

## L'INIST/CNRS, du document scientifique à la veille, un accès privilégié à l'information biomédicale internationale

*The INIST/CNRS : From Scientific Document to Scientific Watch,  
Exceptional Access to International Biomedical Information*

Fagherazzi-Pagel H (\*), Fohanno V (\*\*), Morel-Pair C (\*\*\*)

### Résumé

L'Institut de l'information scientifique et technique (INIST-CNRS) est à la fois lieu d'expertise en information scientifique et technique (IST) et lieu de développement de nouvelles technologies de l'information et de la communication (NTIC).

Son ambition est la collecte, le traitement et la diffusion de l'IST, en intégrant la multidisciplinarité des domaines scientifiques, près de 40 % des documents concernant le domaine biomédical. Sa mission est d'augmenter la visibilité des recherches par l'intermédiaire de ses deux grandes bases de données uniques en Europe et de son fonds documentaire provenant du monde entier, mais aussi de donner au plus grand nombre accès à la recherche internationale, notamment par le développement de tous ses services sur Internet (fourniture de documents, interrogation de ses bases catalographiques et bibliographiques...).

Ses prestations de veille, développées en interne, qu'elles soient automatiques, personnalisées ou plus spécifiques, constituent une voie innovante pour filtrer, analyser et expertiser la masse d'information disponible, écrite ou électronique, qui se multiplie de façon exponentielle.

**Rev Med Ass Maladie 2001;32,2:167-75**

*Mots clés* : médecine, pharmacologie ; sciences biologiques ; France ; information scientifique et technique ; infométrie ; scientométrie ; bibliométrie, veille.

### Summary

The Institute of Scientific and Technical Information (INIST/CNRS) provides both expert evaluation of technical and scientific information (IST) and the development of new information and communication technologies (NTIC).

Its aim is to collect, process and distribute technical and scientific information through the integration of multidisciplinary scientific fields. Approximately 40% of the documents concern the biomedical field. Its mission is to increase research visibility through its two important databases which are unique in Europe and to offer access to international research to the greatest number of people, particularly through the development of all its Internet services (providing documentation, access to the catalogue and bibliography databases). Its automatic, more specific or personalised internally developed service of scientific watching constitutes an innovative means for filtering, analysing and expertly appraising the enormous amount of exponentially increasing written and electronic information available.

**Rev Med Ass Maladie 2001;32,2:167-75**

*Key words*: medicine; pharmacology; biologic sciences; France; scientific and technical information, informetrics; scientometrics; bibliometrics; scientific watch.

(\*) Médecin, ingénieur documentaliste, Institut de l'information scientifique et technique (INIST) – Centre national de la recherche scientifique (CNRS).

(\*\*) Spécialiste en information scientifique et technique (IST), ingénieur documentaliste, INIST/CNRS.

(\*\*\*) Médecin, spécialiste en IST, ingénieur documentaliste INIST/CNRS.

Adresse pour correspondance : Dr Hélène Fagherazzi-Pagel, INIST-CNRS, 2, allée du Parc de Brabois, 54514 Vandœuvre-lès-Nancy cedex, e-mail : fagheraz@imist.fr

## I. INTRODUCTION

La première revue scientifique, le *Journal des sçavants*, paraissait en France en 1665. Le monde de la science naissant put alors se répandre, et la communication écrite du savoir se multiplia sur le mode exponentiel.

Le CNRS créait, dès 1939, des centres de documentation multidisciplinaires pour appréhender un espace de diffusion devenu phénoménal, et bâtir une véritable mémoire de la recherche afin d'y puiser toute information permettant d'intervenir plus efficacement dans la compétition mondiale. Aujourd'hui, des centaines de milliers de périodiques parcourent le monde. Quelques milliers de publications sont devenues de réelles institutions où se jouent à la fois les carrières des chercheurs et les choix stratégiques nationaux en matière de recherche.

Grâce aux possibilités offertes par les nouvelles technologies de l'information et de la communication, l'INIST/CNRS, Institut de l'information scientifique et technique, héritier des centres de documentation scientifique du CNRS, lieu d'expertise et de développement de technologies de pointe, voyait le jour en 1988. Sa conception et son fonctionnement novateurs inauguraient le concept du « tout électronique » [1].

