

Les plates-formes de gestion des connaissances de l'Inserm

Une diffusion collective au service de la recherche

Les nouvelles technologies de communication, de plus en plus intégrées à notre quotidien, ont profondément modifié le monde de l'information scientifique et technique.

Nicole Pinhas

Inserm-Département
de l'information scientifique
et de la communication
nicole.pinhas@dicdoc.inserm.fr

On a assisté en quelques années, dans le domaine médical :

- à la diffusion gratuite *via* Internet des bases de données de la National Library of Medicine (Medline, Cancerline...), avec une interface assez performante pour masquer à l'utilisateur la complexité intrinsèque de l'indexation et permettre au chercheur de réaliser sa bibliographie, directement, sans intermédiaires et depuis son poste de travail ;
- à la possibilité de naviguer facilement de la référence bibliographique au texte intégral ;
- à l'apparition de bases de *pre-prints*, nouveau système de publication des résultats de la recherche qui, tout en conservant le processus de soumission et de validation par les *peer reviews*, risquent de bouleverser le processus éditorial qui prédomine jusqu'à présent ;
- à la possibilité pour toute personne de diffuser aisément de l'information avec une audience élargie mais non maîtrisée *via* le web.

Ainsi la chaîne traditionnelle de production auteur, éditeur, bibliothécaire se trouve en pleine mutation et on assiste, comme le souligne Jean-Manuel Bourgois, à « *une désintermé-*

diation », « *c'est-à-dire à la disparition des intermédiaires dans une relation directe entre le producteur ou le vendeur et le consommateur.** »

Le contexte de l'Inserm

L'Inserm (Institut national de la santé et de la recherche médicale) est un organisme qui comporte près de 400 structures de recherche réparties dans toute la France, principalement au sein des hôpitaux.

Ces dernières années, se sont mis en place des regroupements de laboratoires d'origine institutionnelle variée autour d'une *thématique*, avec mise en commun de moyens humains, administratifs et techniques : les instituts fédératifs de recherche (IFR). On en compte actuellement près de 80, et certains peuvent être géographiquement éclatés. C'est dans ce contexte de mise en place de réseaux scientifiques thématiques que, depuis 1996, le Département de l'information scientifique et de la

* Jean-Manuel Bourgois, « L'avenir des bibliothèques et de l'édition dans le contexte de l'édition électronique », *BBF*, 2000, t. 45, n° 6, p. 63.

Nicole Pinhas, docteur en pharmacie, est responsable du pôle gestion des connaissances au sein du Département de l'information scientifique et de la communication de l'Inserm. Elle a publié plusieurs articles ayant trait à la documentation dans le domaine médical.

communication de l'Inserm (DISC) s'est orienté vers la mise en place de plates-formes de gestion des connaissances :

- pour faciliter l'accès à toute l'information nécessaire, non seulement bibliographique, mais aussi textuelle, factuelle, à partir d'un point d'entrée unique ;
- pour répondre aux besoins et aider la communauté scientifique dans le fonctionnement des « collaboratoires », véritables laboratoires sans murs qui commençaient à se mettre en place avec l'apparition des nouvelles technologies de communication (Internet, *groupware*...). En effet, la recherche exige des collaborations de plus en plus nombreuses qu'expliquent une spécialisation de plus en plus poussée, des coûts de plus en plus importants, et une nécessaire mise en commun de matériels techniques et biologiques. Au niveau des projets européens, cela représente même une obligation affichée, nécessaire pour obtenir un financement.

Les plates-formes de gestion des connaissances

On peut les définir comme un ensemble intégré d'outils permettant de gérer et de diffuser l'information d'origine diverse, dans un contexte de travail collaboratif et au moyen d'Internet.

L'objectif affiché est de répertorier, gérer et diffuser les connaissances sur une thématique, en créant une organisation « virtuelle » basée non sur l'institutionnel et l'administratif, mais sur les besoins et les habitudes de travail et d'échanges propres à chaque communauté scientifique concernée. Comme l'indique Jean-Pierre Courtial, « la production

des connaissances scientifiques est liée à un mode de fonctionnement social... Tout travail de recherche se situe à l'intérieur d'une communauté, liant social et cogniti » (cf. bibliographie).

Le DISC, par l'intermédiaire de son réseau de spécialistes, le réseau DISC-DOC, apporte aux chercheurs son savoir-faire technique et méthodologique permettant d'implanter rapidement une infrastructure documentaire de travail et d'échange, ciblée et adaptée à leurs besoins, pour en préserver l'expertise et faciliter le partage des données de toutes sortes.

