

Etude bibliométrique de la visibilité internationale de la production scientifique de l'INSA.

DEWITTE Pierre-Emmanuel

Sous la direction de Jean Michel Jolion et Monique JOLY
Directeur Adjoint de la Direction de la Recherche
Directrice de DOC'INSA – INSA de Lyon



Remerciements

Je souhaite remercier toutes les personnes qui, de près ou de loin, m'ont aidé durant le stage.

Je suis reconnaissant plus particulièrement envers :

- Monique Joly, directrice de Doc'INSA, et Jean-Michel Jolion, directeur adjoint à la Recherche de l'INSA, qui m'ont encadré, guidé, appuyé et conseillé tout au long de cette étude.
- Thierry Lafouge, maître de conférences à l'ENSSIB, et Geneviève Lallich-Boidin, maître de conférence du laboratoire RECODOC, pour leurs précieux conseils et pour m'avoir accueilli pendant un mois dans leurs locaux.
- Gilles Brochet et Bernard Pariot, du service informatique de DOC'INSA, pour leurs assistances techniques.
- Guillaume Eyraud du Centre des Ressources Informatiques pour son aide, son soutien et ses conseils.

Fatima Nadji, responsable périodiques et prêt en bibliothèques à DOC'INSA, pour son aide dans mes recherches.

Résumé : Etude bibliométrique de la visibilité internationale de la production scientifique de l'INSA.

La bibliométrie permet d'analyser et d'apprécier la production scientifique d'un établissement et le rayonnement de sa recherche scientifique. En 2004, Le Pôle Universitaire Lyonnais a réalisé une première étude bibliométrique (Epsilon2004) destinée à évaluer, dans la base de données des *Current Contents*, la visibilité internationale des articles de périodiques scientifiques des établissements lyonnais. Cette étude a mis en évidence un manque de visibilité de ces publications. L'objectif de notre étude bibliométrique est de confronter les données de l'étude Epsilon2004 avec les données internes de l'INSA de Lyon pour confirmer les causes du manque de visibilité des publications et proposer d'éventuelles solutions.

Descripteurs : bibliométrie, études bibliométriques, publication scientifique, visibilité, base de données, article périodique scientifique, valorisation production scientifique.

Toute reproduction sans accord express de l'auteur à des fins autres que strictement personnelles est prohibée.

Abstract : Bibliometric study about international visibility of INSA scientific production.

Bibliometric studies are a way to analyse the scientific production of universities and their research diffusion. In 2004, the "*Pôle Universitaire Lyonnais*" launched a bibliométric study (Epsilon2004) to estimate the international visibility of scientific articles by the Universities of Lyon, in the Current Contents database. The results of this study provided evidence of a lack of visibility of those productions. The present study is a comparison between Epsilon2004 study data and *Insa de Lyon's* own data to confirm the reasons for the lack of scientific article visibility and to propose possible solutions

Keywords : bibliometrics, bibliometric studies, visibility, database, scientific publication, article, scientific production valorization.

Sommaire

INTRODUCTION.....	7
PRÉSENTATION DE L'INSA DE LYON	13
1. GÉNÉRALITÉS	13
2. LA RECHERCHE À L'INSA DE LYON.....	13
3. LA VALORISATION DE LA RECHERCHE : INSAVALOR	15
4. L'INFORMATION SCIENTIFIQUE ET TECHNIQUE : DOC'INSA	15
RAPPELS DES RÉSULTATS ET CONCLUSIONS DU RAPPORT	
EPSILON2004.....	17
1. OBJECTIFS DE L'ÉTUDE	17
2. MÉTHODOLOGIE UTILISÉE	18
3. RÉSULTATS.....	19
4. DISCUSSION ET CONCLUSION.....	20
5. REMARQUES SUR L'ÉTUDE D'EPSILON2004.....	20
SOURCES DES DONNÉES INTERNES DE L'INSA.....	23
1. LES RAPPORTS QUADRIENNAUX	23
2. LES BASES DE DONNÉES DE L'INSA : LABINFO, LABINSA.....	24
2.1. <i>Présentation des bases de données</i>	24
2.2. <i>Modifications apportées à la structure de la base LabINSA</i>	27
2.3. <i>Quelques mots sur le transfert des données</i>	27
MÉTHODOLOGIE.....	30
1. COMPARAISON QUANTITATIVE DES RAPPORTS QUADRIENNAUX, DE LA BASE LABINSA ET DU CORPUS D'EPSILON2004	30
2. COMPARAISON QUANTITATIVE ET QUALITATIVE DU CORPUS DE LA BASE LABINSA ET DU CORPUS D'EPSILON2004.....	30
2.1. <i>Répartition des articles entre la base LabINSA et Epsilon2004</i>	31
2.2. <i>Choix méthodologiques et réalisation technique</i>	32
2.2.1 <i>Clef de dédoublement</i>	32

2.2.2	Constitution d'une table des périodiques scientifiques de LabINSA	33
2.2.3	Les articles dont au moins un auteur est rattaché à l'INSA	33
2.2.4	Comparaison des corpus LabINSA et Epsilon2004	37
2.2.5	Analyse du corpus D	38
RÉSULTATS ET DISCUSSION		40
1.	TROIS VUES POSSIBLES DE LA PRODUCTION SCIENTIFIQUE DE L'INSA	40
1.1.	<i>Articles présentés par les rapports d'activité</i>	40
1.2.	<i>Articles "INSA" présents dans LabINSA</i>	42
1.3.	<i>Articles "INSA" par l'étude d'Epsilon2004</i>	43
1.4.	<i>Résultats Globaux et Conclusion</i>	43
2.	COMPARAISON DES CORPUS DE LABINSA ET D'EPSILON2004	44
2.1.	<i>La Clef de dédoublement</i>	44
2.2.	<i>Les périodiques de la base LabINSA</i>	45
2.3.	<i>Les articles "INSA" de la base LabINSA</i>	45
2.4.	<i>Les articles "INSA" présents dans le corpus d'Epsilon2004</i>	46
2.5.	<i>Les articles "INSA" de la base LabINSA absents du corpus Epsilon2004</i>	49
2.6.	<i>Analyse du Corpus D</i>	51
2.6.1	<i>Répartition des articles "INSA" selon les laboratoires et les pôles de compétences</i>	51
2.6.2	<i>Analyse de la revue des articles du corpus D</i>	53
3.	PROPOSITIONS DE SOLUTIONS AUX PROBLÈMES POSÉS PAR LA BASE LABINSA	57
4.	LES CAUSES DU MANQUE DE VISIBILITÉ INTERNATIONALE DE L'INSA DE LYON	59
CONCLUSION		62
BIBLIOGRAPHIE PAR ORDRE ALPHABÉTIQUE D'AUTEURS		64
BIBLIOGRAPHIE PAR RUBRIQUE		65
TABLE DES ANNEXES		67

Introduction

Les prémisses du concept de bibliométrie remonteraient au début du XIX siècle (ROSTAING 1996). Mais ce sont les travaux de Bradford qui ont vraiment donné son essor à cette discipline. Bradford considérait que la gestion de bibliothèque est soumise au "chaos documentaire" de la littérature. Il a donc cherché à modéliser la répartition des journaux selon leur aptitude à représenter un domaine scientifique, par des analyses statistiques des productions écrites. Par la suite, la plupart des travaux bibliométriques étaient consacrés à la recherche de formules mathématiques s'ajustant le mieux possibles aux données bibliographiques expérimentales.

