Accéder aux œuvres du passé :

→ ENTRE L'INDUSTRIE DE MASSE ET L'ARTISANAT RELATIONNEL

Hervé Le Crosnier
herve.lecrosnier@unicaen.fr

Après avoir été dix ans conservateur de bibliothèque, **Hervé Le Crosnier** *est actuellement maître de conférence à l'université de Caen Basse-Normandie, où il enseigne les technologies de l'internet et la culture numérique. Il est actuellement en délégation à l'ISCC (Institut des sciences de la communication du CNRS). Créateur de la liste Biblio-fr, il est aussi éditeur multimédia chez C&F éditions (*http://cfeditions.com*). Sa recherche porte sur l'impact de l'internet sur l'organisation sociale et culturelle et l'extension du domaine des biens communs de la connaissance :*
http://herve.perso.info.unicaen.fr

Une approche quantitative de la numérisation ?

Le numérique rend les objets culturels virtuels mobiles et accumulables. Sur un rayonnage numérique, ces objets peuvent être entreposés, et y rester longtemps, même si les lectures (ou achats dans le cas des documents numériques en vente) sont faibles. En première approche, le coût de stockage tend vers zéro, et l'investissement ne se compte qu'une fois, au moment de la numérisation et du dépôt dans la bibliothèque numérique. Certes, des maintenances régulières sont nécessaires pour rafraîchir les données, adapter les fichiers aux nouveaux formats, mais comme elles sont confiées en traitement par lots aux ordinateurs, leur coût apparaît comme magiquement négligeable. De cette situation radicalement nouvelle, qui permet de rendre accessible « tout le savoir du monde au bout du clavier », nous en sommes venus à confondre l'accès et le service de la lecture avec le culte des grands nombres.

Les divers projets de bibliothèques numériques se mesurent trop souvent à la taille de leur catalogue. L'article de Wikipédia (en anglais)[1] consacré au projet Google Books se termine ainsi par un paragraphe « Compétition », qui cite les principales bibliothèques numériques mondiales (Internet Archive, Hathi Trust, Europeana) en les décrivant par le nombre d'ouvrages numérisés qu'elles proposent au public.

Or, dans la tradition des bibliothèques, la question de la « collection » est centrale : une collection est le produit cartésien d'une mission et d'un public. À quoi servent des millions de livres quand on s'adresse à des lycéens ? Pour les chercheurs spécialisés, les documents qu'ils/elles recherchent sont justement en dehors des millions largement accessibles, mais dans les recoins des bibliothèques spécialisées. En réalité, on a confondu le *back office* que constitue le stock de livres numérisés, et la bibliothèque elle-même, qui est avant tout un service offert à un public, chaque fois spécifique.

Cette confusion nous aveugle. Les décideurs politiques et financiers ne jugent plus leur intervention qu'en soutien à une quantité de pages numérisées. On dévie d'un objectif bibliothéconomique vers un objectif industriel. Ce caractère industriel des sites de numérisation et de dépôt est évidemment nécessaire, tout comme des bâtiments solides doivent abriter des bibliothèques physiques. Les nuages de serveurs, qui permettent de stocker des quantités énormes de données, de les répliquer pour assurer leur conservation et de les servir dans les temps les plus courts à des usagers partout dans le monde sont utiles. D'autant plus qu'étant centralisés, ils peuvent intégrer dans leur conception même des économies d'énergie, et l'usage d'énergies renouvelables[2].

1. http://en.wikipedia.org/wiki/Google_Books
2. Hervé Le Crosnier, « De l'immatériel énergivore à l'énergie sociale des réseaux de communication », *Ecorev'*, n° 37, septembre 2011.

Mais cela reste un enjeu industriel, dans lequel des acteurs peu nombreux et fortement concentrés se sont positionnés au point de faire de l'ombre aux décisions des États. Ceux-ci se sentent rassurés par l'appel à ce type de prestataires sans mettre en balance d'autres éléments d'une politique documentaire, notamment la propriété des données (et cela d'autant plus qu'il s'agirait de données publiques), l'organisation des collections en fonction des publics, le rayonnement culturel et l'éducation de masse.

La notion de « domaine public »

Pour entrer dans ce débat d'une autre manière, revenons sur la notion de domaine public de la création. Beaucoup de confusion règne sur ce point. Le domaine public est avant tout un ensemble de contenus dans lequel chacun peut puiser pour inventer des usages nouveaux. Deux distinctions doivent ici être faites.

