



# Etude bibliométrique des différents aspects de la mort subite du nourrisson traités dans les revues scientifiques

**Cet article présente un exemple d'analyse bibliométrique menée avec l'aide d'un chercheur effectuant un travail expérimental sur la mort subite du nourrisson - sujet auquel aucune revue n'est exclusivement consacrée. L'étude bibliométrique consiste ici à rechercher, dans un corpus de références bibliographiques téléchargées, quels sont les aspects de la mort subite du nourrisson traités par telle ou telle revue scientifique : il s'agit d'évaluer le contenu thématique de ces publications lorsqu'elles abordent ce sujet. Une telle méthode, qui peut évidemment être appliquée à d'autres sujets, pourrait aider bibliothèques et centres de documentation à faire des choix d'abonnements et à orienter des recherches bibliographiques.**

J E A N - D O M I N I Q U E P I E R R E T  
G É R A R D H I L A I R E  
L U C Q U O N I A M , H E N R I D O U

Jean-Dominique Pierret travaille à la Direction médicale et scientifique de FDM Pharma, Les espaces de Sophia, 80 route des Lucioles, 06560 Sophia Antipolis, téléphone 92 96 68 09.

Luc Quoniam et Henri Dou sont chercheurs au Centre de recherche et de rétrospective de Marseille (CRRM), Centre scientifique Saint-Jérôme, Université d'Aix-Marseille III, 13397 Marseille Cedex 20, téléphone 91 28 87 48.

Gérard Hilaire travaille dans la même université, au sein de l'équipe de recherche Biologie des rythmes et du développement (ERS CNRS 061).

La mort subite du nourrisson (MSN) est un sujet peu médiatisé [1]. Elle est cependant la première cause de mortalité infantile dans les pays occidentaux, avec un taux de 1,2 décès pour mille naissances en France (1 444 victimes en 1987). Les recherches sur la MSN se font dans plusieurs directions, dont voici les principales :

- aspect respiratoire : la mort est causée par un étouffement dû soit à une mauvaise régulation de la respiration par le système nerveux central (immaturité, mauvais développement

- ou dysfonctionnement des structures nerveuses impliquées dans le contrôle de la respiration), soit par une anomalie de l'appareil respiratoire ;
- état de vigilance, contrôle des fonctions vitales pendant le sommeil ;
- études épidémiologiques : facteurs de risque selon la position du nouveau-né lorsqu'il dort, antécédents, environnement social et physique

- (par exemple le matériau dans lequel est fabriqué la couverture du lit) ;
- contrôle cardiovasculaire ;
- reflux gastro-œsophagien (passage d'une partie du contenu de l'estomac dans l'œsophage).

Bien entendu, il est fréquent de trouver des articles traitant de plusieurs de ces aspects en même temps : les aspects vigilance et respiration sont souvent traités de front, ainsi que la vigilance et les reflux gastro-œsophagiens. De même les méthodes employées varient, et peuvent être combinées entre elles (anatomie, autopsie, électrophysiologie, biologie moléculaire, enquête, radiologie, monitoring, biochimie, etc.).

L'étude bibliométrique de ce sujet ne peut se concevoir sans l'aide d'un spécialiste du domaine, qui va définir les points critiques à étudier, comme dans toute opération de veille technologique [2]. Partant d'un téléchargement de références bibliographiques sur la MSN, cet article présente le travail effectué pour réaliser l'analyse demandée par le spécialiste : quelles revues traitent de la MSN, et sous quels aspects ? Une analyse en composante principale (ACP) est réalisée sur une matrice de fréquence contenant en colonne les noms des journaux, et en ligne les descripteurs (mots-clés).

## Matériel, outils et méthodes

Le matériel utilisé est un ordinateur PC 486 SX 25, avec MS-DOS 5.0 et les logiciels suivants :

- CrossTalk XVI (1), pour réaliser le téléchargement ;
- InfoTrans (2), pour reformater le champ « Source » (SO dans la base *Pascal*) ;
- DataView (3), pour extraire et éditer les formes (noms des revues ou descripteurs), ou les paires de formes, pour calculer la répartition des formes sur les références, et construire la matrice de fréquence journaux/descripteurs ;
- STAT-ITCF (4), pour réaliser l'analyse en composantes principales ;
- l'éditeur du MS-DOS (de Microsoft) pour effectuer les corrections.