Ces applications ont été développées à partir du logiciel de *group-*

ware Lotus Notes qui possède un traducteur HTML (Hypertext Markup Language) intégré, Domino, avec une infrastructure informatique comportant un serveur central situé, pour des raisons historiques et techniques, à Strasbourg, un site miroir au Kremlin-Bicêtre, et des logiciels clients implantés sur les ordinateurs des membres « actifs » des réseaux : Lotus repose sur une architecture client-serveur.

Actuellement, une dizaine de plates-formes sont opérationnelles et disponibles sur le web, quelques-unes en accès réservé. Elles couvrent des thématiques variées comme l'éthique (Réseau Rodin), le vieillissement, la recherche clinique ou l'économie de la santé (voir tableau).

Les plates-formes de gestion des connaissances Inserm	
<p>Les encéphalopathies spongiformes subaiguës transmissibles (ESST) et les prions En partenariat avec le Groupement d'intérêt scientifique « infections à prions »</p>	http://www.inserm.fr/prions
<p>Éthique médicale et de santé publique En partenariat avec le laboratoire d'Éthique médicale et de santé publique de la Faculté de médecine de Necker</p>	http://www.inserm.fr/ethique
<p>Conditions de vie des aînés En partenariat avec le Réseau Fédératif de Recherche n° 12 « Santé, Vieillesse, Société »</p>	http://www.inserm.fr/age
<p>Économie de la Santé Analyse détaillée des études publiées dans le champ de l'évaluation économique des actions de santé portant sur la France En partenariat avec le Collège des Économistes de la Santé</p>	http://www.inserm.fr/codecs
<p>Centre de recherche sur les enjeux contemporains en santé publique En partenariat avec l'université Paris 13 Nord</p>	http://www.inserm.fr/cresp
<p>Inégalités sociales de santé Réalisée dans le cadre de l'Intercommission 6 Inserm « Analyse et évaluation des systèmes de soins et de prise en charge de prévention et protection sociale » et autour de l'ouvrage édité par La Découverte</p>	http://www.inserm.fr/serveur/Dispsoc.nsf
<p>Recherche clinique • Inserm : Promoteur de projet de recherche biomédicale • Centres d'investigation clinique</p>	http://www.inserm.fr/rbm http://www.inserm.fr/cic
<p>Éthique et Qualité des recherches à l'Inserm • Qualité • Recherche clinique et non clinique • Étude sur l'animal</p>	http://www.inserm.fr/eqri

Un partenariat indispensable

Il faut, pour lancer un projet de plate-forme de gestion des connaissances, que la communauté scientifique soit prête à diffuser et échanger des informations, qu'elle manifeste une volonté de partage, qu'elle soit « demandeuse ». Souvent, au début, elle aura du mal à exprimer de manière claire et précise ses besoins. Notre rôle est alors de la sensibiliser à toutes les possibilités offertes par les nouvelles technologies de communication en s'appuyant sur nos expériences précédentes.

Un réel partenariat doit s'instaurer entre les trois maillons essentiels du dispositif :

- les scientifiques, à la fois producteurs et utilisateurs des informations ;
- les documentalistes, gestionnaires des systèmes d'informations ;
- les informaticiens qui vont adapter ou modifier les systèmes existants en fonction de la problématique propre à la communauté scientifique concernée, selon les besoins exprimés ou ceux qui vont apparaître à l'usage.

Les chercheurs sont eux-mêmes organisés en réseaux formels ou informels qui vont venir se greffer au noyau central précédemment décrit.

Pour rendre opérationnel le dispositif, les responsabilités et les tâches de chacun des acteurs doivent être clairement définies. Un comité technique regroupera le ou les scientifiques les plus motivés, l'informaticien et le spécialiste en information. Pour chaque rubrique, un responsable sera nommé en charge de l'animation, du suivi et de la qualité du contenu informationnel.

À côté de ce comité technique, véritable « cheville ouvrière », seront mis en place un comité scientifique garant de la qualité scientifique du projet, et un comité éditorial qui veillera au suivi et à l'évolution de ce projet. L'expérience montre que les chercheurs, en dehors de leur équipe au sens le plus restreint, n'ont pas comme habitude de diffuser l'infor-

mation reçue, et il faudra absolument obtenir leur adhésion et surtout la volonté affirmée du directeur de laboratoire dont l'engagement dans le projet est stratégiquement et politiquement indispensable. Il est évident que le « relationnel » représente un élément important dans la réussite du projet, et que le facteur humain, souvent sous-estimé et masqué par rapport aux facteurs technologiques, ne doit pas être négligé.