Son ambition était et demeure internationale, et sa volonté de transférer rapidement l'information repose toujours sur le concept de la transversalité de la connaissance. Le champ de la santé, qui bénéficie d'innovations biomédicales majeures en s'appuyant constamment sur l'apport d'autres disciplines, des sciences « dures » aux sciences humaines, est emblématique de cette intégration des savoirs. Connaître l'état des recherches, même hors de son secteur d'activité, est essentiel pour établir une stratégie : c'est l'« *informaction* » [2]... L'apport des réseaux, puis celui d'Internet, a fait évoluer et surtout modifié considérablement les pratiques, celles de la communication dans le temps et l'espace, celles de l'écriture scientifique elle-même, voire de ses règles.

Premier centre européen intégré d'information scientifique et technique, l'INIST-CNRS constitue pour cette raison une source unique de *collecte*, de *traitement* et de *diffusion* de l'information scientifique et technique expertisée, à destination de tous : praticiens, chercheurs, étudiants, universitaires, industriels, et décideurs. Il se dote de tous les moyens de développer l'information électronique et la veille, en partenariat avec les acteurs actuels et potentiels d'un marché en pleine mutation.

## II. TRAITER ET COLLECTER L'INFORMATION

L'INIST détient 23 000 titres de périodiques (dont 8 600 abonnements en cours) et 240 000 monogra-

phies, provenant du monde entier : 16 % en médecine et pharmacologie, 3 % en psychologie et 19 % en biologie, les sciences « dures » et les sciences humaines et sociales représentant le complément.

Dans chaque domaine, les spécialistes de l'INIST vont analyser une partie de ces documents (articles, comptes rendus de congrès, thèses, etc.) dans la plupart des domaines de la connaissance pour alimenter les bases de données bibliographiques multidisciplinaires et multilingues PASCAL et FRANCIS. Chaque année, ces experts vont réviser la couverture des périodiques analysés dans les bases de données de l'INIST en sélectionnant les titres à supprimer ou à intégrer.

### 1. Bases PASCAL et FRANCIS

La base PASCAL constitue un réservoir de 14 millions de références bibliographiques en sciences, technologie et médecine depuis 1973. PASCAL couvre la littérature internationale tout en donnant une priorité au traitement de la littérature française et européenne qui représente plus de 40 % des documents.

L'INIST analyse pour sa base PASCAL environ 5 000 titres internationaux dont 1 500 en sciences médicales (médecine, pharmacologie, psychologie) et 800 en biologie (soit respectivement 39 % et 18 % de la base PASCAL).

Les références bibliographiques de la base PASCAL contiennent des mots clés (français, anglais, espagnols) et des résumés (dans 80 % des cas). La présence des affiliations (noms des laboratoires d'appartenance des auteurs) depuis 1995 est un point fort car cela permet d'effectuer des études bibliométriques pour repérer des centres d'expertise et contacter les chercheurs.

Un autre atout de la base PASCAL est sa multidisciplinarité qui permet à l'utilisateur de trouver des références bibliographiques provenant de domaines connexes non couverts par des bases de données spécialisées. C'est un élément particulièrement appréciable dans le cadre d'une recherche de première approche ou sur un thème transversal. De grands domaines interdisciplinaires en sciences médicales y sont d'ailleurs bien représentés : médecine tropicale, santé publique, génie biomédical, etc.

PASCAL est interrogeable par Minitel (08 36 29 36 01), en ligne sur serveurs (Questel. Orbit, DIALOG, Datastar, etc.) et sur cédéroms. Un cédérom spécifique PASCAL Biomed a été développé en collaboration avec SilverPlatter. Il constitue une extraction de la base PASCAL depuis 1987 dans le domaine des sciences médicales (6 500 titres dans PASCAL Biomed dont 3 200 régulièrement analysés dans le domaine biomédical). Ce cédérom permet la consultation des données via Intranet avec le logiciel Webspirs et une interrogation simul-

tanée sur plusieurs bases ; PASCAL Biomed est le complément idéal des autres bases du catalogue Silverplatter : Medline, PsychInfo, Embase... PASCAL Biomed est aussi accessible en ligne sur le serveur français QWAM (<http://www.qwam.com>).

Une interrogation complémentaire sur la base FRANCIS peut aussi permettre une recherche transversale notamment dans le domaine du social étroitement lié à la santé. La base FRANCIS, unique au monde, signale plus de 2,5 millions de références bibliographiques en sciences humaines, sociales et économiques depuis 1972 (3 700 titres de périodiques analysés). Comme PASCAL, FRANCIS est accessible en ligne et sur cédéroms.

