

Bibliothèques numériques : les aiguilleurs du Web

par Yannick Maignien¹

Les bibliothèques sont à la croisée des chemins numériques², mais plutôt que de frémir aux risques des

mutations, il apparaît plus conséquent de voir combien important leur rôle nouveau d'aiguilleur du Web pourrait apparaître au fur et à mesure du développement d'Internet³.

et de coopération, les bibliothèques peuvent à l'avenir mieux définir leur rôle sur Internet.

1. L'auteur est responsable de la mission scientifique de numérisation de la Bibliothèque nationale de France, mais, selon la formule consacrée, si les propos tenus ici doivent beaucoup à l'expérience accumulée ces dernières années dans ces fonctions, ils ne sauraient engager l'institution de la BNF, et ne relèvent bien souvent que de considérations et réflexions personnelles de l'auteur.

2. On fera grâce ici au lecteur de la pléthore de références à ces craintes, à la mort du Livre, au « Ceci tuera cela » transposé de V. Hugo, à la disparition des bibliothécaires, etc., thème récurrent de toute nouvelle avancée technologique.

Sur les trois dimensions suivantes : politique documentaire, d'accès et de réseau

3. De nombreux rapports sur les autoroutes de l'information ont alerté sur le rôle que peuvent jouer les bibliothèques au sein de cette mutation, mais souvent de façon allusive : voir Alain Minc, Rapport Therry, le Programme « Livre électronique », GIS, Le Club de l'Arche, etc.

Ces trois dimensions sont loin de dessiner d'emblée le triangle vertueux d'une politique de numérisation. Nombre d'écueils en feraient un passage aussi dangereux ou indéterminé que celui des Bermudes, et rien dans les technologies existantes ne prédit un avenir certain ou radieux. C'est plutôt du côté du réexamen, de la réévaluation des missions pérennes des bibliothèques qu'il importe de s'interroger.

Politique documentaire

Elle est souvent passée sous silence ; le choix et l'attention accordés aux contenus, l'élection des documents pour la numérisation, déterminent l'apport des bibliothèques sur un Internet encore relativement pauvre en information primaire, au profit de beaucoup de bruit secondaire. Chacune d'elles, sur un secteur spécialisé, régional, ancien, ou simplement bien signalé ou décrit, a matière à diffuser largement, peut-être au niveau planétaire, telle ou telle collection, tel ou tel fonds, pour tel ou tel segment de public. C'est d'abord en tant que producteur, émetteur, serveur, que la bibliothèque s'engage dans la numérisation. La reproduction est ici nouvelle production, car nouvelle diffusion et nouvelle lecture, et le Web favorise ici la fourniture du différent et du spécifique.

Nous ne rentrerons pas ici dans le détail de ce que doit ou peut être une telle politique, encyclopédique ou spécialisée, patrimoniale ou documentaire de façon active, tant elle dépend des ressources ou des missions propres. Rappelons simplement qu'elle doit conjuguer les facteurs suivants :

- critères de sélection liés à l'importance intellectuelle, mais aussi à la rareté ou à la difficulté d'accès (liée elle-même autant à des critères physiques, éditoriaux que disciplinaires) ;
- évaluation de l'ampleur de la diffusion et donc participation d'une politique de sauvegarde ;
- récurrence des accès et déclinaisons de nouveaux services (tirage à la demande, aide aux accès pour handicapés, réassort, etc.) ;
- multiplicité des supports et cohérence multimédia (mais cohérence aussi des signalements) ;
- domaine public et documents sous droits, pour autant que les réseaux de diffusion ou les services associés le permettent, etc.

Il s'agit toujours, ne serait-ce que pour gérer des ressources rares, d'optimiser un certain rapport offre/demande.

La numérisation reste liée à l'existence du document physique, papier ou microforme : la scannérisation des documents reste dans son ensemble fonction d'une disponibilité des documents ; trop souvent la reprographie est vécue comme une atteinte à l'intégrité du document, sinon du patrimoine, et l'on aboutirait rapidement au

paradoxe que les documents les plus rares seraient les moins reproductibles. La manipulation unique, soigneuse et scientifiquement adaptée de reproduction est l'atout maître d'une politique de sauvegarde. La difficulté d'accès aux documents (qui est bien une de leurs définitions intrinsèques) continue de conditionner, c'est-à-dire de restreindre leur numérisation⁴.