Les livres

Les livres eux-mêmes restent propriété de leur acheteur ou dépositaire, par exemple les bibliothèques, de même que les tableaux appartiennent aux musées. Ce qui accède au domaine public est le contenu. On doit donc distinguer le fait qu'une œuvre soit dans le domaine public et le fait qu'elle puisse atteindre un lecteur, auditeur ou spectateur. Car un « contenu » ne se transmet qu'une fois reproduit sous une forme spécifique[3]. Longtemps, avant la numérisation, la réédition par impression d'un nouveau volume avec le contenu appartenant au domaine public a été le seul moyen. On doit même se féliciter que des éditeurs aient pu gagner de l'argent avec les grands classiques, sinon nous les aurions certainement perdus. Quand un éditeur décide de rendre un contenu disponible à un nouveau public, il est tributaire de l'existence de ce domaine public (du *back office*) qui propose des œuvres pour lesquelles il n'a d'autorisation à demander à personne. Mais le succès de sa réédition dépend de sa propre volonté de présenter ce contenu différemment, afin de toucher un nouveau public, ou de réhabiliter une œuvre abandonnée, de la traduire[4], ou encore de construire sur le domaine public de nouvelles approches, par exemple sous la forme de préfaces, de postfaces ou de commentaires.

Le domaine public permet de « *se jucher sur les épaules des géants* » suivant l'expression d'Isaac Newton, pour produire des éléments de culture qui soient utilisables ici et maintenant. Car la culture ne se transmet pas « en masse », mais livre par livre, dans un combat permanent pour faire penser, faire vibrer et faire aimer. Les parents le savent bien, qui ont toujours vu leurs enfants rechigner devant les livres qui ont marqué leur jeunesse, et qu'ils essaient de faire apprécier à leur progéniture… Malheureusement, la typographie de l'exemplaire qu'ils possèdent est marquée, le papier, les illustrations, bref tout l'appareillage éditorial est daté… et c'est souvent en achetant la réédition façonnée pour le lectorat actuel que l'on peut faciliter la redécouverte et assurer la transmission. Ce qui se vit régulièrement dans les familles est généralisable à toute la société, et l'industrie éditoriale n'a pas manqué d'en tirer parti, avec les rééditions, le « remastering digital », ou des versions « intégrales » des films du patrimoine.

Considérer le domaine public comme un dépôt ouvert à la réédition, à la mise en valeur, que ce soit par l'appareil éditorial, mais aussi par l'école ou les bibliothèques, nous offre de nouvelles perspectives. Nul n'arrivera à numériser l'ensemble du domaine public, mais, plus encore, nul n'arrivera à l'exploiter au bénéfice de la culture et des savoirs. Le nombre ne fait rien à l'affaire. Que m'importe le nombre de documents accessibles au travers des systèmes informatisés si je n'ai nulle part d'incitation, d'encouragement, si aucun appareil critique ne vient me suggérer de lire une œuvre (ou d'écouter une musique, de voir un film…) ?

Le numérique

Or, avec le numérique, ce qui semblait aller de soi dans l'univers des reproductions onéreuses et nécessitant une chaîne de prestation de service (l'impression, la diffusion et les bibliothèques) aurait disparu car « chacun(e) » pourrait accéder aux trésors précédemment étouffés de poussière sur les rayonnages des bibliothèques.

Je maintiens ici que la facilité de reproduction n'est qu'un des aspects de l'aventure éditoriale qui permet à une œuvre de toucher un public. Des professions intermédiaires sont utiles pour ne présenter au public qu'un extrait restreint de tout ce qui existe… mais avec une approche culturelle qui rend accessibles les documents créés dans des circonstances et avec des lecteurs imaginés très différents de ceux qui peuvent y accéder actuellement.

Pour autant, pour que cette réédition/sélection/promotion soit possible, il faut que ces intermédiaires puissent accéder au grand fleuve du *back office*. Et puissent replacer ce qui y coule dans le flux actuel, que ce soit par réédition imprimée ou numérique, mais aussi dans les blogs, les sites, les sélections destinées aux étudiants par leurs enseignants ou les mises en avant par les bibliothécaires.

Ce phénomène est non seulement nécessaire, mais il est surtout la garantie d'un maintien, autant que faire se peut, de la diversité culturelle. Car si on rapporte les œuvres aux conditions de l'accès, on se trouve aussi désemparé que le néophyte qui ne sait quoi choisir sur un site de streaming musical… et se réfugie sur les listes des

3. Sur le rapport entre le texte et les formes du document, voir : Jean-Michel Salaün, *Vu, lu, su : les architectes de l'information face à l'oligopole du Web*, La Découverte, 2012.
Voir la critique de l'ouvrage par Yves Desrichard dans ce numéro du *BBF*, p. 99-100.