La littérature grise occupe aussi une place importante au sein de la base : sous ce terme, on désigne les documents non publiés ou à diffusion restreinte, qui échappent aux circuits traditionnels de l'édition et aux dispositifs de contrôles bibliographiques : congrès, rapports de recherche, etc. L'INIST est membre de l'association EAGLE et produit la partie française de la base européenne de littérature grise SIGLE. L'INIST attribue également les ISRN (*International Standard Report Number*), code unique d'identification des rapports, en tant qu'agence nationale ISRN.

## 2. Coopération entre l'INIST/CNRS et la BDSP en santé publique

La coopération entre l'INIST et la Banque de données de santé publique (BDSP) est basée sur un échange de références bibliographiques, dans le cadre d'un réseau d'une trentaine de centres spécialisés en santé publique, et permet un enrichissement respectif des bases PASCAL et BDSP dans ce domaine.

L'INIST analyse la littérature internationale (anglophone) issue des périodiques cœur en santé publique. La BDSP (<http://www.bdsp.tm.fr>), par l'intermédiaire des autres membres du réseau, permet d'étendre la couverture de PASCAL plus spécifiquement en information francophone et en littérature grise (rapports des observatoires régionaux de la santé, des services déconcentrés de l'Etat, statistiques nationales, documents de l'Assemblée nationale ou du Sénat, mais aussi thèses, ouvrages, etc.).

## III. DIFFUSER L'INFORMATION

Une des principales missions de l'INIST est de diffuser les travaux des scientifiques français et européens dans le monde et de donner à ceux-ci accès à la recherche mondiale. L'INIST traite 700 000 commandes par an soit 3 000 documents par jour pour 6 000 clients en France et dans le monde.

Pour les documents indisponibles dans son propre fonds, l'INIST les recherche auprès d'un réseau de

plus de 150 bibliothèques françaises et européennes partenaires. Les utilisateurs des services de l'INIST sont principalement les organismes de recherche publics, l'enseignement supérieur et les centres de recherche des grandes entreprises (laboratoires pharmaceutiques, compagnies pétrolières, etc.). Le développement de services sur le Web permet d'étendre aujourd'hui l'offre aux particuliers.

### 1. Article@inist

L'évolution technologique, et en particulier l'avènement du document électronique, a conduit l'INIST à développer ses accès à l'information électronique via Internet.

Il s'agit d'offrir à l'utilisateur les capacités d'identifier et de localiser un document dans une multitude de sources d'information.

Article@inist, accessible sur le site de l'INIST, est un ensemble de bases catalographiques. Ce service permet de consulter gratuitement les catalogues du fonds de l'INIST pour retrouver une référence d'article, de revue ou de monographie (près de 6 millions de références depuis 1990) et de commander ensuite le document directement en ligne. ArticleSciences (<http://articlesciences.inist.fr/accueil.phhtml>), est une variante de la base Article, facilitant la commande ponctuelle d'un document, et son règlement en ligne sans ouverture de compte au préalable.

Pour améliorer les délais d'obtention des documents à domicile, un nouveau service de livraison électronique est également proposé à l'INIST depuis le début de l'année 2001. Il nécessite l'acquisition du logiciel Ariel®, développé par le *Research Library Group* (RLG), pour recevoir le document souhaité sur son micro-ordinateur.

### 2. ConnectSciences : portail CNRS d'information scientifique et technique

Au centre de son dispositif d'offre de services sur Internet, le portail CNRS en Information scientifique et technique est né de la volonté de l'INIST-CNRS de mettre en ligne des ressources en IST, fruit de ses traitements mais aussi issues du monde de ses partenaires et de l'Internet.

Ce portail (<http://connectsciences.inist.fr>) est en libre accès depuis mai 2001. La possibilité de personnaliser sa connexion permet à l'utilisateur de bénéficier d'une gamme de produits et services correspondant à ses besoins dans une ou plusieurs disciplines scientifiques, par exemple.

#### • ArticleSciences

Outre la consultation de son catalogue d'articles (ArticleSciences : 7 millions de références d'articles depuis 1990) et la commande de copies d'ar-

tibles en ligne, le paiement sera désormais possible par carte bancaire. ArticleSciences est présenté sous une interface très conviviale, avec mise à jour quotidienne.

- **Actualités de PASCAL et FRANCIS**

ConnectSciences offre un accès gratuit à la mise à jour de PASCAL et FRANCIS. Les trois derniers mois de PASCAL, avec chargement hebdomadaire de 3 000 notices, et ceux de FRANCIS sont interrogeables via le logiciel LexiMine. Cet outil d'aide à la formulation de requêtes permet d'extraire par simples clics des concepts issus de ces bases de données (titre, mots clés et résumé en français et en anglais, sur une période d'un an).