De plus, passer du papier au numérique, en évitant au maximum la numérisation des documents, demande la mise en place d'un circuit de collecte parfois complexe qui doit être formalisé et connu de tous les membres, permanents ou de passage, chercheurs, secrétaires, étudiants ou stagiaires.

Des réunions régulières, très fréquentes au départ, permettront de dynamiser le projet et de créer le ciment humain indispensable : on échange plus facilement quand on se connaît bien.

Les comptes rendus des réunions, l'état d'avancement et le suivi du projet sont gérés par le même système et accessibles à tous en accès réservé depuis Internet.

Au sein de ce partenariat, le spécialiste en information doit jouer un rôle de médiateur et souvent de chef de projet. À l'écoute de ses partenaires, il doit comprendre et évaluer leurs attentes et leurs besoins qui peuvent évoluer en fonction des usages. Il doit être capable, en partenariat avec l'informaticien, de concevoir et proposer des solutions qui donneront lieu à de nouveaux développements informatiques. Il aura aussi pour mission, tout au long du projet, de remédier aux écueils du travail partagé en améliorant au maximum la collecte des données, en harmonisant les formats pour les rendre compatibles avec les normes en vigueur, et en proposant des outils documentaires adaptés. Plus la chaîne de production sera automatisée, plus les règles seront bien définies et normalisées à la saisie, plus le travail de mise en ligne en sera facilité.

Nous devons assurer la formation des participants aux outils et aux contraintes du projet, même si les outils doivent rester simples, facilement utilisables à l'aide des fonctions habituelles de traitement de texte (coller/copier).

Une architecture à la carte

La mise en place du réseau humain va retentir sur la conception de l'architecture informatique. Les membres du noyau central, le plus souvent membres du comité technique, auront le droit d'ajouter, de modifier ou de supprimer. Ils peuvent intervenir directement par l'intermédiaire du logiciel client présent sur leurs ordinateurs : leurs droits sont déterminés par leur identifiant (ID), et gérés par l'administrateur du serveur. Les autres participants ne pourront intervenir que de manière indirecte, par messagerie, pour transmettre informations et commentaires. On peut aussi décider que toute information avant sa diffusion sur Internet nécessite une validation par le responsable scientifique. Mais la mise en place d'un tel circuit a été jugée inutile et trop lourde dans les applications développées jusqu'à présent.

L'infrastructure interne des plates-formes s'appuie sur l'intégration d'un ensemble de bases de données permettant de gérer l'information structurée (références bibliographiques, réglementations, mémoires et synthèses de maîtrises, DEA (diplôme d'études approfondies), thèses, cours/enseignements, congrès, ouvrages...). À chaque type de document correspondent un format propre et une grille de saisie bien définie. Il s'agit d'un ensemble de briques qui sont assemblées à la demande, interrogeables séparément ou globalement. Pour chaque base, il faudra définir les « vues », c'est-à-dire indiquer les entrées possibles pour un même document (par exemple vue par auteur, par date, par théma-

tique...), pour rendre plus aisé et convivial, l'affichage de l'information (fig. 1).

À cet ensemble de bases de données sont intégrés des services annexes : annuaires, forum de discussion, planning des réunions, agenda, gestion de projets...

Ainsi le réseau Rodin, dont l'objectif est de promouvoir le débat social à partir d'un contenu en éthique médicale, riche et validé, est constitué de 8 bases de données, d'un annuaire des membres du réseau et de plusieurs forums de discussion. Il comporte plus de 20 000 documents enregistrés et indexés (fig. 2).

À partir de cette organisation interne basée sur la typologie des documents, on peut envisager une organisation de diffusion qui ne s'appuiera plus sur des caractéristiques documentaires, mais sur des critères d'usage définis par les chercheurs à partir du contenu informationnel. Une base « structure » permet de présenter à la carte les divers documents en totalité ou en partie grâce à la pose de liens hypertexte.