Ajoutons, et c'est le principal : la politique documentaire de chaque établissement peut être ouverte, évolutive, dynamique. En effet, tout plaide pour l'association de comités scientifiques qui peuvent accompagner telle évolution de collection, justifier de façon interactive telle mise en relief, tel accroissement, tel désherbage virtuel et temporaire. On le soulignera ensuite, la limite entre politique documentaire et politique éditoriale est mince.

Enfin, la numérisation et le Web ouvrent la politique documentaire sur une question essentielle : l'organisation des connaissances. On pourrait citer ici M. Foucault, pour qui le document valait plus pour son statut « monumental » que pour son contenu subjectif, plus comme « masse d'éléments qu'il s'agit d'isoler, de grouper, de rendre pertinents, de mettre en relations, de constituer en ensembles » que comme « l'heureux instrument d'une histoire qui serait en elle-même et de plein droit *mémoire*. » Il est certain que de vastes collections numériques réalisent ou du moins permettent « la mise en question du document ». C'est en effet un certain rapport de l'Histoire au document, déjà en œuvre culturellement, que la numérisation vient appliquer et développer : « L'histoire, c'est une certaine manière pour une société de donner statut et élaboration à une masse documentaire dont elle ne se sépare pas⁵. »

4. Nous avons discuté ailleurs cette idée dans « L'œuvre d'art à l'époque de sa reproduction numérisée », *Bulletin des Bibliothèques de France*, n° 46, 1996.

5. M. Foucault, *L'Archéologie du savoir*, Paris, Gallimard, 1969, p. 14, et de nombreux autres passages qui insistent sur l'histoire comme rapport au « tissu documentaire », qui « l'organise, le découpe, le distribue, l'ordonne, le répartit en niveaux, établit des séries, distingue ce qui est pertinent de ce qui ne l'est pas, repère des éléments, définit des unités, décrit des relations. » Nous avons tenté de développer ce rapport de la structure documentaire à l'hypertexte avec J. Virbel dans la catalogue de l'exposition de la BNF : « Tous les savoirs du monde ». Voir « Encyclopédisme et hypermédia : de la difficulté d'être à la complexité du dire ».

Politique d'accès

Par la scannérisation, la numérisation est une sorte de mise à plat, de révélation du dénominateur commun de la « matérialité documentaire (livres, textes, récits, registres, actes, édifices, institutions, règlements, techniques, objets, coutumes, etc.) » pour continuer à citer ou détourner M. Foucault. Cette ère de la reproduction a pour contrepartie une assistance technique, informatique, qui, si elle a pris pied depuis longtemps dans la gestion et le signalement bibliographiques, atteint maintenant le contenu de ce tissu, de cette matérialité documentaire, et culmine dans la lecture assistée par ordinateur.

L'architecture logique de la bibliothèque numérique vers laquelle tendent de nombreuses recherches dans le monde peut se dessiner ainsi : le modèle définit une matrice qui croise des entrées fonctionnelles d'usage et des entrées techniques d'infrastructures informatiques.

Les entrées d'usages peuvent être ramenées à trois items principaux : le catalogue, la collection personnelle, le bureau.

Les entrées techniques peuvent être aussi résumées autour de trois axes : architecture serveur, protocole réseau, fonctionnalités du terminal.

Nous n'entrerons pas ici dans les développements complexes de chacun des croisements de cette matrice, mais pour l'essentiel, c'est une réflexion sur celle-ci qui commande la place que le génie logiciel doit prendre sur chacune de ces instances techniques ou de ces fonctionnalités d'usage⁶.

Limitons-nous à préciser la logique utilisateur requise. Le cœur de la question de la bibliothèque numérique est repérée par la littérature internationale par les fonctionnalités que le lecteur final peut réellement utiliser sur son terminal « The library is the workstation⁷ » écrit un collègue suédois, au sens où l'ordinateur est à la fois :

– A : lieu de l'identification et interrogation sur le Web des serveurs et catalogues de

6. Un programme européen, Memoria, avait permis d'approfondir les réflexions présentées rapidement ici.

7. Ceci n'enlève rien au retard actuel de la France : 7 à 10 % des foyers seulement sont équipés d'ordinateurs (et 1 % seulement de modem), contre 37 % aux États-Unis, ni aux tarifications dissuasives en matière de télécommunication.