- **Autres sources**

Une recherche sur les bases PASCAL ou FRANCIS, à partir de LexiMine, permet d'extraire et d'étudier la terminologie scientifique employée par les auteurs, puis d'envoyer une requête successivement sur un ensemble de bases ou moteurs de recherche disponibles (brevets, bases concurrentes, moteurs de recherche, etc.). Le navigateur ouvre alors une nouvelle fenêtre indépendante du portail, pour se connecter aux bases et moteurs sélectionnés.

- **Base Thésa**

La base Thésa est une base de données française et gratuite. Elle signale les sujets de thèse en cours dans les grandes écoles appartenant à la Conférence des grandes écoles (CGE). Le signalement est conservé jusqu'à un an après la soutenance. Cette base permet de connaître les thèmes des recherches menées dans ces laboratoires et d'identifier les acteurs de ces travaux.

- **Base interactive de liens et d'Agenda-Évènements**

ConnectSciences propose une base de données interactive de liens, véritable répertoire d'adresses Internet au contenu à caractère scientifique et technique, multidisciplinaires, multilingues, validées par des scientifiques CNRS, ainsi qu'une base agenda d'évènements scientifiques (congrès, conférences, etc.). Devant la prolifération des sites et les dérives possibles, l'information en santé demande une analyse et une attention soutenues. Dans tous les cas, l'information est validée.

- **Bibliosciences**

Grâce à un accord entre le CNRS et SilverPlatter, éditeur de cédéroms scientifiques, les 1 300 unités CNRS peuvent bénéficier gratuitement du service BiblioSciences, accès illimité en ligne sur le serveur de l'INIST aux différentes bases de données essentielles dans leur domaine (INSPEC, Medline, Current Contents, PASCAL, FRANCIS...).

- **Outils documentaires et dossier**

On peut consulter en ligne : la terminologie utilisée par l'INIST, les titres des revues du fonds de l'INIST, ainsi que ceux analysés dans les bases PASCAL/FRANCIS, les plans de classement. Ces codifications, liées à une thématique scientifique donnée, représentent un critère de recherche supplémentaire important pour interroger les produits serveurs et les cédéroms de l'INIST.

Un service personnalisé « Dossier » permet de manipuler des éléments très divers, tels que des références bibliographiques ou des URL (adresses Internet), qui peuvent alors être stockés en vue d'un traitement ultérieur : tri, impression, transfert vers un autre utilisateur du portail.

## IV. DÉVELOPPER LA VEILLE ET L'INNOVATION EN IST

### 1. Objectifs et contexte

L'information est, plus que jamais en ce début de millénaire, indispensable à l'élaboration de stratégies de recherche et d'innovation, de production ou même de travail de terrain, si l'on considère l'évolution rapide des sciences et technologies ou des problématiques sociales, juridiques et économiques et leur mondialisation.

Ce ne sont pas les sources d'information qui manquent ; on recense à ce jour, par exemple, près d'un million de revues scientifiques et techniques publiées dans le monde, 450 bases de données contenant chacune plusieurs millions de références sur le seul serveur américain *Dialog*, des centaines de millions de pages Internet répertoriées par les moteurs de recherche. Face à cette multiplication des sources, écrites ou électroniques, validées ou non, complémentaires mais aussi redondantes, l'entreprise, le laboratoire ou l'organisme administratif ont besoin de trouver rapidement ce qui leur sera réellement utile. C'est le but de la « veille » de fournir « *la bonne information à la bonne personne au bon moment pour une bonne décision* », selon une définition donnée par M. Porter ; l'information disponible est traitée comme une matière première pour en extraire l'information utile.

La recherche documentaire classique, réalisée de façon ponctuelle (recherche rétrospective) ou régulièrement renouvelée (diffusion sélective de l'information), fait un premier tri en ramenant les documents ou les notices bibliographiques à partir d'une question comportant les mots clés souhaités. Cependant, les résultats sont présentés sous la forme d'une liste brute de notices, en général par ordre chronologique, sans critère de pertinence ni aide à l'analyse des résultats.

Les prestations de veille ont pour objectif d'apporter une valeur ajoutée dans le filtrage, l'analyse et la synthèse de l'information mise à disposition ; elles se déclinent finement en fonction des besoins des utilisateurs, depuis la représentation de l'information contenue dans de gros corpus documentaires après analyse automatisée (infométrie) jusqu'à la recherche ciblée et totalement personnalisée des quelques sources les plus pertinentes, documents ou experts du domaine, avec éventuellement une synthèse des résultats. Ces produits permettent donc à la fois d'obtenir une vue d'ensemble et une première approche d'un sujet, de suivre et analyser l'évolution d'un domaine, de repérer des thèmes en émergence, et constituent un outil précieux de connaissance et de décision.