Dans les années soixante, un américain De Solla Price développa l'idée que l'activité scientifique est régie par des lois sociologiques. Il renonça à l'emploi de l'outil statistique selon la rigueur mathématique exigée, et introduisit des facteurs sociologiques. Cela lui permit de formuler plusieurs lois encore reconnues aujourd'hui (ROSTAING 1996). Si les techniques statistiques utilisées étaient les mêmes que celles de la bibliométrie, les objectifs étaient différents. Aussi les chercheurs appelèrent ces techniques "scientométrie". Parallèlement, E. Garfield fonda au début des années 1960, l'*Institute of Scientific Information* (ou ISI). Garfield a contribué à accroître les connaissances statistiques sur les articles de périodiques scientifiques en créant en 1963 le *Science Citation Index* (ou SCI). En 1978, la scientométrie s'est même dotée d'une revue *Scientometrics* et le premier colloque international de scientométrie a eu lieu en 1987.

L'abondance des travaux et surtout la variété des objectifs ont rendu confuses les terminologies. Aussi donnons-nous ci-après la définition de quelques termes utilisés aujourd'hui :

Bibliométrie : ensemble des méthodes statistiques ou mathématiques appliquées à des références bibliographiques.

Scientométrie : application de méthodes statistiques à des données quantitatives (économiques, humaines, bibliographiques) caractéristiques de l'état de la science.

Infométrie : ensemble des activités métriques relatives à l'information, couvrant aussi bien la bibliométrie que la scientométrie.

Les outils d'analyse bibliométrique ont été initialement imaginés comme une aide à la compréhension de l'évolution de la communauté scientifique. Mais rapidement les instances dirigeantes ont vu en eux la possibilité de disposer d'un système d'évaluation de la recherche. Les indicateurs bibliométriques permettent d'évaluer l'activité de la recherche, c'est-à-dire la productivité et la position stratégique des différents acteurs participant à cette recherche. Ceci influe aujourd'hui sur les prises de décision dans les politiques de recherche.

En février 1995, sept établissements français habilités à délivrer des diplômes nationaux (Université Claude Bernard Lyon 1, Université Lumière Lyon 2, Université Jean moulin Lyon 3, Ecole Centrale de Lyon, Ecole Normale Supérieure de Lyon, Ecole Normale Supérieure Lettres et Sciences Humaines, Institut National des Sciences Appliquées de Lyon) se sont regroupés au sein du Pôle Universitaire de Lyon (ou PUL) afin de concerter leurs efforts dans différents secteurs et de donner une meilleure lisibilité des différentes activités des Etablissements d'Enseignement Supérieur (ou EES) : formation, recherche, valorisation de la recherche, culture scientifique et technique, etc. Ils ont été rejoints depuis par 9 autres établissements.

Le PUL est un Groupement d'Intérêt Public, régit par la loi 1901 relative aux associations. Il regroupe actuellement 16 établissements d'enseignement supérieur (voir la présentation du PUL en annexe 1). A travers le développement de projets communs, le PUL remplit ses objectifs qui sont :

- De favoriser l'ouverture, la politique d'accueil et le rayonnement international de l'Université de Lyon (sous cette désignation sont regroupés tous les établissements du PUL).
- De renforcer l'information et la qualité de vie pour les étudiants.
- De développer la communication et la culture scientifique.
- De valoriser et insérer la recherche dans le développement économique local.
- D'accompagner le développement des nouvelles technologies pour l'enseignement.

La nécessité de positionner la production scientifique des EES représente une part importante des objectifs du le PUL. Un positionnement de la production scientifique était nécessaire pour doter les EES d'indicateurs fiables et lisibles afin de fournir des données en réponse à les études menées au niveau international, comme par exemple le classement mondiale des universités de l'Université de Shanghai. (<http://ed.sjtu.edu.cn/ranking.htm>). La bibliométrie est susceptible de fournir de tels indicateurs.

Depuis 1999, et à plusieurs reprises ensuite, le PUL a obtenu un soutien financier de la région Rhône-Alpes pour conduire le projet Epsilon. Ce projet avait pour but de construire une méthodologie pour ensuite évaluer, à l'aide d'indicateurs bibliométriques pertinents, la production scientifique des établissements du PUL. Une première étude d'Isabelle Marchand (MARCHAND 1999) analysait, dans le secteur Biologie-santé, la production scientifique de deux établissements : l'Université Claude Bernard Lyon1 et l'Institut Nationale des Sciences Appliquées. Plusieurs sources de données avaient été testées (rapports d'activités des laboratoires, Internet, bases de données informatisées) et une équation de recherche avait été mise au point pour interroger les bases de données. Isabelle Marchand souligne le manque de normalisation des adresses indiquées par les auteurs dans leurs publications scientifiques. La principale conclusion de ce rapport était que la base de données des *Current Contents* de l'ISI, constituait le meilleur outil pour analyser la production scientifique des établissements.

La base de données des *Current Contents* couvre tous les domaines scientifiques, au niveau mondial. Elle contient les notices bibliographiques de plus de 8500 périodiques dépouillés. La collecte des publications à travers une équation de recherche portant sur les adresses s'avère fiable et relativement exhaustive : 70% des publications sont retrouvées en croisant avec les rapports d'activité des laboratoires (BEDAGUE 2001). La particularité du champs "institution" des notices bibliographiques de cette base, est que l'adresse de tous les auteurs est indiquée. En revanche ce champs n'est pas normalisé, ce qui conduit à insérer dans l'équation de recherche tous les termes se rapportant aux EES. Le principal indicateur de la production scientifique de cette base est l'article de périodique scientifique. Les brevets, ouvrages et congrès sont en grande majorité ignorés. Par

conséquent la visibilité de la production scientifique est analysée au travers des articles.

En 2001 Isabelle Bedague (BEDAGUE 2001) a étudié les différentes étapes à mettre en place pour identifier la production du PUL. Elle s'est beaucoup intéressée au problème de la non normalisation des adresses des productions scientifiques. Elle a travaillé sur une table des correspondances des adresses d'établissements lyonnais importée du Ministère de la Recherche. En 2002, Adel Larouci (LAROUCI 2002) a enrichi l'équation de recherche établie en 1999 par Isabelle Marchand et réalisé l'identification de la production scientifique de 5 établissements du PUL (Université Lyon1, INSA de Lyon, Ecole Centrale de Lyon, Ecole Normale Supérieure de Lyon), dans le domaine de la physique.