références bibliographiques. L'icône ici peut être la « bibliothèque » ou le catalogue ;
 – B : outils de présélection pour constituer une collection personnelle (consultation des tables des matières, possibilités de moteurs de recherche sur ces tables des matières ou tout autre extrait significatif et interrogations d'autres aides bibliographiques). L'icône peut être ici l'« étagère » ;
 – C : travail sur les documents ouverts sur le bureau et ensemble des fonctionnalités de lecture, tel que les logiciels du terminal peuvent l'assister : identifier l'information, l'annoter, la structurer, la baliser par des chemins de lecture, permettre l'OCR et mettre en œuvre des outils linguistiques ou statistiques, etc. L'icône est ici le « bureau ».

Ces trois instances « catalogue », « étagère », « bureau » sont emboîtées, sinon hiérarchisées par la plus ou moins grande appropriation du document d'un côté, par la plus ou moins précise description bibliographique de l'autre, de la part des bibliothèques ou centres responsables de fonds.

Le parcours du lecteur est souvent complexe : des catalogues (A) à la lecture directe (C) pour retenir ou non en (B) tel document reconnu par indexation ou classement ; mais aussi référence trouvée pendant la lecture (C), motivant l'interrogation de (A), et nécessitant de rassembler des extraits plein texte en (B) pour sélectionner et lire en (C), etc.

Mais en même temps, cette complexité est proche des règles de « bon voisinage » qui optimisent le regroupement documentaire.

Aussi, la gestion de cette « profondeur » de lecture et d'exploration doit-elle se rattacher de façon très intuitive à ces trois niveaux ou icônes.

Architecture logique et structuration

Si la logique de l'utilisateur final est bien ce qui commande ce modèle, il n'en reste pas moins pour les bibliothèques de savoir jusqu'où doit aller l'effort de structuration des collections en amont. On sait que la question de l'adoption de SGML, pour compléter, sinon remplacer les formats traditionnels (MARC) est au centre des discussions.

Aussi, centrale est la question des protocoles d'interrogation client-serveur (Z39.50) qui peuvent être implémentés

sur le réseau indépendamment des formats d'origine.

Plus globalement, la question de la méta-information, et la finesse de sa granularité descriptive, est un enjeu majeur, par exemple avec la mise en œuvre de l'encodage de la Text Encoding Initiative. La productivité attendue est au croisement de la méta-information automatisée et de l'information sémantique fine traitée en aval par le lecteur.

Enfin, « la collection personnelle » est la question sensible, en tant qu'elle marque l'appropriation par le lecteur de *certain*s documents, mais aussi en ce qu'elle mélange l'information lue, image du texte, texte alphanumérique où des moteurs de recherche s'y appliquent, et la méta-information qui décrit le document.

Le schéma A.B.C. présenté plus haut n'a que les apparences de la simplicité. Pour être pleinement productif, il faut considérer qu'il met en jeu une structure hypertexte complexe : liens des rubriques catalogues aux différentes structures bibliographiques (d'autorité, d'indexation, de classification) ; liens de celles-ci aux différents extraits para-textuels utilisés dans les extraits enrichis (tables des matières, mais – pourquoi pas ? – bibliographies, index, fréquence des citations, etc.) ; liens enfin à l'information lue et repérée de façon linguistique, sémantique ou morpho-syntaxique. HTML est ici autant indispensable que vraisemblablement réducteur, par l'impossibilité de typer ces familles de liens.

À cet égard, les métadonnées qui permettent de lier des ressources distantes sont aussi les adresses et identificateurs PURL (Persistent Uniform Resources Locators), URC (Uniform Resources Characteristics) et URN (Uniform Resources Names).

Enfin, une des questions ouvertes riche de promesses est celle de la « charge » d'interactivité possible attachée aux documents, aux différents niveaux logiques A.B.C., telle que les définissent les langages « objets » ou Java.

Être lecteur de telle bibliothèque, ce ne sera plus seulement être assidu à telle salle de lecture, mais être assidu (depuis chez soi, son lycée ou son université) à tel site Web de telle bibliothèque, valant par la qualité de ses ressources propres, bibliographiques et documentaires en ligne, valant par la qualité des liens avec

d'autres fonds répartis, mais aussi par la fiabilité et les performances des outils de structuration et de travail utilisables directement sur ce terminal distant.

