraire une distribution d'expertise dans le monde entier.

Les perspectives européennes

Le projet de bibliothèque européenne à l'instigation de Jean-Noël Jeanneney, ancien président de la Bibliothèque nationale de France, rassemble 23 bibliothèques nationales européennes sur 25 avec l'objectif de numériser entre un et deux millions d'ouvrages sur les six millions représentant la culture européenne. Des industriels soutiennent le projet en France (Thomson, Editis, les éditions Eyrolles).

Au début, il s'agissait de numériser les livres mais aussi de faire pièce au projet de Google de se constituer en moteur de recherche universel. Toutefois, vu l'ampleur des financements nécessaires, la priorité a été mise sur la numérisation, et non pas sur le moteur de recherche, lequel reste du domaine de la recherche et développement avec *Quaero*. En effet, l'expérience de *Gallica,* le site numérisé de la BNF, qui compte aujourd'hui quelque 80 000 documents et plus de 70 000 images, démontre la difficulté d'utiliser les métadonnées de format commun et ouvert qui exigent de remplir les champs correspondants manuellement en attendant une indexation automatique : trente documentalistes sont employés à temps plein à cet effet depuis cinq ans. Au rythme de l'indexation manuelle, il n'y a aucune chance d'arriver à remplir les champs des métadonnées concernant les deux millions d'ouvrages à moins que chaque pays participant au projet de bibliothèque numérique européenne ne fournisse un effort considérable dans les cinq ans à venir.

Malgré l'impulsion politique initiale qui cherchait la meilleure manière de lui résister, Google réapparaît comme partenaire potentiel de la bibliothèque numérique européenne avec ses algorithmes d'indexation automatique. Une politique publique européenne de soutien à l'édition automatique de métadonnées sous forme de développement de routines standardisées paraît plus que jamais indispensable. Dans

ce contexte, la question de l'accès libre aux bases de connaissances européennes n'est pas tranchée. Il serait souhaitable d'obtenir un accord de tous les partenaires concernés dans un partenariat public-privé qui garantirait à la fois l'intérêt des partenaires privés et le bien public informationnel conçu comme la mise à disposition de tout utilisateur des réseaux des savoirs en ligne de son choix avec les outils semi-automatique d'indexation, les moteurs de recherche et les plates-formes de formation les mieux adaptés.

Conclusion

À la demande croissante d'apprentissage tout au long de la vie l'offre publique doit répondre selon des modalités favorisant la richesse des contenus et l'égalité d'accès. Aussi est-il important d'explorer, comme nous avons commencé de le faire ici, les méthodes et moyens à mettre en œuvre pour remplir ces conditions.

Le principe qui ressort des remarques qui viennent d'être faites est le suivant : il convient d'identifier des types de savoirs à préserver de la marchandisation en les déclarant biens publics informationnels, savoirs utiles à la survie, à l'éducation de base, aux compétences indispensables pour s'adapter aux contraintes du développement durable. Ce n'est pas tout, cependant : l'information sur l'information étant un facteur déterminant pour les conditions d'accès au savoir, il s'agit d'inclure aussi dans le bien public informationnel les dispositifs d'indexation et de recherche.

La normalisation des outils et des dispositifs d'accès au savoir en ligne est, en effet, le lieu où se prennent de grandes décisions pour l'avenir. Les pouvoirs publics doivent donc se donner les moyens de réguler la chaîne de production, d'indexation et de mise en ligne des contenus, avec des archives ouvertes, des métadonnées interopérables, des moteurs de recherche utilisant ces métadonnées de manière efficace, des plates-formes de formation ouverte et à distance permettant de réutiliser les contenus, afin de garantir des procédures d'apprentissage optimisées propres à faciliter la construction de connaissances de tout apprenant.

La construction individuelle et collective des connaissances est liée à la maîtrise des outils de communication, nous l'avons dit. Et cette maîtrise dépend elle-même des dispositifs de démocratisation des usages et de la contribution d'enseignants facilitant l'accès au savoir pour le plus grand nombre et, d'une manière générale, oeuvrant à une meilleure répartition au niveau mondial des compétences et des richesses. Dans le cadre de la promotion de l'usage des métadonnées normées, des communautés d'enseignants et de responsables administratifs ainsi que de documentalistes sont à constituer afin de faciliter l'accord international sur les listes d'autorité et les contenus des champs des métadonnées.

Il est intéressant, à cet égard, d'observer l'émergence d'un tiers secteur scientifique, qui participe à la constitution d'un espace public comme lieu de négociation démocratique des innovations et des choix scientifiques. L'ambition est

de bâtir des savoirs alternatifs, produits selon un mode participatif, où le rapport de délégation des profanes aux experts est remplacé par le dialogue et la co-construction des connaissances. Dans la mise en place de cet espace public de la science, il est évident que puissance publique et communautés d'utilisateurs doivent être mises semblablement à contribution.

Bibliographie

Aigrain P., *Cause commune. L'information entre bien commun et propriété*, Fayard, Paris, 2005, 284 p.

Arnaud M., *Liberté, égalité, fraternité dans la société de l'information*, L'Harmattan, Paris, 2007, 178 p.

Bourdieu P., Passeron J.-C ., *La Reproduction : éléments d'une théorie du système d'enseignement*, Editions de Minuit, Paris, 1970, 284 p.

Halttunen K., "Students' conceptions of information retrieval: Implications for the design of learning environments". *Library & Information Science Research*, 25, 2003, pp. 307-322

Lessig L., *L'Avenir des idées. Le sort des biens communs à l'heure des réseaux numériques,* PUL, Lyon, 2005, 414 p.

Meirieu P., *Apprendre...oui, mais comment ? C*ollection Pédagogies, ESF Editeur, Paris, 1999, 192 p.

Perret-Clermont A.-N., *La Construction de l'intelligence dans l'interaction sociale*, Peter Lang, Berne, 1979, 256 p.

Perriault J., *L'accès au savoir en ligne,* Paris, Odile Jacob, 2002, 255 p.

Piaget J., *La Psychologie de l'intelligence*, Armand Colin, Paris, 1967, 234 p.

Rouet J.-F., « La compréhension des documents électroniques » in Gaonac'h, D. & Fayol, M. (Eds.), *Aider les élèves à comprendre*. Paris, Hachette Education, 2003, 256 p., pp. 15-35

Sennett R., *La Culture du nouveau capitalisme*, Albin Michel, Paris, 2006, 160 p.

Stiglitz J., *La grande désillusion*, Poche, Paris, 2003, 407 p.

Tricot A., « L'utilisation d'Internet permet-elle de nouveaux apprentissages documentaires ? », *Colloque Politique documentaire des EPLA et ressources numériques*, CRDP Amiens, novembre 2006