Le travail à réaliser est le suivant : faire une matrice de fréquence jour-

### Stratégie d'interrogation de Pascal

*Question 1* : Sudden death ou mort subite

2 637 réponses

*Question 2* : infant ou nourrisson

36 890 réponses

*Question 3* : (sudden death ou mort subite) et (infant ou nourrisson)

904 réponses

*Question 4* : infant ou nourrisson ou foetus

75 594 réponses

*Question 5* : (infant ou nourrisson ou foetus) et (sudden death ou mort subite)

911 réponses

Sont sélectionnées les références dont la date d'incorporation dans *Pascal* est comprise entre 1990 et 1992. Après avoir effectué cette sélection, la ré-exécution des questions donne :

*Question 1* : 622 réponses

*Question 2* : 6 558 réponses

*Question 3* : 201 réponses

*Question 4* : 14 424 réponses

*Question 5* : 203 réponses

Les 203 références de la question 5 sont téléchargées. Elles sont limitées à 1990-1992, et ont comme mots-clés :

- sudden death ou mort subite
- et
- infant ou nourrisson ou foetus.

naux/descripteurs, pour effectuer, avec le logiciel STAT-ITCF, une analyse en composantes principales.

L'étude porte sur des références extraites de la banque de données *Pascal*, produite par l'INIST (5), sur le serveur Questel (6). Cette banque est scientifique et multidisciplinaire. Elle n'est peut-être pas aussi complète dans les domaines qu'elle aborde (physique, chimie, biologie, géologie, etc.) que le sont des banques thématiques comme *Chemical Abstracts* (chimie) ou *Medline* (médecine, biologie). Mais *Pascal* présente l'avantage d'être une banque de données française, et donc son contenu accordera une part plus importante aux travaux d'équipes françaises que ne l'aurait fait un fichier américain par exemple.

L'interrogation porte sur des mots-clés, et non sur les codes INIST (encadré ci-dessus). Ces derniers ne permettent pas une recherche sélective sur le sujet de la MSN.

La première opération à réaliser est la mise en forme correcte du téléchargement [4]. Des modifications de

deux types principaux sont apportées :

- correction des noms des journaux ;
- traitement des descripteurs français : ils proviennent d'un thésaurus qui n'est pas performant en ce qui concerne ces références. Par exemple, on peut trouver dans deux références différentes les deux descripteurs suivants : *rythme respiratoire*, et *fréquence respiratoire*, qui décrivent le même phénomène !

(1) Logiciel distribué par Digital Communications Associates Inc., 1000 Aldeman Dr., Alpharetta, GA 30 202-4199, Etats-Unis.

(2) Logiciel distribué par I+K France, 9 avenue de Villepreux, 78340 Les Clayes-sous-Bois.

(3) Logiciel conçu par le CRRM, voir référence [3].

(4) Logiciel conçu par l'Institut technique des céréales et des fourrages (ITCF), 8 avenue du Président Wilson, 75116 Paris.

(5) Institut de l'information scientifique et technique, 2 allée du Parc de Brabois, 54500 Vandœuvre-lès-Nancy.

(6) Questel, Immeuble Le Capitole, 55 avenue des Champs-Pierreux, 92012 Nanterre cedex.

Nous avons donc traité ce champ en opérant un éclatement de tous les descripteurs, puis en regroupant les termes par affinité sémantique.

Le champ « Source » des références contient, en plus du nom de la revue, le nombre de pages de l'article, le volume, le numéro et l'année de parution. Seul le nom de la revue est

nécessaire, les autres renseignements parasitant le traitement informatique. Le logiciel InfoTrans permet de reformater le champ source, en ne conservant que le nom de la revue.

Dans le fichier télédéchargé, le nom d'un même journal est parfois écrit de plusieurs façons différentes. Par exemple, les deux titres suivants

désignent le même périodique : *Lancet (The) (British edition)* et *Lancet : (British édition)*. Pour connaître les différentes formes graphiques que prend le nom d'un même journal, il faut éditer, avec le logiciel Dataview, la liste des noms des journaux, triés par ordre alphabétique. La consultation du CD-Rom *Myriade*, qui est une version abrégée du *Catalogue collectif national des publications en série* (CCN), permet de connaître le nom exact de chacun d'entre eux.

Le logiciel STAT-ITCF accepte des matrices dont les intitulés contiennent au maximum trois lettres. Dataview réalise automatiquement cette opération.