La veille stratégique est actuellement une priorité du ministère de la Recherche ; la Direction de la technologie a lancé en avril 2001 un appel d'offres sur des projets permettant la maîtrise de la « sur-information textuelle » sur Internet par des méthodes de filtrage et de résumé de l'information disponible.

## 2. Produits innovants et de veille en développement à l'INIST/CNRS [3, 4, 5, 6, 7, 8]

Trois niveaux de prestations de veille ont été définis à l'INIST, essentiellement des produits électroniques : des prestations *automatiques*, *personnalisées* ou *plus spécifiques*.

### • Prestations automatiques

Ces prestations, mises à disposition par l'intermédiaire du portail de l'INIST, utilisent notamment les fonctions de personnalisation et d'espace communautaire développées dans le portail et la technologie « push-mail », qui permet une livraison électronique automatique régulière à l'adresse souhaitée.

Dans un premier temps sont envisagés :

- une revue de sommaires sur une collection de périodiques définie par l'utilisateur en fonction de ses besoins de surveillance ;
- un service d'alerte régulière sur les notices nouvellement entrées dans nos bases et portant sur les sujets définis par l'utilisateur ;
- la constitution de carnets d'adresse Internet à partir de ses thèmes d'intérêt, avec possibilité d'alerte quand il y a des nouveautés. Selon leur niveau, elles seront gratuites ou payantes.

### • Prestations personnalisées, avec des projets dans deux directions complémentaires

Seront proposés des produits multibases, issus de l'interrogation de plusieurs bases de données choisies en fonction du sujet et de leur complémentarité

avec les bases produites par l'INIST, et des produits multisources, qui intègrent de plus la recherche sur Internet, avec possibilité d'une synthèse des résultats ; ces produits permettront d'assurer une meilleure exhaustivité et une surveillance rapprochée d'un sujet. Les recherches effectuées pourront être ponctuelles ou actualisées régulièrement dans le but d'une veille sur le sujet.

Par exemple, l'application *Vetech* développée à l'INIST permettra l'acquisition, l'analyse et la diffusion sélective d'information par axes de surveillance dans tous les domaines, avec interrogation de bases de données généralistes et spécialisées et exploration des ressources Internet pertinentes, choisies au départ et évaluées régulièrement, qu'il s'agisse de sites, de listes de discussion ou de forums. Ces produits entrent dans le cadre de la législation sur la propriété intellectuelle, ce qui demande de résoudre préalablement à leur mise sur le marché un ensemble de questions juridiques et commerciales.

Seront proposés en parallèle des « plates-formes d'investigation documentaire » : celles-ci ont pour objectif d'analyser et de visualiser la répartition des notices d'un corpus documentaire de gros volume selon différents critères, contenu scientifique (mots clés, mots du titre ou du résumé...) et contenu factuel (auteurs, organismes d'affiliation, source documentaire...).

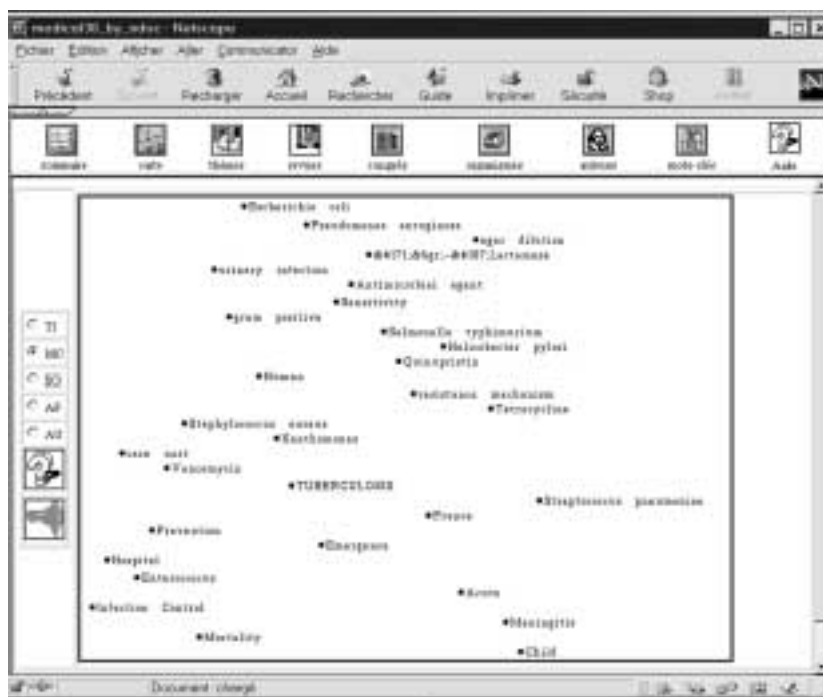
Utilisant des méthodes infométriques de calcul de proximité des documents, elles génèrent des sous-ensembles de documents voisins entre eux, nommés « clusters » ou « agglomérats », et une représentation graphique des résultats, avec des liens hypertexte permettant de naviguer entre et dans les sous-ensembles, jusqu'aux documents les plus pertinents pour l'utilisateur. Ces plates-formes permettent donc d'explorer de manière aisée, synthétique, guidée et dynamique l'information recelée par le corpus, notamment les sous-thématiques d'un sujet, le vocabulaire utilisé, les auteurs ou organismes souvent associés, les sources documentaires principales, les idées ou associations émergentes...