S'appuyant sur les travaux précédents, Camille Prime-Claverie a mené, début 2004, une étude portant sur 5 établissements du PUL : l'Université Claude Bernard Lyon1, l'Institut National des Sciences Appliquées, l'Ecole Centrale de Lyon, l'Ecole Normale Supérieure-Sciences, l'Ecole Nationale Vétérinaire de Lyon. L'objectif n'était pas d'analyser la production scientifique en la collectant à la source (dans les laboratoires des établissements) mais de la capter dans les sources nationales et internationales de nature bibliographique. Puis des indicateurs bibliométriques ont été mis au point pour rendre compte de cette production et permettre des comparaisons avec d'autres pôles de recherche en France. Cette étude ouvre une discussion sur la façon dont les établissements peuvent améliorer leur visibilité internationale. Nous avons détaillé précisément cette étude dans la suite de notre rapport car elle est à la base de notre travail.

La demande qui nous a été soumise par Monique Joly, Directrice de DOC'INSA et Jean Michel Jolion Directeur Adjoint à la Direction de la Recherche de l'INSA de Lyon, consistait à comparer les résultats de l'étude Epsilon2004, aux données internes de l'INSA. La Direction de la Recherche de l'INSA possède depuis 1995 une base de données dans laquelle sont renseignées (en théorie) toutes les publications scientifiques des laboratoires de l'INSA.

L'objectif de notre travail était donc de:

- Comparer ces deux corpus de publications, en nous limitant aux articles de périodiques scientifiques de l'INSA. Ce sont les articles où au moins un auteur fait partie de l'INSA.
- Déterminer les causes du manque de visibilité de l'INSA dans les Current Contents.
- Proposer des pistes de solutions pour y remédier.

Deux hypothèses étaient avancées alors, pour expliquer ce manque de visibilité :

- Le manque de normalisation de l'adressage des publications par les chercheurs. Cela se traduit par une grande diversité d'adresses pour désigner un même établissement.
- Le décalage entre les domaines d'activités des laboratoires de l'INSA et les domaines scientifiques qualifiés de "fondamentaux" par l'ISI. Certaines sciences, telles que la Physique ou la Santé, sont présentées par l'ISI comme le "cœur" de la science mondiale. Cela se répercute dans le choix des revues dépouillées.

Nous tenterons de vérifier ces deux hypothèses tout au long de notre étude.

La base LabINSA de la Direction de la Recherche est mise à jour à partir d'extractions des données des bases propres, appelées bases LabINFO, de chaque laboratoire de l'INSA. Or la base LABINSA n'ayant pas été pensée pour réaliser des études bibliométriques, notre premier travail s'est attaché à faire des propositions et à mettre en œuvre des modifications de la structure de la base pour pouvoir disposer des données bibliographiques nécessaires. Puis nous avons activement participé à la mise à jour des données par les laboratoires. Enfin nous avons comparé le corpus des articles de périodiques scientifiques de l'INSA en interne avec le corpus de l'étude Epsilon2004.

Nous décrivons l'ensemble de l'étude menée pendant notre stage, selon cinq parties. Tout d'abord nous présentons l'INSA de Lyon en tant que grande école d'ingénieurs française. Dans une deuxième partie, nous rappelons les résultats de l'étude menée par Camille Prime-Claverie. Cette étude étant notre point de comparaison, nous l'avons analysée de façon très précise. Troisièmement nous exposons nos sources d'informations, notamment les bases de données LabINFO

des laboratoires et la base LabINSA. Nous proposons des modifications de structures de cette base afin de mener à bien notre travail. Quatrièmement, nous détaillons notre méthodologie ainsi que les choix que nous avons dus faire en fonction des données qui nous ont été fournies. Dans une cinquième partie, nous présentons nos résultats et nous les analysons. Nous les confrontons également avec ceux des études menées précédemment dans le cadre de l'analyse de la production scientifique des établissements du PUL. Pour finir, nous concluons sur les causes du manque de visibilité de l'INSA de Lyon et sur les perspectives d'avenir de la bibliométrie. Tout au long de notre rapport nous indiquons également les difficultés que nous avons rencontrées et les solutions qui nous semblent adaptées.

Présentation de l'INSA de Lyon

1. Généralités

Fondé en 1957, l'Institut National des Sciences Appliquées de Lyon (ou INSA de Lyon, <http://www.insa-lyon.fr/>) appartient au réseau national des INSA : Lyon, Rennes, Rouen, Strasbourg, Toulouse. L'INSA de Lyon fait partie des Grandes Ecoles d'Ingénieurs françaises et représente l'un des meilleurs établissements d'Enseignement Supérieur en Europe. Avec un cycle de formation sur 5 ans, plus de 800 ingénieurs sont diplômés chaque année dans 10 spécialités. De plus l'INSA de Lyon est tourné vers l'International avec plus de 25% d'étudiants étrangers accueillis par an.

2. La Recherche à l'INSA de Lyon

La Recherche est un domaine essentiel de l'INSA de Lyon. Elle est caractérisée par son aspect pluridisciplinaire avec une grande importance accordée à la Recherche Technologique et aux Sciences pour l'Ingénieur. Les 27 laboratoires que compte l'INSA de Lyon sont répartis selon 5 pôles de compétences :

- Le pôle Biologie et Santé. Ce pôle est "original" dans l'Institut. La mise en place de l'opération transversale bio-informatique en est une expression concrète. Elle s'appuie sur une formation ingénieur en biochimie assez peu courante dans les écoles d'ingénieurs.
- Le Pôle Matériaux. L'activité dans le domaine des matériaux de structure, des matériaux de fonction et des matériaux polymères constitue un axe fort, un des piliers de la recherche à l'INSA de Lyon. Trois fédérations CNRS (Matériaux de structure, micro-nano technologies et polyméristes lyonnais) en sont la traduction.
- Le pôle Energie-Environnement. L'environnement pour le développement durable et la maîtrise de l'énergie constituent un domaine prioritaire fédérateur

de recherche tant au sein de l'Institut que sur le site lyonnais. La réflexion en cours pour la mise en place prochainement d'un pôle global de recherche et d'expertise en environnement pour le développement durable traduit une volonté politique forte dans le domaine.

- Le pôle Mécanique. La mécanique constitue une tradition forte de l'INSA de Lyon. Elle a su se structurer et faire émerger, en synergie avec le pôle matériaux, des thématiques fortes et reconnues. La mécanique du solide et des structures, la tribologie, la dynamique des machines et l'acoustique en sont l'expression visible.
- Le pôle Sciences et Technologies de l'Information et de la Communication. Le fort développement des STIC (Sciences des Technologies de l'Information et la Communication) en formation est accompagné par celui de la recherche. L'émergence d'une thématique STIC-Santé visible dans les laboratoires de ce domaine apparaît clairement. La thématique STIC représente actuellement 40% du potentiel recherche de l'institut.

La création de ces cinq pôles de compétences est accompagnée par la mise en place d'opérations transversales fédératives visant à renforcer et faire vivre la multidisciplinarité, la transversalité et l'interdisciplinarité.