Les documents non structurés sont répertoriés avec identification (titre, date de création, auteurs) et indexation par mots clés, puis le texte

intégral est mémorisé sous forme de fichiers PDF (Portable Document Format) ou HTML. La mise en ligne de fichiers PDF, simple et rapide, permet d'éviter le découpage fastidieux des documents en pages, mais convient peu à la lecture à l'écran. Cependant, le format PDF offre une grande qualité de reproduction et conserve les mises en forme d'origine : il convient parfaitement à la diffusion de docu-

ments à conserver dans une forme identique aux originaux. La version 5 de Lotus Notes, intégrant le format XML (Extensible Markup Language), devrait faciliter la mise en ligne de documents importants (thèses, rapports). Nous avons jusqu'à présent privilégié le choix d'un format de consultation interactive et non d'édition.

Un contenu validé et régulièrement enrichi

Pour constituer cette bibliothèque électronique, il faut, dans un premier temps, numériser les documents anciens jugés indispensables aux travaux de recherche actuels. On retrouve, à cette étape, le travail documentaire traditionnel : collecte, classement, indexation.

La collecte et le référencement seront élargis, après validation, aux documents disponibles sur le web. Une veille thématique sera régulièrement réalisée pour repérer les nouvelles pages, les mises à jour et identifier de nouveaux sites. Une fois validée, cette information est cataloguée, indexée et un lien est établi avec le site d'origine. Des outils spécifiques permettent de suivre les sites

repérés et d'être alerté de toute modification intervenant sur une page.

Les documents stockés vont pouvoir être enrichis par la pose de liens hypertextes qui vont permettre de relier les informations pour en faciliter la lecture et proposer des enrichissements complémentaires (liens intra-document, lien de type *voir aussi*, lien permettant d'accéder à la bibliographie ou à une définition, liens entre documents...). L'utilisateur peut ainsi naviguer selon son propre choix, sans suivre obligatoirement un cheminement linéaire (fig. 3).

Des synthèses, des résumés de cours ou de séminaires, et des articles sont écrits spécialement pour la plateforme et référencés par une adresse URL (Uniform Resource Locator) construite comme une référence bibliographique (comportant auteur, date, thématique, numéro d'identification), pour permettre de citer la page de manière explicite et de localiser directement le document sur le site.

Une rubrique nouveautés/actualités permet d'individualiser les derniers documents enregistrés sur la plateforme lors du dernier trimestre.

La rubrique documentation constitue une véritable bibliothèque électronique thématique. Elle propose un

catalogue des revues électroniques pouvant intéresser les chercheurs avec un repérage de leur localisation sur Internet et une présentation des informations mises à disposition par l'éditeur (sommaries, consignes aux auteurs...). En fonction des accords existants avec certains éditeurs, comme celui qui existe entre l'Inserm et Elsevier (Sciencedirect), il est possible d'aller de la référence bibliographique au texte intégral après identification préalable.

Pour l'accès aux références bibliographiques internationales, on peut décider de créer une base thématique constituée, le plus souvent, par la mise en commun de bases existantes, ou par importation automatique avec profils de fichiers provenant d'Embase, Medline, Pascal...

Des bibliographies sélectives permettent à l'utilisateur non spécialiste de s'informer rapidement sur un sujet en consultant la sélection proposée, extraite de la masse d'informations disponibles. Ces bibliographies peuvent être analysées, commentées et enrichies par la pose de liens hypertextes permettant depuis le nom d'un auteur d'accéder à toute sa bibliographie ou depuis un terme d'accéder au glossaire.

Dans la plate-forme Codecs, en économie de la santé, les références sont liées à une grille de lecture qui permet de recueillir les diverses données extraites de l'analyse détaillée d'articles sélectionnés.

Comme dans une bibliothèque, le site sera animé, certains documents mis en avant et des dossiers spécifiques réalisés à la demande.

Le spécialiste en information doit veiller à la qualité structurelle des documents et à la bonne application des règles définies en commun. Par sa formation technique, il doit assurer la responsabilité de la rubrique « liens vers d'autres sites », véritable carnet d'adresses spécialisé qui sera enrichi et validé collectivement par tous les membres du réseau.

Aujourd'hui plus que jamais les documents électroniques exploités sur les réseaux facilitent la circulation et la diffusion des connaissances. Mais il faut en assurer la visibilité, la qualité et la mise à jour régulière.