Trois outils d'investigation documentaire sont disponibles à l'INIST :

- la plate-forme *STANALYST*, qui utilise des méthodes infométriques riches et variées, issues soit des méthodes statistiques et mathématiques, soit de la recherche sur l'intelligence artificielle, et présente les résultats sous forme de cartographies illustrant la proximité ou la distance entre les groupes de documents (Tableau I) ;
- la plate-forme *DILIB*, qui présente les résultats sous forme de tableaux et de graphiques en fonction de la taille des sous-ensembles et permet d'envoyer automatiquement sur Internet les requêtes

Tableau I  
Microbiologie : cartographie sous HENoch, interface graphique de STANALYST

Cette carte présente une étude en microbiologie sur la résistance aux antibiotiques, étudiée du point de vue médical, sur 2 642 résumés d'articles publiés entre 1997 et 1999.



#### Commentaire

La plupart des classes obtenues regroupe les différents aspects : épidémiologie, pathologie et traitement, caractérisation de la résistance par l'intermédiaire des tests de sensibilité médicamenteuse (TSM) qui se réfèrent à l'intitulé de la classe.

La carte ci-dessus permet d'avoir une vue d'ensemble des classes obtenues. Les aspects épidémiologie, prévention et environnement se regroupent vers le coin gauche en bas de la carte (« Care unit », Hospital » et « Infection control »). Sur la partie droite et vers le bas de la carte, s'isolent 3 classes : « Streptococcus pneumoniae », « Child » et « Meningitis ». En effet, « Streptococcus pneumoniae » provoque des méningites chez les enfants. De plus, cette bactérie provoque également des otites chez les enfants, problématique qui représente une part importante des documents associés au thème « acute » ce qui explique donc la position de cette classe dans cette zone de la carte.

Dans la partie supérieure de la carte sont regroupées les classes de la catégorie des antibiotiques se mêlant avec les classes de la catégorie des micro-organismes. Dans cette partie de carte, la proximité entre les classes : « Tetracycline » et « Agar dilution » s'explique par le fait que les documents associés à ces deux classes portent sur l'aspect méthodologique des tests de sensibilité (marqueur : tétracycline, milieu : agar).

correspondant aux sous-thématiques intéressant l'utilisateur (Tableau II) ;

– le logiciel *LexiMine*, outil d'exploration des termes associés dans un corpus, d'aide à la formulation de questions et d'interrogation d'Internet (Tableau III).

#### • Prestations d'infométrie plus spécifiques

Elles correspondent à un objectif de recherche dans un domaine, à l'élaboration d'indicateurs scientifiques ou demandant un travail personnalisé d'ingénierie documentaire pour répondre à la question posée.

Ce type d'étude est réalisé essentiellement dans le cadre d'accords institutionnels avec de grands organismes de recherche français ou étrangers. Ces applications ont ainsi permis d'effectuer pour le compte de différents organismes une trentaine d'études validées par des experts du domaine proposant des classifications automatiques et des représentations graphiques sur des ensembles de notices bibliographiques. L'INIST présente également ces plates-formes d'investigation documentaires lors des rencontres professionnelles organisées mensuellement pour ses clients, sur différents thèmes de recherche proposés par ceux-ci, ou à l'occasion de salons et congrès. Différents

Tableau II  
Dossier médical : regroupement thématique de notices avec la plate-forme DILIB

Cet extrait de l'analyse fournie par DILIB sur le sujet « Dossier médical » est le résultat du traitement d'un corpus de 900 notices, sur la période 1995-2001.