Politique de réseau et de coopération

Elle est essentielle dans la définition même de la notion de bibliothèque numérique : se développer d'emblée dans une stratégie de réseau coopératif entre établissements et avec une répartition de compétences. C'est le cas, on le sait du National Digital Library Project américain.

Là encore nous retrouvons la question de la continuité des missions des bibliothèques : peut-on avec le numérique se conformer à un modèle connu de rassemblement, de conservation et d'accessibilité des connaissances ? ou faut-il faire place à un réseau multipolaire, de compétences partagées ? S'il apparaît que les bibliothèques doivent avoir un rôle renforcé, du moins culturellement, d'orientation, d'authentification, d'aiguilleurs du Web, et pas seulement de producteurs ou diffuseurs d'informations parmi d'autres, un centre unique de réseau en étoile ne peut jouer ce rôle.

De nouvelles pratiques éditoriales, de nouveaux objets d'écritures, numériques avant même d'être imprimés⁸, ou même purement numériques, supposent de nouvelles digithèques. Et que l'écriture du Web pour l'essentiel échappe au dépôt légal souligne combien n'est pas réglée la conservation nouvelle, la pérennité d'une telle dynamique d'écriture.

Acteurs sur le réseau, les bibliothèques doivent-elles avoir pour autant une politique éditoriale ?

Droits d'auteurs et missions publiques de diffusion de la connaissance

La plupart des forums ouverts sur les Digital Libraries⁹ mettent en exergue

8. Voir l'excellent article de Vivette Pouzet « La revue des Revues » s'interrogeant sur l'évolution de format des revues sur Internet.

9. Voir le serveur et groupe de discussion Diglib de l'IFLA modéré par T. Kunny.

l'acuité de la contradiction entre politique de numérisation et obstacles à la diffusion des documents numériques couverts par des droits.

Certes, chacun convient que le problème est complexe, que les intérêts en jeu sont importants, au moins autant que le flou

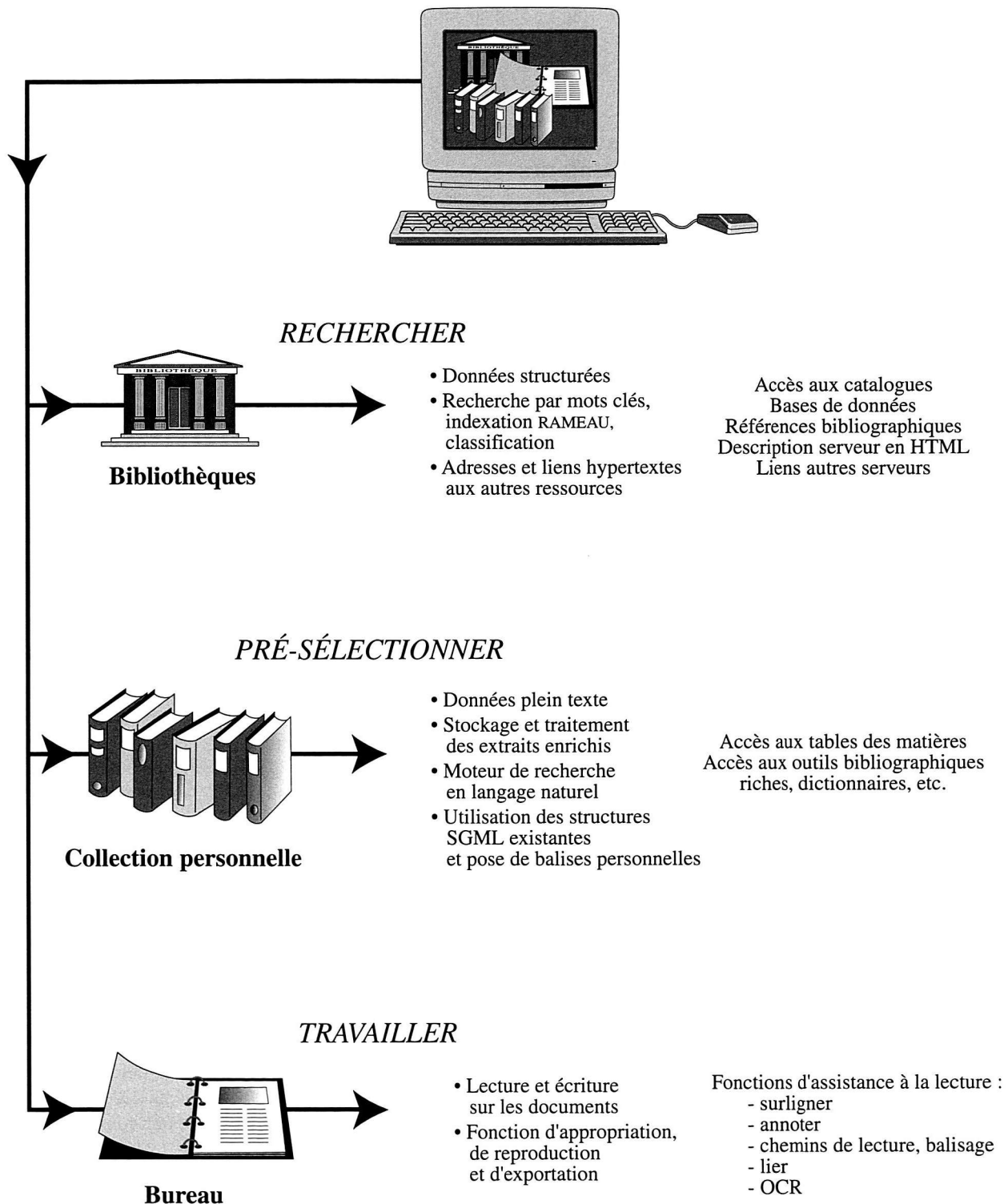
prospectif pour l'édition concernant l'avenir des gains ou les risques de pertes !