En observant la répartition des termes du thésaurus de *Pascal* (figure 1), on remarque que 90 % des termes ont une fréquence inférieure à 5. Ces termes sont composés soit de simples mots ou unitermes (comme *respiration*), soit de plusieurs mots ou pluritermes (*appareil respiratoire pathologie*). Si on « éclate » les termes du thésaurus, par exemple, *appareil respiratoire pathologie* donnerait lieu à trois unitermes : *appareil*, *respiratoire* et *pathologie*.

D'un pluriterme du thésaurus, on passe à trois unitermes, dont deux ont une signification pour la MSN (*respiration* et *pathologie*). Puis si on observe la répartition des « termes éclatés » du thésaurus (figure 2), on se rend compte que, par cet artifice, on augmente le nombre de mots à forte fréquence (87% des termes ont une fréquence inférieure à 5).

Dans le télédéchargement, 606 descripteurs différents sont employés. En « éclatant » les descripteurs du thésaurus, on trouve que 633 mots différents sont utilisés pour composer ces 606 descripteurs.

Avec l'avis d'experts du domaine, nous avons regroupé certains termes issus de l'éclatement des descripteurs sous un même terme générique, par exemple :

Figure 1 - Graphique des termes du thésaurus

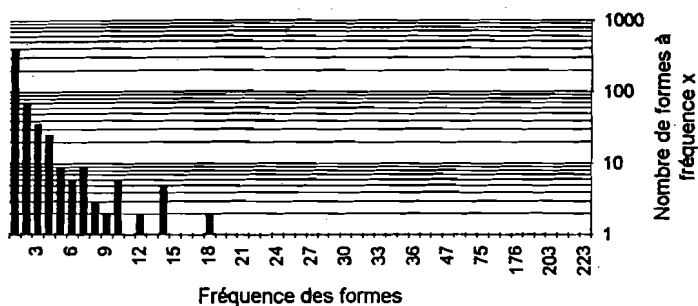


Figure 2 - Graphique des termes « éclatés » du thésaurus

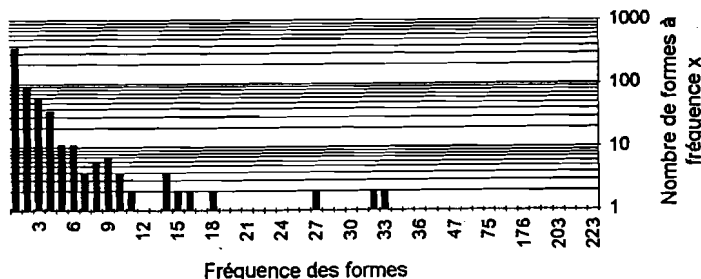
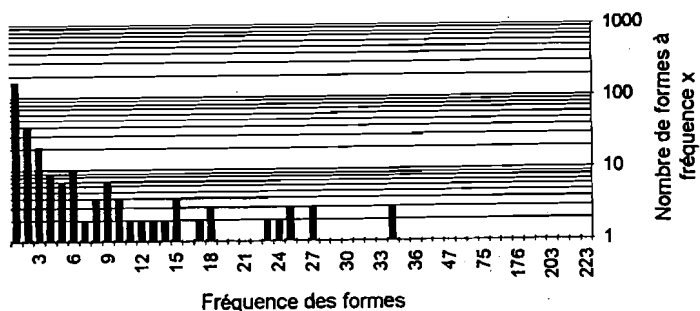


Figure 3 - Graphique des formes simples



- *respiration* englobe les termes *apnée, respiration, respiratoire ventilation, etc.* ;

- *micro-organisme* englobe les termes *aureus, bactériose, bactérie, micro-organisme, cytomégalovirus, etc.* ;

- *cérébral* englobe les termes *bulbe, cervelet, calleux, cérébelleux, cérébral, encéphale, etc.* ;

- *immunité* englobe les termes *C4, IGA, IGM, IGG, immune, immunité, immunologie, macrophage, etc.*

Ces regroupements, effectués avec l'aide de spécialistes du domaine, permettent, comme le montre le graphique des formes simples (figure 3), d'augmenter le nombre de termes à forte fréquence (73 % des termes ont une fréquence inférieure à 5). Cette opération augmente le nombre de descripteurs à fréquence élevée, donc le nombre de descripteurs potentiellement significatifs.