L'INSA de Lyon développe également un partenariat étroit avec les autres établissements lyonnais (principalement Ecole Centrale de Lyon, Ecole Normale Supérieure-Sciences et Université Claude Bernard Lyon1) au travers de laboratoires inter-établissements. Enfin 14 laboratoires sont contractualisés avec des Etablissements Public à caractère Scientifique et Technique (ou EPST) : 12 sont affiliés CNRS, 1 INSERM et 1 INRA. Le descriptif des 27 laboratoires concernant leurs statuts, leurs effectifs etc. est présenté en annexe 2.

Tout ceci concourt à faire de l'INSA de Lyon, un important complexe de Recherche en France.

La Recherche à l'INSA de Lyon en quelques chiffres :

- 511 Enseignants-Chercheurs et Chercheurs (338 dans des laboratoires EPST).
- 240 étudiants en DEA dont 94 élèves ingénieurs INSA.
- 450 doctorants.
- 120 thèses soutenues par an.

- 1500 publications et communications internationales.
- Deuxième école en recherche et développement avec un chiffre d'affaires de 14,7 millions d'euros (Industrie et Technologies – novembre 2003).

3. La Valorisation de la Recherche : INSAVALOR

Dans le but de rapprocher le monde de la recherche et le monde de l'entreprise, l'INSA de Lyon a créé, en 1988, une filiale : INSAVALOR (<http://www.insavalor.fr/p/propos.php>). Cette société favorise le transfert et la mise en valeur des connaissances et des compétences des laboratoires vers les entreprises. Ainsi celles-ci peuvent s'engager dans des travaux de recherche en partenariat avec l'INSA de Lyon. Avec 1000 contrats de partenariat et un chiffre d'affaire de 10 millions d'euros en 2002/2003, INSAVALOR est aujourd'hui parmi les premières filiales de valorisation de la recherche universitaire en France.

4. L'Information Scientifique et Technique : DOC'INSA

Dès sa création, l'INSA de Lyon a valorisé l'information à travers le développement de deux centres documentaires : DOC'INSA et la bibliothèque des Humanités. Au fil du temps, des bibliothèques propres aux laboratoires se sont développées. Ce sont les bibliothèques de composantes. Deux sont ouvertes au public : la bibliothèque du département Informatique et la bibliothèque du département Génie Industriel.

Actuellement DOC'INSA est le plus grand centre de ressources en Sciences pour l'Ingénieur de la Région Rhône-Alpes. DOC'INSA propose de nombreux services aux usagers. Un portail documentaire permet d'accéder à l'ensemble des ressources en Sciences de l'Ingénieur et en Humanités. Le renseignement documentaire est un service spécialisé pour conseiller et/ou informer sur les recherches documentaires, sur place ou à distance. Le catalogue collectif permet de rechercher un document dans les bibliothèques de l'INSA. Il a été construit, en réseau, à partir du catalogue universitaire national (SUDOC).

Le prêt entre bibliothèques est également proposé. Enfin un ensemble de formations à la recherche documentaire est dispensé tout au long du cursus des ingénieurs, en fonction de leur spécialisation.

DOC'INSA en quelques chiffres :

- 100 000 ouvrages.
- 350 abonnements à des périodiques papiers.
- 2700 périodiques électroniques.
- Accroissement annuel de 4000 ouvrages.

Rappels des résultats et conclusions du rapport Epsilon2004

1. Objectifs de l'étude

L'objectif de l'étude d'Epsilon2004, menée par Camille Prime-Claverie pour le Groupement d'intérêt Public Pôle Universitaire Lyonnais (ou PUL), était d'évaluer la visibilité internationale de la production scientifique des établissements du PUL, soit l'Université Claude Bernard Lyon1, l'INSA, l'Ecole Centrale de Lyon, l'Ecole Normale Supérieure-Sciences, l'Ecole Nationale Vétérinaire de Lyon. Cette évaluation est importante non seulement pour connaître l'activité scientifique des établissements et ainsi aider à la décision politique et budgétaire ; mais aussi et surtout pour connaître le rayonnement mondial des établissements lyonnais et la façon dont ils sont perçus par la communauté scientifique internationale.

Le principe consiste à évaluer cette visibilité en analysant les bases de données scientifiques dans lesquels ils sont signalés. La base de données retenue est les *Current Contents* de l'*Institute of Scientific Information* (ou ISI), base bibliographique riche avec mention de l'affiliation de chaque auteur (voir un exemple de notice bibliographique en annexe 3). Le principe de comptage adopté est celui utilisé par l'Observatoire des Sciences et des Technologies (ou OST, présentation en annexe 4), à savoir le compte de présence : chaque fois qu'un établissement étudié apparaît dans le champs "institutions", la publication est comptabilisée. En théorie, plus le nombre de publications d'un établissement, trouvées dans la base de l'ISI, est important par rapport à la production totale, plus l'établissement a une bonne dimension internationale.

Le principal souci de Camille Prime-Claverie a été de réduire le bruit et le silence inhérents à toute recherche d'information. Dans ce cas la principale cause du bruit était le manque de précision lors de l'interrogation sur l'affiliation des auteurs. En

effet le champs "Institution", qui contient l'affiliation des auteurs, est multi-valué dans les *Current Contents*. La cause principale du silence était le manque de normalisation des adresses de la part des auteurs. Il en sera donné des exemples plus loin. Tout ceci constitue des biais, liés à la base des *Current Contents*, qui influent sur la qualité des résultats.

2. Méthodologie utilisée

Camille Prime-Claverie a mené ses interrogations avec deux équations de recherche. La première (équation 1, en annexe 5) est issue d'études antérieures (MARCHAND 1999, LAROUCI 2002) et utilise les différentes dénominations des établissements. Cette équation permet une interrogation assez précise mais ne résout pas le problème lié à la grande diversité d'adresses utilisées par les auteurs dans leurs publications. La deuxième équation (équation 2, en annexe 5) est basée sur la position géographique : c'est à dire le nom des villes où sont situés les établissements. Cette équation si elle est plus large, a l'avantage de passer outre la non-normalisation des adresses ; étant considéré que l'orthographe des villes est la même d'un auteur à l'autre.

Ensuite le champs "Institution", de chaque référence bibliographique rapatriée par les équations de recherche, a été scindé pour séparer les différentes affiliations et deux filtres "informatiques" ont été mis au point pour affecter les références à chaque établissement. Le premier (filtre 1, en annexe 5) s'appuie sur des "expressions régulières" englobant les différentes appellations des établissements. Le deuxième filtre (filtre 2, en annexe 5) vérifie si le champs "Institution" contient une affiliation à une unité du CNRS (grâce à la mention du numéro d'UMR) ; et le cas échéant l'affecte à son ou ses organismes(s) de tutelle si celui-ci (ceux-ci) est un établissement étudié. Le filtre 2 s'avère nécessaire par le fait que des établissements du PUL comme l'INSA possèdent des laboratoires affiliés au CNRS ayant plusieurs organismes de tutelle (Université, Ecole Normale, INSA, etc.) et dans lesquels les chercheurs ne mentionnent pas toujours leur établissement d'appartenance.