Faut-il pour autant condamner les bibliothèques à ne mettre en réseau que le domaine public ? Quand bien même ? Comment ferait-elle à l'inverse reconnaître, sinon protéger l'énorme investissement

intellectuel mis à sélectionner et signaler des collections pertinentes. Il apparaît donc urgent que les pouvoirs publics ou les législateurs aient des propositions.

Se placer sur un terrain strictement juridique semble stérile, si celui-ci circonscrit une seule pratique, celle de la diffusion,

Logique d'accès aux fonds numérisés



de la reproduction d'un document essentiellement matériel, physique, quantitativement évaluable : l'imprimé. Rien d'étonnant que le problème soit bloqué par des intérêts existants divergents, d'ailleurs assez spécifiques aux pratiques éditoriales et bibliothécaires de différents pays ou cultures (les missions de service public des bibliothèques ne sont pas définies de la même façon en France, dans l'Europe du Nord ou aux États-Unis, de même pour le droit d'auteur ou le copyright).

Il nous semble qu'il y aurait intérêt pour tous, éditeurs et bibliothèques, à se situer résolument sur un terrain nouveau, à le traiter comme tel, de façon expérimentale en faisant table rase de références à l'imprimé, et à observer et évaluer précisément ce qui se passe sur des domaines bien cernés. En effet, rien n'infère qu'il faille penser en volumes constants, où les effets de substitution par rapport à l'édition papier seraient à somme nulle. Autrement dit, quels secteurs éditoriaux électroniques sont à fort coefficient d'évolution ? quels autres ne sont pas notablement changés (mais au contraire valorisés ou promus) par une large diffusion électronique de service public ?

Autres perspectives concrètes : quelle complémentarité entre des extraits significatifs largement mis en accès sur fonds publics et œuvres complètes diffusées seulement de façon commerciale ? Quels outils statistiques et de suivi de la demande peuvent être fournis aux éditeurs grâce à la mise en réseau ? Quelles coéditions ou coproductions entre établissements publics et éditeurs privés sont-elles possibles ? Quelle baisse du coût unitaire d'accès à l'information est à la fois rentable pour l'éditeur, souhaitable pour le lecteur et supportable pour les missions et services publics des bibliothèques ? Que signifie gratuité d'accès à l'information si ne sont pas comptabilisés les coûts consolidés d'investissement dans la numérisation ? Y a-t-il possibilité de développer une offre numérique en proposant aux éditeurs de numériser sous licence des collections patrimoniales des

bibliothèques ? Qu'en est-il de l'évolution de l'édition de service public¹⁰ ?