Nom de cluster	F	nb de doc
Cluster information - Information	82	430
Base de données - Base de données	40	212
Informations - Informations	31	260
Informations - Accessibilité	27	206
Base de données - Informations	24	219
Cluster - Sécurité	15	26
Médecine travail - Exposition professionnelle	15	18
Informations - Informations	9	116
Base de données - Base de données	9	114
Médecine - Santé publique	8	28
Cluster - Sécurité	7	15
Informations - Informations	7	40
Base de données - Base de données	7	111
Application médicale - Base de données	7	51
Cluster - Santé	6	8
Cluster - Transfusion	6	29
Cluster - Santé	6	13
Cluster - Santé	6	36

#### Commentaire

Les clusters constitués automatiquement en fonction de l'occurrence de mots clés proches dans les notices montrent bien l'intérêt actuel pour l'informatisation du dossier médical dans le cadre de la médicalisation du système d'information de l'hôpital ; ils mettent en évidence :

- beaucoup d'aspects techniques et technologiques, centrés sur les mots « base (ou enregistrement) donnée, téléinformatique, réseau, technologie, innovation, architecture système » ;
- des implications sur l'accessibilité et la confidentialité des données personnelles médicales (« réglementation, sécurité, accessibilité ») et sur les pratiques des personnels soignants ;
- des applications en santé publique et épidémiologie, un questionnement en psychiatrie, un intérêt pour la prise de décision.

Le cluster « médecine du travail, exposition professionnelle » regroupe deux types de notices, les unes sur le dossier médical et la collecte des données d'exposition professionnelle, les autres sur l'informatisation des données.

Le cluster « transfusion, sang » rassemble des notices sur l'importance du dossier médical dans la traçabilité des produits sanguins.

(La colonne Fij donne le nombre de notices contenant les deux termes les plus fréquents dans le cluster, la colonne « nbre de doc » affiche la taille totale du cluster constitué de proche en proche.)

sujets médicaux, comme le PMSI, le dossier médical, le suicide de l'adolescent ou la thérapie génique, ont donné des résultats intéressants. Des exemples sont accessibles librement sur le site de l'INIST dans la rubrique « Bibliosites » et seront présents ultérieurement dans le portail.

### 3. Les acteurs veille et innovation à l'INIST

Ces activités de veille et d'innovation en information scientifique et technique mettent en œuvre des compétences diverses, documentaires, scientifiques et informatiques en étroite collaboration avec l'utilisateur. À l'INIST, les activités de recherche et de développement de produits de

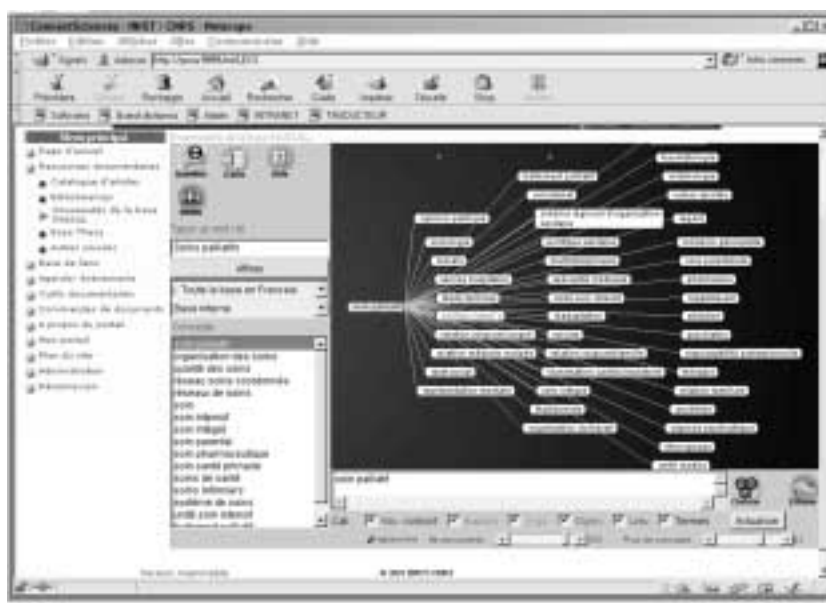
veille sont réparties dans différents départements aux activités complémentaires.