Un langage commun

Bien qu'un moteur de recherche (Topic de Verity) permette une interrogation sur tous les textes mémorisés, il est parfois commode d'indexer les documents avec des termes généraux pour permettre une vue globale par grandes thématiques. Ce classement doit être celui de l'« usager », correspondre à son cheminement intellectuel et celui-ci doit se l'approprier. Dans la plate-forme Rodin sur l'éthique, dix grandes thématiques ont été identifiées par les chercheurs, points forts des activités de recherche du réseau.

Lors de la saisie, chaque document subit un double classement, par type et par thématique, permettant ainsi, par filtrage automatique, la visualisation sur Internet selon ces deux possibilités au choix.

Mais, pour traiter l'hétérogénéité de l'information, il faut faire appel, en complément, à d'autres outils plus sophistiqués qui, associant analyse

statistique et linguistique, vont permettre de classer automatiquement les informations, d'extraire après traitement les concepts, de les visualiser sous forme de cartes dynamiques et d'établir des relations entre eux.

La visualisation de ces graphes de concepts offre à l'utilisateur une autre vision de l'information enfouie dans les documents et peut aider les experts à expliciter leurs connaissances.

Les spécialistes en information doivent assurer une veille technologique, tester, choisir et intégrer les nouveaux outils d'analyse textuelle, de « *text mining* » qui apparaissent sur le marché en nombre de plus en plus important.

Bilan

Réaliser une plate-forme thématique riche et de qualité représente pour les laboratoires de recherche une visibilité et une reconnaissance internationales qu'il ne faut pas négliger. Leurs travaux se trouvent valorisés par une diffusion élargie à d'autres « lecteurs » que ceux appartenant à la communauté habituelle. Le taux de fréquentation rend compte de cette notoriété et de cet impact.

La gestion et la diffusion sur le web d'un contenu validé, enrichi d'élé-

ments annexes représentent aussi un outil indispensable pour l'enseignement. En préservant la mémoire du laboratoire, il facilite l'appropriation des savoirs antérieurs par les étudiants. Le projet Rodin fait d'ailleurs partie du programme « Université du III^e millénaire ».

Depuis six ans, l'Inserm a développé un savoir-faire et des outils permettant aux chercheurs de travailler selon un processus coopératif, à partir d'une bibliothèque électronique adaptée à leurs besoins spécifiques et enrichie de banques de données complémentaires, le tout constituant un ensemble homogène, à point d'accès unique.

Au cœur de ce dispositif où les chercheurs sont à la fois acteurs et utilisateurs, le rôle des ingénieurs en information est primordial. Outre leurs compétences traditionnelles de documentalistes qui leur permettent de gérer, d'organiser et de classer des contenus de nature variée, nécessaires aux chercheurs, il leur faut désormais savoir animer des groupes de travail, travailler en réseau et pouvoir analyser le contenu : d'intermédiaires, ils doivent devenir de véritables partenaires au sein des réseaux de recherche.

Mars 2001

BIBLIOGRAPHIE

- COURTIAL**, Jean-Pierre, « Construction des connaissances scientifiques, construction de soi et communication sociale », *Solaris*, 1995
<http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2courtial.html>.
- DELAINE**, Virginie, « La gestion au quotidien des revues électroniques sur Intranet », *Documentaliste-Sciences de l'information*, 2000, 37 (3-4), p. 182-191.
- ERMINE**, J.-L., « Capter et créer le capital savoir, Réalités industrielles », *Annales de l'Ecole des Mines*, 1998, p. 82-86.
- KOCK**, N. ; **MC QUEEN**, R.J., « Groupware support as a moderator of interdepartmental knowledge communication in process improvement groups : an action research study », *Info Systems J*, 1998 (8), p. 83-198.
- LEVASSEUR**, D., « Intranet comme architecture documentaire privilégiée en entreprise : illustration d'une implantation simple dans un contexte ISO9000 », *Documentation et bibliothèques*, avril-juin 1999, p. 69-76.
- PINHAS**, Nicole ; **RASSER**, H., « Un partenariat documentalistes/chercheurs pour une gestion en réseau de l'information scientifique à l'Inserm », IDT, 8-10 juin 1999, Paris.
- LAU**, F., « Building a virtual network in a community health research training program », *Journal of the American Informatics Association*, 2000, 7 (4), p. 361-377.
- SOBIESKI**, N., *Le rôle du centre de documentation par rapport à la gestion des connaissances : un savoir-faire participant au mouvement de la connaissance*, mémoire de maîtrise des Sciences de l'information et de la documentation, Université Paris I-Sorbonne, septembre 2000.