Les matrices « creuses » sont difficiles à analyser avec les méthodes classiques d'analyse de données [5]. Seuls les journaux qui apparaissent trois fois et plus dans les références sont pris en compte. Ils sont au nombre de 21 et représentent 59,7 % des références. Parmi ces journaux, on relève *Neurochemistry International* qui apparaît dans 5 références. Il se trouve que ces 5 références proviennent du même numéro de ce journal, qui est soit un numéro spécial, soit un numéro composé suite à un colloque. Cela ne veut pas dire que ce journal traite particulièrement de la MSN. Cet exemple montre que les résultats de l'analyse bibliométrique dépendent du corpus de références traité. *Lancet* qui apparaît 20 fois sur les références, et *Sleep (New York)* qui n'apparaît que 3 fois ne seront pas analysés avec la même finesse.

Les descripteurs pris en compte ont une fréquence allant de 5 à 84 apparitions dans les références. Cela représente une palette de 73 descripteurs, sur les 323 recensés (soit 22,6 % des descripteurs). Les trois descripteurs suivants : *mort, subite* et

*nourrisson*, ne sont pas pris en compte, car ils sont communs à toutes les références. Si on tient compte du nombre de fois qu'un descripteur est utilisé, les 73 descripteurs retenus représentent alors 53,5 % des utilisations.

Dataview permet de sélectionner les journaux et les descripteurs à prendre en compte pour la construction de la matrice selon leur fréquence. Cette matrice comporte 21 journaux (colonnes) et 71 descripteurs (lignes).

La méthode d'analyse de données choisie est l'analyse en composantes principales (ACP) [6]. L'ACP qui est une ordination en espace réduit, propose une interprétation de la matrice avec un nombre réduit de dimensions, toutes indépendantes, c'est-à-dire orthogonales entre elles. Elle va tendre à regrouper les journaux qui utilisent sensiblement les mêmes descripteurs. Ces regroupements mettront en évidence, ou non, des ensembles de journaux qui abordent la MSN pratiquement sous le même angle.

Dans notre cas, l'ACP explique à 64,4 % la répartition des journaux en fonction des sujets traités. Donc 35,6 % de cette répartition ne sont pas expliqués par cette analyse. En bibliométrie, à partir de 60 % d'interprétation, on peut avoir une bonne analyse.

### Résultats : interprétation de l'analyse

Dans l'encadré ci-contre, le pourcentage entre parenthèses, après le nom du journal, représente la part d'explication fournie par les commentaires qui suivent. Par exemple, *Sleep (New York)* (67 %), signifie que les commentaires qui suivent expliquent 67 % des aspects abordés par ce journal, concernant la MSN.

Les principaux thèmes traités sont les suivants :

- aspect épidémiologique, abordant la position pendant le sommeil, les différents facteurs de risque, la localisation géographique ;
- aspect respiration et pathologie ;
- aspect vigilance et réanimation ;
- aspect rythme cardiaque, avec souvent une approche statistique ou comparative des phénomènes ;
- aspect système nerveux, principalement orienté vers le système nerveux central.

L'aspect respiration est aussi lié à l'aspect vigilance, et l'aspect réanimation est lié aux aspects pathologie et étiologie.

D'autres aspects plus secondaires apparaissent : électroencéphalographie, prévention et étiologie.

Deux groupes principaux de journaux se détachent, abordant principalement soit l'aspect épidémiologie, soit l'aspect respiration et pathologie.

#### Liste des journaux traités

Le nombre entre parenthèses correspond à la fréquence du journal dans le téléchargement.

*Lancet (British Edition)* (20)  
*Pediatrics (Evanston)* (12)  
*Archives of Disease in Childhood* (10)  
*Journal of Pædiatrics and Child Health* (9)  
*Early Human Development* (7)  
*Klinische Pædiatrie* (6)  
*British Medical Journal* (6)  
*Neurochemistry International* (5)  
*European Journal of Pediatrics* (5)  
*Pédiatrie (Marseille)* (4)  
*Pediatric Research* (4)  
*Medical Hypotheses* (4)  
*Journal of Clinical Pathology* (4)  
*Forensic Science International* (4)  
*Biology of the Neonate* (4)  
*Australian Pædiatric Journal* (4)  
*Sleep (New York)* (3)  
*La presse médicale (1983)* (3)  
*Journal of Inherited Metabolic Disease* (3)  
*International Journal of Epidemiology* (3)  
*Archives françaises de pédiatrie* (3)

Nombre de journaux à fréquence 2 : 12  
 Nombre de journaux à fréquence 1 : 56  
 Nombre total de journaux : 89

Les journaux qui ne sont pas dans un des deux groupes sont isolés. Ils abordent la MSN sous un aspect original par rapport à l'ensemble. Il est aussi possible de distinguer deux autres groupes : d'une part, les journaux moyennement ou bien représentés dans l'analyse (à partir de 60 %) et, d'autre part, les journaux médiocrement représentés (en dessous de 50 %).