3. Résultats

Tout d'abord nous devons préciser que nous présentons ci-après uniquement les résultats qui nous intéressent pour notre propre étude. Ce sont des résultats globaux pour chaque établissement. Camille Prime-Claverie présente par ailleurs, des résultats selon les champs disciplinaires ou "Classification Code" de l'ISI ; selon les sujets de journaux ou "Journal Subject" de l'ISI ; ainsi que les collaborations internationales des établissements.

Le principe de comptage adopté est le compte de présence. Chaque fois qu'un des établissements étudiés apparaît dans le champs "Institution" d'une référence, il reçoit la note "+1". Si plusieurs laboratoires du même établissement apparaissent, la note reste "+1" (l'article n'est compté qu'une fois pour cet établissement).

Les résultats sont donnés pour la période 2000/2003. L'équation 2, avec 14 966 références, apporte plus de réponses que l'équation 1 (12 888 références) ; et tend ainsi à atténuer le silence. Cette atténuation est constatée également avec le double filtrage. L'application du filtre 1 sur les résultats de l'équation 2 aboutit à 10151 références, alors que l'application des filtres 1 et 2 donne 10585 références. Le deuxième filtre sélectionne les publications des Unités CNRS ; ce que ne permet pas le premier filtre. La réduction du silence était l'un des principaux buts de l'étude.

L'utilisation des deux filtres permet également la diminution du bruit. Le double filtrage diminue le nombre de références de 14 966 à 10 585 références.

Le tableau 1 présente les résultats pour l'ensemble du PUL.

Tableau 1 : Etablissements du PUL pour la période 2000/2003.

		2000 à 2003	2000	2001	2002	2003
EQUATION 1	filtre 1	9797	2538	2458	2501	2300
EQUATION 1	filtres 1 et 2	10004	2590	2511	2549	2354
EQUATION 2	filtre 1	10151	2581	2506	2541	2523
EQUATION 2	filtres 1 et 2	10585	2686	2605	2651	2643

Le tableau 2 présente les résultats pour l'INSA.

Tableau 2 : Institut National des Sciences Appliquées.

		2000 à 2003	2000	2001	2002	2003
EQUATION 1	filtre 1	1091	252	262	294	283
EQUATION 1	filtres 1 et 2	1150	263	278	306	303
EQUATION 2	filtre 1	1109	252	262	294	301
EQUATION 2	Filtres 1 et 2	1195	269	283	313	330

4. Discussion et Conclusion

Camille Prime-Claverie souligne que l'étude qu'elle a menée nécessite une très bonne connaissance du terrain. En effet, la réalisation des équations de recherche et des filtres nécessite la connaissance du nom de toutes les villes où sont situés les établissements qui sont parfois éclatés en différents lieux géographiques. De plus il faut déterminer toutes les appellations utilisées pour désigner ces mêmes établissements. Ainsi dans le cadre d'une évaluation globale sur le plan national et international, une telle étude, avec utilisation de filtres, ne peut être envisagée. Camille Prime-Claverie insiste sur l'obligation de normaliser les adresses.

Camille Prime-Claverie fait également remarquer les biais introduits par la base des *Current Contents* de l'ISI. Ces biais que nous évoquerons plus loin portent entre autre sur le choix des revues dépouillées, l'utilisation de l'article dit "scientifique" comme seul indicateur de production scientifique.

5. Remarques sur l'étude d'Epsilon2004

Notre présent travail s'attachera à analyser les résultats de l'étude Epsilon2004 pour l'INSA de Lyon. Or l'étude Epsilon2004 a été réalisée pour l'ensemble des établissements du PUL et cela pose des difficultés méthodologiques et introduit des biais. Nous précisons qu'il ne s'agit aucunement de remettre en cause le travail de Camille Prime-Claverie, parfaitement cohérent en lui-même ; mais les résultats de ce travail étant notre point de comparaison, nous nous devons d'évoquer les biais possibles engendrés par cette étude.

Trois biais nous apparaissent. Le premier vient de l'utilisation de la position géographique des établissements. Certains chercheurs peuvent ne pas indiquer le nom de la ville où se situe leur laboratoire. En particulier pour les laboratoires contractualisés EPST, les chercheurs peuvent ne signer que du nom du laboratoire et du numéro d'UMR. Le corpus des publications scientifiques de l'INSA peut avoir été sous-évalué.

Le deuxième provient de la note "+1" affectée à un établissement par sa présence dans le champs "Institution" des références bibliographiques. Pour bien comprendre le problème, il faut avoir conscience de la spécificité de la recherche française où les établissements EPST sont distincts des universités et des écoles. Nous avons dit, lors de notre présentation de l'INSA de Lyon, que les partenariats entre établissements étaient multiples. Or, le manque d'information concernant l'affiliation des chercheurs des EPST est important et pose des difficultés supplémentaires à notre étude. En effet, l'analyse du corpus des publications des laboratoires inter-établissements, nécessite une très bonne connaissance du personnel.

Prenons un exemple : le laboratoire CREATIS. Ce laboratoire est contractualisé CNRS, mais possède également du personnel INSERM et est rattaché à l'INSA et à l'Université Claude Bernard Lyon1. Pour notre étude nous ne nous préoccupons que des articles, parus dans des périodiques scientifiques, pour lesquels au moins un auteur est rattaché à l'INSA. Or certains articles peuvent n'être publiés que par des auteurs rattachés à l'Université Lyon 1. Par conséquent, pour ce type de laboratoire, l'étude du corpus des publications passe par une analyse des auteurs. Celle-ci nécessite de connaître précisément le personnel rattaché à l'INSA.

Pour le personnel EPST, la Direction de la Recherche est en mesure de nous donner la liste des personnels EPST rattachés à l'INSA. Pour distinguer les personnels des différents établissements au sein d'un même laboratoire, nous verrons dans la partie 3 que chaque laboratoire possède une base de données dans laquelle est renseigné le personnel. Chaque personne est caractérisée par une série de codes qui permettent de connaître son rattachement.

Si nous revenons maintenant à la note "+1" attribuée à chaque établissement présent dans le champs "institution" des références bibliographiques. Pour les

laboratoires inter-établissements, les publications signées uniquement du nom du laboratoire donnent la note "+1" à tous les établissements de tutelle alors que tous les auteurs peuvent n'appartenir qu'à un seul de ces établissements ; comme vu dans l'exemple du laboratoire CREATIS ci-dessus. Par conséquent, le corpus des publications scientifiques de l'INSA peut avoir été surévalué dans le rapport d'Epsilon2004.

Enfin, Camille Prime-Claverie a construit son corpus de départ avec l'équation 2 et a traité ce corpus de notices bibliographiques à travers deux filtres informatiques. Par conséquent, les filtres 1 et 2 peuvent ne pas avoir sélectionné toutes les productions scientifiques de l'INSA, initialement présentes dans le corpus obtenu avec l'équation 2.

Sources des données internes de l'INSA

Le principe de notre étude est la comparaison des données interne à l'INSA avec les données externes obtenues par Camille Prime-Claverie lors de son interrogation des *Current Contents*. Nous présenterons tout d'abord les sources des données internes de l'INSA à savoir les rapports quadriennaux et les bases de données LabINFO et LabINSA.