L'unité de recherche et d'innovation (URI) a une activité de recherche propre et une capacité d'innovation technologique dans les domaines des mathématiques appliquées à l'analyse de l'information, du traitement informatique du langage naturel en gros corpus, de l'application des techniques symboliques et numériques de l'intelligence artificielle.

Le département des produits et services (DPS) dispose d'une cellule de veille destinée à mettre au point de nouveaux produits et services à partir des outils développés en interne ou à l'extérieur et des besoins des clients.

Tableau III  
Soins palliatifs, analyse des termes associés sous LexiMine

Cette carte représente l'analyse des termes associés au concept « soins palliatifs » dans un corpus de notices multidisciplinaires PASCAL, sur les « trois derniers mois ».



#### Commentaire

Le logiciel LexiMine donne une liste des termes cooccurrents ou mots associés (retrouvés fréquemment ensemble). « Carte » affiche l'écran principal de LexiMine, « Stats » permet de représenter sous forme graphique l'évolution dans le temps de la fréquence des termes. Un clic sur les termes souhaités permet d'élaborer une équation de recherche et d'interroger d'autres sources sur Internet.

L'aspect santé publique (politique sanitaire) de l'organisation des soins palliatifs, en France et dans d'autres pays, notamment en Angleterre, semble être largement publié dans la littérature courant 2001, de même que l'aspect social du « mourir ».

On peut discerner l'intérêt de la couverture en psychologie, les références sur le vécu de la personne hospitalisée, les relations du soignant avec le malade en fin de vie et sa famille. Une partie des notices traite de la réanimation (soins intensifs), et des prises de décision concernant l'indication des traitements invasifs en stade terminal.

Le département des prototypes et moyens informatiques (DPMI) assure le développement de prototypes, l'industrialisation, l'intégration et la maintenance des autres outils.

Le département prospective et relations extérieures (DPRE) étudie les besoins des utilisateurs et négocie les partenariats nécessaires à l'évolution des services.

#### V. CONCLUSION

Après trois siècles de communication écrite puis numérique, les scientifiques du monde entier ont désormais à leur disposition moult services et produits toujours plus performants : mise à disposition d'un patrimoine, rapidité de son exploration, interactivité des interfaces d'accès.

De la structuration de l'information à l'évaluation globale de la recherche (infométrie), en passant par la quête de l'information pertinente, tous ces traitements « à valeur ajoutée », devraient s'accompagner d'échanges où visibilité des travaux et lisibilité des « écrits » devraient toujours coexister.

Devant la course aux publications et la partialité d'Internet, souhaitons que le but premier des publications, celui de diffuser les connaissances, ne soit pas trop altéré par ses modes de gestion. A l'ère de l'Information scientifique et technique foisonnante de matière et d'idées, l'INIST-CNRS propose ses moyens humains, juridiques et organisationnels pour offrir au plus grand nombre : outils, conseils et technologies, afin de mieux choisir.

*Remerciements à Roger Hung Kung Sow, médecin, ingénieur documentaliste, INIST-CNRS.*



---

## RÉFÉRENCES

1. Côme C, Lacroix V, Fagherazzi-Pagel H, Arnould F, Petitjean S. L'INIST-CNRS : un accès privilégié à la documentation sociale. *Vie sociale. Sources de la documentation sociale* 2000;3:65-77.
2. Frochot D, Colnard C. Vous avez dit photocopillage ? Nous disons « information » ! *Documentaliste Science de l'Information* 1994;31 ; 1:35-7.
3. Ducloy J, Charpentier P, François C, Grivel L. Une boîte à outils pour le traitement de l'Information. *Scientifique et Technique. Génie logiciel* 1991;25:80-90.
4. François C, Grivel L. Les outils de l'unité de recherche et innovation de l'INIST. *La lettre de l'URFIST* 1999;21:10-8.
5. Grivel L, HENOCH, un outil d'analyse de corpus d'information scientifique et technique. *Le Micro Bulletin Thématique, CNRS-DSI L'information scientifique et technique et l'outil Internet. Expériences, recherches et enjeux pour les professionnels de l'IST* 1999;3:27-44.
6. Polanco X. Aux sources de la scientométrie. *SOLARIS* 1995;2:13-79.
7. Polanco X. Plus que d'un système d'information : il s'agit de transformer l'information en connaissance et la connaissance en action. *Le Micro Bulletin Thématique, CNRS-DSI L'information scientifique et technique et l'outil Internet* 1999;3:15-25.
8. Polanco X, François C, Keim JP. Artificial neural network technology for the classification and cartography of scientific and technical information. *Scientometrics* 1999;41;1:69-82.