Voici la description de chacun des 21 journaux analysés, en commençant par ceux qui sont moyennement ou bien représentés.

• **Journaux abordant entre autres l'aspect épidémiologie :**

- *Lancet (British Edition)* (77,4 %) : traite de la vigilance, de la réanimation, de la position, de l'épidémiologie, et des facteurs de risque ;
- *Journal of Paediatrics and Child Health* (62,7 %) : traite principalement de l'épidémiologie, des facteurs de risque, de la position et de la localisation géographique. Traite secondairement de la réanimation et de la vigilance ;
- *British Medical Journal* (76,4 %) : traite de l'aspect épidémiologie, des facteurs de risque, de la position, de la localisation géographique, de la réanimation et de la vigilance ;
- *Archives françaises de pédiatrie* (59 %) : traite de l'aspect épidémiologie, des facteurs de risque, de la position, et de la localisation géographique. Ne semble pas aborder l'aspect respiration, pathologie.

• **Journaux abordant entre autres l'aspect respiration :**

- *Pediatrics (Evanston)* (60,3 %) : traite principalement de la respiration, et de la vigilance, et secondairement de réanimation et de pathologie ;
- *Archives of Disease in Childhood* (65,1 %) : traite de la respiration et de pathologie, mais ne semble pas traiter de l'aspect épidémiologie, position, facteurs de risque et localisation géographique. Semble s'en tenir aux articles qui relatent une approche expérimentale ou clinique ;
- *Early Human Development* (59,2 %) : traite principalement de la

respiration et de pathologie, secondairement aborde l'aspect rythme cardiaque, électrocardiologie, étude statistique ;

- *Pediatric Research* (70,9 %) : traite de la respiration et de pathologie, aborde peu l'aspect épidémiologie ;
- *La presse médicale (1983)* (68,8 %) : traite principalement de la respiration et de pathologie, secondairement de l'aspect système nerveux, aspect cérébral. Ne semble pas traiter de l'épidémiologie, de position, des facteurs de risque, de localisation géographique, ni d'étiologie ;
- *European Journal of Pediatrics* (52,4 %) : traite de l'aspect respiration et pathologie.

• **Autres journaux :**

- *Sleep (New York)* (67 %) : aborde principalement l'aspect électroencéphalogramme, vigilance et respiration, secondairement, l'aspect rythme cardiaque, étude statistique ;
- *Medical Hypothese* (66,9 %) : traite de l'aspect étiologie. Ne semble pas aborder l'aspect cardiaque, ni les aspects épidémiologie, respiration, position, facteurs de risque, localisation géographique ;
- *Biology of the Neonate* (53,8 %) : traite principalement de l'aspect cardiaque et étude statistique, secondairement de l'aspect pathologie et respiration ;
- *Neurochemistry International* (50,3 %) : traite du système nerveux, de l'aspect cérébral.

• **Les journaux suivants ne sont**

**pas affectés à l'une ou l'autre des catégories précédentes**, car trop peu « expliqués » par l'analyse (< 50 %) :

- *Klinische Paediatric*,
- *Pédiatrie (Marseille)*,
- *Journal of Clinical Pathology*,
- *Forensic Science International*,
- *Australian Paediatric Journal*,
- *Journal of Inherited Metabolic Disease*,
- *International Journal of Epidemiology*.

**Discussion et conclusion**

La difficulté majeure de cette étude a été de reformater et retraiter l'information téléchargée. Une bonne connaissance des logiciels utilisés est indispensable. Pour réaliser une analyse statistique correcte, il faut que le fichier soit informatiquement « propre ». Le zéro défaut est certes utopique, mais il faut faire bonne mesure entre le temps passé à la correction et la qualité demandée pour les résultats. En effet, il arrive un moment où toute correction est inutile si l'on compare le temps passé à la réaliser à la qualité d'interprétation gagnée. Avec de la pratique et des moyens adéquats, ce travail doit pouvoir être réalisé en une semaine, par une personne à plein temps, à condition d'être assuré de la collaboration d'un spécialiste du sujet traité pour réaliser les regroupements sur le champ des descripteurs.