1. Les Rapports Quadriennaux

Les rapports quadriennaux sont rédigés tous les 4 ans par les laboratoires, à la demande du ministère de la recherche, pour justifier de leurs demandes de reconnaissance (label et financement). Ces rapports contiennent toutes les informations concernant le fonctionnement du laboratoire : son personnel, ses axes de recherche, ses activités, ses contrats et bien sûr l'ensemble des productions scientifiques (articles, brevets, actes de congrès etc.). Ces rapports d'activités sont disponibles à la Direction de la Recherche de l'INSA. Ils sont une source importante d'informations car ils permettent de connaître, de façon fiable, et quasi exhaustive, toutes les productions scientifiques d'un laboratoire sur une période donnée.

En revanche ils présentent deux inconvénients majeurs qui les rendent peu exploitables pour notre travail :

- Il n'existe pas de normalisation quant à la forme et au rendu de ces rapports : supports papiers et/ou électronique ; mise en page variable.
- Les publications de l'année en cours n'apparaissent pas toutes dans le rapport : environ un an de publication est perdu tous les 4 ans (LAROUCI 2002).

Néanmoins nous utiliserons ces rapports pour avoir un aperçu quantitatif du nombre de productions des laboratoires sur une période donnée.

2. Les Bases de Données de l'INSA : LabINFO, LabINSA

2.1. Présentation des bases de données

Chacun des 27 laboratoires de l'INSA possède sa propre base de données interne appelée LabINFO, dans laquelle sont saisies toutes les données concernant "la vie" du laboratoire.

On peut trouver la table "PRODUCTION" qui contient toutes les données bibliographiques de l'ensemble des productions scientifiques du laboratoire, à savoir les articles de périodiques mais aussi les brevets, les thèses, les actes de congrès, les livres, les rapports internes, les séminaires etc. Concernant les laboratoires inter-établissements, le fait que cette table contienne toutes les productions du laboratoire signifie que certaines productions peuvent n'avoir aucun auteur INSA.

Reprenons l'exemple du laboratoire CREATIS déjà cité (voir page 21). Ce laboratoire est sous tutelle de l'INSA et de L'Université Lyon1. Toutes les productions de ce laboratoire sont saisies dans la table PRODUCTION. Certaines d'entre elles peuvent n'avoir que des auteurs rattachés à l'Université Lyon1. Or nous ne nous intéressons qu'aux articles de périodiques scientifiques ayant au moins un auteur rattaché à l'INSA. Une recherche sur les auteurs a donc été nécessaire pour déterminer leur établissement de rattachement. Les auteurs sont saisis dans la table PERSONNEL.

La table "PERSONNEL" renseigne sur l'ensemble du personnel qu'il soit permanent ou temporaire. Les noms, prénoms, adresses, téléphones, période de présence, grades, fonctions etc. sont indiqués.

D'autres tables informent sur les implantations du laboratoire, les partenariats avec d'autres établissements, les projets, les types d'activités etc.

Toutes ces tables d'autorité sont liées entre elles soit par l'importation de clefs étrangères (relations (0,N) ou (1,N)) ; soit par la création de tables de relation (N,M). Le Modèle Conceptuel de Données des bases LabINFO, fourni par le Centre des Ressources Informatiques (ou CRI), est en annexe 6.

Une remarque cependant : beaucoup de champs sont facultatifs pour des raisons de bon fonctionnement de la base LabINFO. Mais l'inconvénient est que ces champs

sont souvent non renseignés. Prenons l'exemple du code établissement. Dans la table "PERSONNEL" sont renseignés de façon obligatoire, les nom, prénoms, numéro de dossier (automatique), le code qualité (chercheurs, doctorants, extérieur etc.). Mais le code de l'établissement d'appartenance de la personne est facultatif. Ceci est compréhensible par le fait que la personne qui saisit le dossier, ne connaît pas forcément l'établissement d'appartenance. Mais en conséquence, pour un chercheur donné, si ce code n'est pas renseigné, on ne peut savoir à quel établissement il appartient. Cela peut poser des problèmes, comme nous le verrons plus loin, pour l'exploitation des données lorsque nous cherchons à savoir à quel établissement appartient un auteur.

La même remarque peut être faite pour la table "PERIODIQUE" où le code ISSN n'est pas obligatoire pour la même raison la même raison et aussi parce que les chercheurs ne connaissent le plus souvent pas ce code et ne le fournissent donc pas aux personnes chargées de saisir les données dans la base (le plus souvent le secrétariat du laboratoire).

Concernant l'ensemble des LabINFO, certaines tables sont propres à chaque laboratoire comme la table "PERSONNEL", la table "PRODUCTION", la table "PERIODIQUE" etc. A l'inverse certaines tables dites tables de références, sont communes à tous les laboratoires pour apporter une certaine homogénéité dans la saisie. Exemple la table "QUALITE" présente les différentes "qualités" possibles du personnel, à savoir chercheur, stagiaire, doctorants etc. Ces tables de références sont contenues dans une base de données spécifique appelée base des références.

Pour notre propre étude, nous n'avons pas directement accès aux différentes bases de données LabINFO. En revanche nous avons accès à la base de données de la Direction de la Recherche appelée LabINSA. De façon simpliste, cette base de données regroupe une partie des données des bases LabINFO. Elle est alimentée à partir d'extractions des différents LabINFO, qui ont lieu deux fois par an en février et en juillet. Pendant cette extraction, l'intégrité de structure est conservée des bases LabINFO vers la base LabINSA. Pour les tables propres à chaque laboratoire, la table qui résulte dans LabINSA est la concaténation de toutes ces tables. Pour les tables de références une image de la base des références est disponible dans la base LabINSA.

Là encore deux remarques nous paraissent importantes. Tout d'abord le fait que certaines tables soient propres aux laboratoires et non communes posent des difficultés pour mener des études bibliométriques. Nous pensons tout particulièrement à la table "PERIODIQUE". La clef primaire de cette table est formée du code affecté au laboratoire et du code affecté au périodique. Si l'on veut faire une étude globale des revues dans lesquelles publient les laboratoires de l'INSA (ce qui est notre cas) la table "PERIODIQUE" de la base LabINSA résultante de la concaténation contiendra des doublons à cause du code du laboratoire. De plus seuls les libellés sont le plus souvent saisis et non les ISSN, d'où un risque de fautes d'orthographe important qui interdit toute comparaison avec une autre table contenant les périodiques d'un éditeur par exemple. La liste des tables de la base LabINSA est en annexe 7.

La deuxième remarque est **la possibilité de doublons dans LabINSA** dans la table "PRODUCTION". Nous rappelons que cette table contient toutes les productions scientifiques des laboratoires de l'INSA. La clef primaire de cette table est formée du code du laboratoire et d'un code production. Mais au moment de la concaténation dans la base LabINSA cette clef n'est plus valable. En effet, si deux laboratoires ont collaboré à la même production, ils saisissent chacun dans leur table "PRODUCTION", cette production avec leur code laboratoire et un code production. Lors de la concaténation une même production avec un code laboratoire différent donc deux clefs primaires différentes est intégrée dans la base LabINSA.