4. Voir à ce sujet la polémique entre Gallimard et François Bon sur une nouvelle traduction du *Vieil homme et la mer* d'Hemingway, par exemple : Hubert Guillaud, « Nous n'échapperons pas à reposer la question du droit », *La feuille*, 17 février 2012 : http://lafeuille.blog.lemonde.fr/2012/02/17/nous-nechapperons-pas-a-reposer-la-question-du-droit

morceaux « les plus écoutés ». La distinction par la recommandation reste un moment essentiel de l'activité de transmission culturelle. Et si le nombre d'écoutes, de liens, de pages citantes ou citées peut effectivement être un critère, efficace pour les revenus financiers des activités culturelles, chacun conçoit aisément que cela ne peut pas être le seul critère, malgré la vogue nouvelle des évaluations chiffrées qui envahit toutes les sphères [5].

Les bibliothèques, acteurs de la numérisation

Limiter la numérisation ?

Les bibliothèques, par le biais de l'indexation, de la classification, mais aussi de l'organisation des espaces en libre accès, sont d'autres acteurs, avec leur spécificité propre, de cette construction d'une « offre », nécessairement orientée vers les publics, c'est-à-dire sélectionnée pour le débutant ou l'élève, plus large pour l'étudiant, et se contentant d'un service d'accompagnement pour la découverte par le spécialiste. Au fond, une bibliothèque se construit principalement autour du service au lecteur, bien plus sûrement qu'en mettant les ouvrages au centre. Et ce qui est une banalité dans le domaine des bibliothèques d'ouvrages physiques reste d'actualité dans les bibliothèques numériques. Comme toute approche qualitative, le service, la relation, la communication ne se mesurent pas aisément à des statistiques, aux lois des grands nombres, et à la quantité de pages, signes, ou mots indexés... qui sont trop souvent le critère numéro un quand on aborde le contenu des serveurs ou des banques de données.

Est-ce à dire qu'il faut limiter la numérisation aux œuvres décidées utiles ? Évidemment non. Rappelons que la numérisation est une chaîne industrielle, qui comme telle a des processus d'optimisation (traiter en bloc une étagère est plus efficace que de décider livre par livre de l'utilité). Mais la numérisation construit un *back office* qui se distingue de l'usage des fichiers numérisés par les bibliothèques, fussent-elles numériques.

Ce qui est important, c'est ce qui est fait ensuite des documents une fois qu'ils ont été remisés sous une nouvelle forme sur des rayonnages électroniques. Deux éléments deviennent déterminants : l'indexation, la classification, la mise au catalogue, les « accroches » comme on dirait dans la presse, qui font qu'un lecteur va pouvoir trouver un document en fonction de ses besoins propres (niveau, langue, date de publication, genre documentaire...) ; et la disponibilité « juridique » du document retrouvé pour les usages ultérieurs.

Une approche calculatoire

Dans les opérations de sélection, constitution de collection, présentation, intermédiation..., le bibliothécaire est à nouveau confronté à un appareil industriel du traitement automatique des documents. Certains pensent que le document une fois numérisé, on pourra reconnaître le texte, le découper en mots et l'indexer dans un moteur de recherche pour qu'il soit enfin « libéré » de sa gangue matérielle et des conditions drastiques de l'accès imposées par des gardiens du temple. Quelques mots dans la fenêtre ouverte d'un moteur de recherche, et les documents sont à nous. C'est oublier en chemin que les listes de réponses sont pré-organisées par ledit moteur de recherche, en fonction d'objectifs qui lui sont propres. Et que seuls les chercheurs obstinés vont au-delà des premières références. Cette « masse » de documents disponibles est finalement réduite à la portion congrue des premiers de liste.

Mais l'opération machinique et calculatoire qui nous délivre cette sélection nous est rendue acceptable par la

5. Alain Abelhauser, Roland Gori et Marie-Jean Sauret, *La folie Évaluation : les nouvelles fabriques de la servitude*, Mille et une nuits, 2011, 208 p.

propagande : ce ne sont pas des humains par volonté de contrôle qui organisent les listes, mais des algorithmes qui découvrent pour vous, en fonction de multiples critères, dont l'applaudimètre (le nombre de lien entrants) et les intérêts publicitaires[6]. Au moins nul n'intervient pour préjuger le contenu lui-même, n'est-ce pas ?

Nous sommes entrés dans l'univers des « big data », pour lesquelles le raisonnement, l'inférence, la démonstration ou même la relation affective doivent s'effacer devant la corrélation. La quantité des données disponibles pour le traitement calculatoire est si importante qu'elle devrait même nous révéler des questions auxquelles nous n'aurions pas pensé, par rapprochements statistiques, analyses de la variance et applications de modèles linguistiques ou sémantiques. Quand les littérateurs ne jurent que par l'originalité, le style, la personne, les ordinateurs se font « *super crunchers*[7] » pour confronter des items, des structures ou des *patterns* en espérant qu'à terme un sens nouveau émergera.