R É F É R E N C E S

[1] The neuropathology of the sudden infant death syndrome : a review / K. Kinney, J. Filliano, R. Harper. — *Journal of neuropathology and experimental neurology*, 1992, vol. 51, p. 115-126.

[2] Pratique de la veille technologique / François Jakobiak. — Paris : Les éditions d'organisation, 1991. — 219 p.

[3] Veille technologique et bibliométrie : concepts, outils, applications / Hervé Rostaing. — 1993. — 353 p. Th. doct. : Aix-Marseille III : 1993.

[4] Systèmes de surveillance systématique pour le management stratégique de l'entreprise / William Nivol. — 1993. — 333 p. Th. doct. : Aix-Marseille III : 1993.

[5] Bibliométrie sur des références bibliographiques : méthodologie / Luc Quoniam. In : La veille technologique / Hélène Desval et Henri Dou. — Paris : Dunod, 1992. — P. 243-262.

[6] Exploration informatique et statistique des données / Michel Jambu. — Paris : Dunod, 1989. — 505 p.

L'analyse de ces journaux est valable dans la mesure où l'on se réfère uniquement aux 203 références. C'est une analyse décrivant le contenu du téléchargement : on observe donc, avec l'ACP, les convergences entre différents journaux, sur une période allant de 1990 à 1992, journaux dont les références bibliographiques sont contenues dans la banque *Pascal*. Cette méthode décrit les sujets traités par différents journaux, mais ne mesure pas la qualité de ces journaux.

Les journaux dont le contenu a été correctement analysé sont ceux dont la fréquence est d'au moins 9 ou 10 apparitions dans le téléchargement (encadré page 208).

Une meilleure analyse aurait été faite sur un nombre plus grand de références, toujours en utilisant les mots-clés de *Pascal*. Avec plus de références, serait apparu plus de bruit (journaux de fréquence 1), mais les

journaux importants auraient eu une fréquence plus forte, et donc l'analyse statistique aurait été meilleure. Pour obtenir un plus grand nombre de références en interrogeant *Pascal*, une solution consiste à étendre la recherche à un intervalle de temps plus long (par exemple de 1988 à 1992).

Dans le cadre du choix d'abonnement à des revues, cette méthode peut être utilisée comme aide à la décision pour chiffrer les sujets abordés. Le traitement devra cependant être réalisé sur un plus grand nombre de notices bibliographiques, et sur plusieurs banques de données, pour avoir une image la plus proche possible de la réalité. Les banques de données n'ayant pas toutes le même format, une étape supplémentaire s'impose qui est le reformatage personnalisé de chaque téléchargement pour avoir en fin de compte un format unique de fichier. Dans ce cadre, il conviendra de limiter la

période sur laquelle on interroge la banque, pour avoir un corpus le plus actuel possible. La répartition des sujets traités actuellement n'est certainement pas la même qu'il y a vingt ans. Le résultat de l'ACP permettra d'orienter la décision.

Ce genre de méthode pourrait être utilisé en bibliothéconomie pour aider à sélectionner les journaux auxquels s'abonner. Elle permettrait de ne pas se fier exclusivement à des spécialistes, en proposant une analyse comparative du contenu des revues, et ce sans lire aucun article, l'analyse étant entièrement automatique. Cette méthode élimine les choix fondés sur des « on dit que tel journal traite plus ce thème que tel autre », en chiffrant les différences ou les ressemblances au sein d'une sélection de journaux.

Janvier 1994

## LOGICIELS - LOGI+ - LOGICIELS

### Gérez votre bibliothèque et vos fonds comme vous l'entendez



**Convivialité :** approches professionnelles et grand public (OPAC) grâce à la souplesse des paramètres.

**Evolutivité :** plus de 10 ans de développement et de suivi à l'écoute du client.

**Ouverture :** échange des données en import et en export - adaptation aux nouvelles technologies.

**Rapport qualité/prix :** sans concurrence.



6, rue de Stockholm - 67000 Strasbourg  
Tél. 88 60 17 60 - Fax 88 61 12 72