En conséquence, une table unique des périodiques pour tous les laboratoires faciliterait grandement le travail des études bibliométriques futures et augmenterait la clarté de lecture de la table "PERIODIQUE" de la base LabINSA. Enfin, une clef de dédoublonnage devrait être mise au point pour éviter les doublons dans la table "PRODUCTION".

Pour conduire notre étude, nous avons noté que l'ensemble des informations contenues dans les LabINFO n'étaient pas intégralement transférées dans LabINSA, notamment les données bibliographiques des productions. Ainsi avant notre étude, la table "PRODUCTION" de la base LabINSA contenait très peu, pour ne pas dire aucune, information bibliographique sur les productions des

laboratoires. La première partie de notre travail a donc consisté à proposer au Centre des Ressources Informatiques un ensemble de modifications à apporter à la base LabINSA afin d'obtenir toutes les informations nécessaires à notre travail.

2.2. Modifications apportées à la structure de la base LabINSA

Pour apporter des propositions de modifications, nous avons étudié de façon approfondie la structure des bases LabINFO et les données qui étaient transférées dans LabINSA. Ceci a été fait avec l'aide de Guillaume Eyraud du CRI. Les modifications proposées devaient, notamment, tenir compte de toutes les données bibliographiques des productions. Nos propositions ont été validées par Monique Joly. Nous avons ensuite transmis les modifications souhaitées à Guillaume Eyraud du CRI qui a réalisé les mises à jour techniques de la base LabINSA.

Pour les besoins de notre travail, nous avons retenu les tables suivantes :

- I_Production
- IC_Type_Production
- *I_Production_Auteur*
- I_Personnel
- I_Periodique
- IC_Qualité
- I_Laboratoire
- IC_Etablissement
- *I_Laboratoire_Etablissement*
- IC_Type_Etablissement

Les tables de relation sont indiquées en *Italique*.

Les tables dont le nom commence par I sont **Internes** au laboratoire les tables dont le nom commence par IC sont **Communes** à tous les laboratoires.

Les champs retenus et importés dans la base LabINSA, sont présentés en annexe 8.

Les tables et champs rajoutés suite aux modifications demandées pour notre travail sont indiqués en *Italique*.

Remarque : d'autres modifications apparaissent dans l'annexe. Elles ont été réalisées pour la Direction de la Recherche et pour DOC'INSA.

2.3. Quelques mots sur le transfert des données

Le transfert des données des bases LabINFO à la base LabINSA reste à l'initiative des laboratoires. Ceux-ci sont sollicités en février et en juillet mais la procédure de transfert est lancée par les laboratoires eux-mêmes, si nécessaire avec l'assistance du CRI. A ce sujet nous avons pu mesurer toutes les difficultés d'un tel transfert. L'INSA ne possède "que" 27 laboratoires et nous avons pu compter sur l'appui de

Jean Michel Jolion, Directeur Adjoint de la Recherche, dans son intervention devant les directeurs de laboratoires, lors de l'assemblée générale de l'INSA. Nous avons également pu bénéficier de l'appui de Guillaume Eyraud qui a assisté les laboratoires pour faciliter le transfert des données. Les modifications techniques de la base LabINSA ont été effectives le 23 juin et signalées aux laboratoires par mail par Guillaume Eyraud. Nous avons nous-mêmes suivi de très près cette période de transfert et relancé quasiment quotidiennement les laboratoires. Malgré cela, les dernières données sont arrivées le 23 juillet soit un mois après les modifications techniques de structures. De plus 7 laboratoires n'ont transféré aucune donnée. Attention : cela ne signifie pas qu'aucune donnée n'existe dans la base LabINSA pour ces laboratoires, mais qu'ils n'ont pas fait le transfert des données prévu en juillet 2004.

Plusieurs choses peuvent expliquer cela. Tout d'abord la contrainte de temps. C'est à ce problème que nous avons été le plus souvent confronté. La majorité des laboratoires fermant le 23 juillet pour les vacances d'été, les personnes chargées du transfert des données (très souvent les secrétaires) ont eu peu de temps à accorder à la mise à jour de LabINSA. Ensuite des contraintes techniques sont également apparues. Certains laboratoires ne possèdent aucun personnel qualifié pour effectuer la procédure de transfert (si simple soit-elle) ; où la personne qui en est chargée, rechigne à le faire car elle ne l'a jamais ou peu pratiquée. Le plus souvent, la contrainte du temps était évoquée pour contourner le problème. Enfin, il nous est apparu de façon plus subtile, une cause plus profonde. Il semble que les laboratoires aient peu conscience de l'intérêt de la mise à jour des bases LabINFO et encore moins de l'intérêt de la remontée des informations dans la base LabINSA. Lors de l'assemblée générale de l'INSA, les directeurs de laboratoires ont paru extrêmement septiques quant à l'utilité de saisir leurs informations dans LabINFO et de les transférer dans la base LabINSA. Quant aux personnels administratifs (secrétaires), il semble que ce soit une contrainte : le transfert est fait "parce qu'on m'a dit de le faire". Nous tenons à préciser ici que tout ce qui a été dit n'est absolument pas une critique envers le personnel mais simplement un exposé des problèmes rencontrés.

En conséquence, une communication plus large pourrait être menée pour expliquer l'intérêt de ces mises à jour et de transfert d'informations. Les études bibliométriques peuvent aider à améliorer la visibilité de l'INSA au niveau international, donc son image et par ricochet l'image des laboratoires. Lors de l'assemblée générale a été évoquée la mise en ligne des publications de chaque laboratoire. Cette idée a reçu un meilleur accueil de la part des directeurs de laboratoire qui voient en cela une façon de valoriser leurs activités. Mais cette mise en ligne sur le Web passe par un transfert régulier et fiable des données des bases LabINFO vers la base LabINSA.