Cette approche calculatoire des « humanités numériques » se distingue de l'approche de construction des « cyberinfrastructures », qui veulent rendre accessibles, annotables et partageables les documents numérisés au service de la recherche. Elle a cependant le vent en poupe, notamment en faisant glisser sur l'expression culturelle et les sciences humaines des travaux déjà épistémologiquement critiquables dans les domaines des sciences du vivant, notamment la génomique. La capacité à « faire » (des calculs, des statistiques, des analyses automatiques…) remplace de plus en plus la réflexion sur l'intérêt culturel et scientifique, au risque de voir émerger des approches déshumanisantes des sciences humaines et sociales[8].

Les contraintes juridiques

L'autre question porte sur les contraintes juridiques qui pèsent sur les documents numérisés en masse et plongés dans ce *back office*. Constituent-ils alors un bien commun dans lequel chacun pourrait puiser, non seulement pour lire, mais aussi pour accélérer le partage, la circulation, éventuellement marchande, ou bien de nouvelles contraintes liées à la numérisation viennent-elle s'interposer ? Si on admet l'argument de cet article selon lequel il ne suffit pas de disposer d'un vaste ensemble de documents numériques pour améliorer l'accès à la culture et la connaissance, mais qu'il s'agit d'offrir le substrat pour les opérations de sélection, d'organisation de collection, de promotion, de remise au goût du jour, etc., il convient de se pencher sur les addenda juridiques que les numériseurs portent sur leurs documents.

Que ce soit Google Books, Europeana ou Gallica, les conditions juridiques limitent l'exploitation, notamment « commerciale ». Est-ce vraiment utile ? Les réflexions ci-dessus sur le domaine public nous incitent à penser qu'une réédition, avec le travail qui doit l'accompagner, mérite que les lecteurs ainsi intéressés par ces actes de remise en lumière puissent accepter de financer le travail complémentaire. Ce qui est déterminant, c'est que des accès libres et gratuits existent, au travers des bibliothèques, ou par le biais d'autres règles de circulation de la culture, notamment le droit à la copie privée. Et cela sans même parler de la concurrence qui permet à tout autre éditeur d'effectuer un travail similaire à partir des mêmes données publiques… ce qui va tendre à favoriser les éditeurs apportant la meilleure plus-value intellectuelle ou de présentation pour le prix le plus faible. Pour le reste, la réhabilitation, la correction des erreurs, la promotion, sont des activités importantes pour la diversité culturelle, qui procèdent d'une autre approche que l'existence d'un *back office* et qui rendent réellement vivant le domaine public. Justement parce qu'un tel travail va jouer sur le « petit nombre », le choix des œuvres, la gestion d'un public…

La culture et la connaissance sont enjeux de transmission. Les relations, la communication, l'encouragement à la lecture, la promotion et la réhabilitation des œuvres portent dans ce cadre beaucoup plus loin que la simple mise à disposition de grandes masses de documents. Nous nous trouvons donc à la croisée de nouveaux chemins.

D'une part, il convient de construire les bases arrière de la circulation accélérée des savoirs sur les réseaux numériques par la numérisation massive, industrielle, des documents. Mais d'autre part, il est indispensable de favoriser l'initiative des intermédiaires pour agir comme passeurs d'idées sur les nouvelles frontières du savoir et de la circulation des émotions et des idées. On ne peut valoriser que ce que l'on aime, en général quelques œuvres, la construction d'un choix défini par des critères particuliers, par exemple les critères d'un enseignant pour son cours, ceux d'un critique pour son journal ou ceux des bibliothèques pour l'organisation des collections.

C'est donc en acceptant l'initiative privée (marchande ou, bien plus encore, non marchande, dans les réseaux interpersonnels de transmission) que nous valoriserons réellement le domaine public et rendrons aux bibliothèques numériques un véritable rôle dans la découverte, la préservation et l'organisation des documents issus du passé. Et que livre par livre nous garderons vivante la longue histoire accumulée dans les bibliothèques. ●

Avril 2012

6. Brigitte Simonnot et Gabriel Gallezot (coord.), *L'entonnoir : Google sous la loupe des sciences de l'information et de la communication*, C&F éditions, 2009, 248 p.
7. Ian Ayres, *Super Crunchers : Why Thinking-by-Numbers Is the New Way to Be Smart*, Bantam, août 2007.
8. dana boyd et Kate Crawford, « Six provocations for big data » : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431