Autant de questions (posées dans différents pays) qui demandent un regard neuf. La réorganisation de tels réseaux de diffusion, les nouvelles formes de partage du savoir, les nouvelles régulations économiques puis juridiques ne peuvent bien sûr pas être devinées et décrétées *a priori*. Elle doivent donc faire l'objet d'« ateliers » de réflexion, associant des experts de bibliothèques et de l'édition, afin de ne pas ajouter aux difficultés techniques des lignes Maginot juridiques dont on connaît la solidité face à l'extérieur.

Réseaux prioritaires d'utilisateurs

Une politique de numérisation, étant donné les investissements conséquents qu'il faut réunir, doit enfin en priorité avoir une demande cible.

À l'évidence, les besoins les plus cruciaux, tant nationaux qu'internationaux, semblent être du côté de l'éducation secondaire et universitaire. La croissance nouvelle, quantitative et qualitative des besoins et des comportements en ce domaine (autoformation et personnalisation des contenus, accès de plus en plus diversifiés, développement de la documentation scientifique et technique, etc.) plaide pour une flexibilité et une forte réactivité des accès que ne permet plus toujours l'édition papier.

La reprographie papier (photocopie) vient d'avoir sa législation et sa régulation économique. Mais comme elle procède maintenant de photocopieurs numériques, la souplesse et l'ubiquité des besoins pousseront l'extension du champ réglementaire au numérique. D'ores et déjà, les grands éditeurs scientifiques anglo-hollandais développent des accès sous licence sur les réseaux locaux des campus. Une cartographie documentaire entre différentes bibliothèques universitaires spécialisées devraient nourrir les réseaux de recherche et de formation¹¹, par exemple en France sur Renater I et II.

Conclusion

Importance intellectuelle de la politique documentaire, efficacité fonctionnelle et informatique de l'architecture d'accès, capacité à intervenir dans les réseaux nationaux et internationaux de coopérations ou de coédition, la numérisation redéfinit assez profondément les missions et fonctions bibliothéconomiques classiques.

Bien sûr, chaque bibliothèque n'a pas vocation à amplifier un nombre toujours croissant et diversifié de sites Web. Le réseau ne cessera de se réorganiser, de se hiérarchiser.

Dès lors que la bibliothèque oriente sur ses ressources propres, mais aussi sur des ressources externes ou associées, le site serveur d'une bibliothèque est bien un poste d'aiguillage pour trier, hiérarchiser des flux complexes. De ce point de vue, la maintenance du site Web va mobiliser de plus en plus de compétences et de ressources humaines au sein des bibliothèques. On a dit plus haut que l'informatisation des documents mettait le lecteur en position centrale dans le dispositif, mais ce serait une erreur de penser qu'il court-circuite ainsi le rôle du bibliothécaire. Le dispositif A.B.C. signalé plus haut doit être le site d'accueil, d'observation, de consultation à partir duquel se situe et travaille le lecteur. L'assistance au lecteur passe largement par les performances du site Web, par l'interopérabilité dont est capable ce site avec les autres, par le repérage plus ou moins riche que ce site a fait *a priori* des autres ressources. Les aiguilleurs du Web doivent être des « fureteurs », à la fois espions et hommes de partage, prédateurs et prodiges à la fois, dans la mesure où, sur le réseau, on ne trouve et prend que par ce qu'on donne et échange...

Les bibliothèques peuvent craindre de se dissoudre dans un réseau planétaire de communication, de voir disparaître ce qui fait leur spécificité patrimoniale. Mais si cette crainte conduit à désertier les questions documentaires, d'accès et de coopération soulevées ici, il y a fort à parier que la bataille est déjà perdue. En revanche, si les bibliothèques s'engagent résolument dans cette transformation, la société de l'information à venir saura reconnaître le rôle indispensable et renouvelé qui est le leur.

10. Les éditions du CNRS, l'INIST, la Documentation française, le Comité national d'éthique, les centres de documentation d'organismes comme l'INSERM, l'ORSTOM, la Réunion des Musées nationaux, les presses universitaires, etc.

11. Voir Les actes du colloque de Jussieu « L'Internet, l'université et l'édition ». Université Paris VII, Denis Diderot. Collection « Vie universitaire », octobre 1995.