Méthodologie

1. Comparaison quantitative des rapports quadriennaux, de la base LabINSA et du corpus d'Epsilon2004

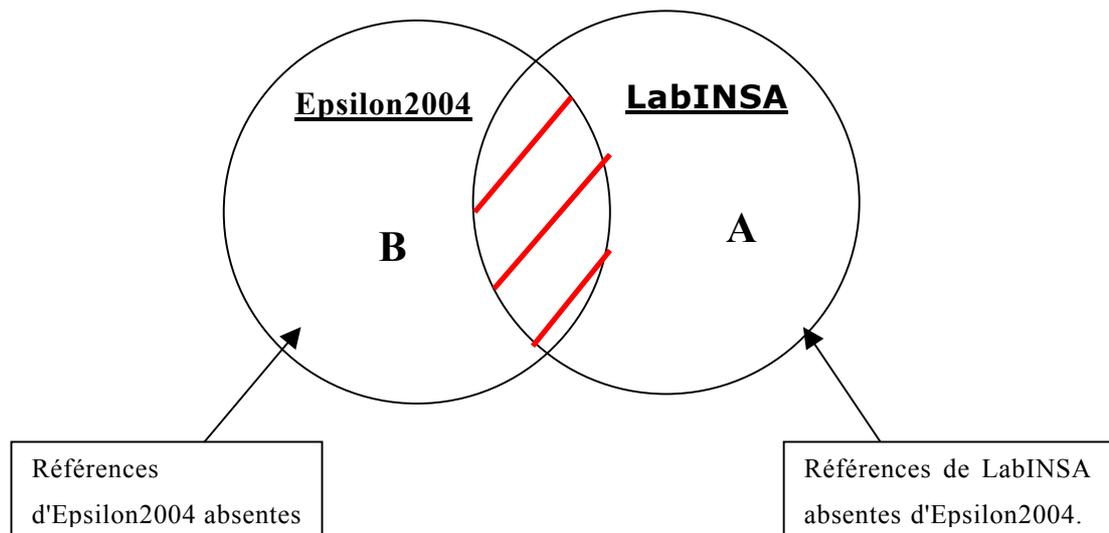
Nous avons effectué une première approche en comparant, sur un plan purement quantitatif, le nombre de publications présentes respectivement dans les rapports d'activités des laboratoires, la base LabINSA après le transfert de données de juillet, et le corpus d'Epsilon2004. Seuls les articles de périodiques scientifiques, ont été comptabilisés ; les autres types de document (brevets, congrès, ouvrages) ne sont pris en compte. Le principe de comptage est le compte de présence : "+1" pour une publication dont au moins un auteur fait parti de l'INSA.

2. Comparaison Quantitative et Qualitative du Corpus de la base LabINSA et du Corpus d'Epsilon2004

L'unité de comparaison des deux corpus est, là aussi, l'article de périodiques scientifiques. En effet il représente la très grande majorité du corpus d'Epsilon2004 (96%). Les années de comparaison seront choisies en fonction du nombre d'articles présents dans LabINSA. Cependant nous avons exclu immédiatement les années 2000 et 2003. En effet, les contacts que nous avons eus avec la Direction de la Recherche nous ont assuré que peu de laboratoires avaient saisi leurs publications en 2000 et qu'en conséquence la base LabINSA risquait d'être très incomplète. Pour l'année 2003, le problème vient du temps de latence des *Current Contents*. Adel Larouci (LAROUCI 2002) a estimé à 30 semaines le délai maximum de saisie des références bibliographiques. Or Camille Prime-Claverie a interrogé les *Current Contents* en février 2004 ; ce qui ne permet pas d'avoir la totalité des références pour l'année 2003. Ainsi notre comparaison s'est effectuée sur la période 2001/2002.

2.1. Répartition des articles entre la base LabINSA et Epsilon2004

Le croisement des données entre les corpus d'Epsilon2004 et la base LabINSA peut être schématisé comme suit :



Répartition des articles entre la base LabINSA et Epsilon2004

La partie B ne concerne pas directement notre étude. L'hypothèse la plus probable qu'un article soit présent dans le corpus d'Epsilon2004 et pas dans la base LabINSA, est qu'il n'ait pas été saisi dans la base LabINSA.

La partie hachurée représente l'ensemble des articles présents à la fois dans le corpus d'Epsilon2004 et dans la base LabINSA. Elle peut nous servir à comprendre pourquoi certains articles sont retenus par les *Current Contents* et d'autres non.

La partie A est celle qui intéresse directement notre étude. Elle regroupe les articles présents dans la base LabINSA et absents du corpus d'Epsilon2004. C'est cette partie que nous avons étudiée très précisément.

Rappels à propos du corpus d'Epsilon2004 :

- Le corpus initial a été obtenu par une interrogation sur les villes où se situent les établissements du PUL. Comme expliqué précédemment (voir page 21), les articles dont la ville n'a pas été précisée, n'ont pas pu être intégrés dans ce corpus. Il sera indiquer plus loin des exemples de tels articles.
- Pour attribuer des publications à l'INSA Camille Prime-Claverie utilise un filtre, affectant "+1" à chaque établissement présent dans le champs "institution" des références. Comme expliqué précédemment (voir page 21), ce filtre peut avoir surévalué le nombre de publications des laboratoires inter-établissements. Une analyse des auteurs des articles de ces laboratoires a été réalisée pour estimer le biais.

Rappels à propos du corpus de LabINSA :

- Pour les laboratoires ayant plusieurs établissements de tutelle, tous les articles sont saisis dans LabINSA. Or certains articles peuvent n'avoir aucun auteur faisant partie de l'INSA (page 24 paragraphe 2.1). Déterminer l'établissement de rattachement d'un auteur a été une des difficultés majeurs de notre étude.
- Compte tenu de la possibilité de doublons (voir page 26) dans le corpus de LabINSA, une **clef de dédoublonnage** sera réalisée pour estimer la proportion de ces doublons.

2.2. Choix méthodologiques et réalisation technique

L'interface de LabINSA sur laquelle nous travaillons est une interface Access97.

2.2.1 Clef de dédoublonnage

La possibilité de doublons dans la table PRODUCTION de la base LabINSA, nous a obligé à réaliser tout d'abord une clef de dédoublonnage. Une production est identifiée de façon unique par le titre, l'année et le type de production. Dans notre cas nous ne nous intéressons qu'aux articles de périodiques, notre clef de dédoublonnage a donc porté sur le titre et l'année (limitée à 2001 et 2002), et nous avons ajouté une condition sur le type de production, à savoir "AP" pour article de périodique. Puis nous avons rédigé la requêtes en mode SQL correspondant à la clef de dédoublonnage (voir annexe 10).

2.2.2 Constitution d'une table des périodiques scientifiques de LabINSA

Pour apporter une solution au problème posé par le fait que chaque laboratoire ait sa propre table PERIODIQUE (voir page 26), il nous a été demandé de réaliser une table contenant tous les libellés avec leurs ISSN, des revues indiquées dans la base LabINSA. La table PERIODIQUE de la base LabINSA contient 1321 libellés de périodiques que nous avons dédoublonnés d'une part, et dont nous avons vérifié ou recherché les ISSN d'autres part. Nous avons complété les ISSN manquants grâce au catalogue du SUDOC (<http://www.sudoc.abes.fr/>).

Puis nous avons désiré croiser notre corpus de revues avec celui des revues dépouillées par l'ISI. Ainsi nous pouvions connaître la proportion de revues, dans lesquelles publient les chercheurs de l'INSA, traitées par l'ISI. Ceci a nécessité la création d'une table contenant les libellés et ISSN des revues dépouillées par l'ISI. Nous avons créé cette table à partir du site de l'ISI (<http://www.isinet.com/cgi-bin/jrnlst/jlresults.cgi?PC=MASTER>). La procédure de création de cette table est en annexe 9.

2.2.3 Les articles dont au moins un auteur est rattaché à l'INSA

Avant de pouvoir faire notre comparaison, nous devons connaître les articles de périodiques scientifiques dont au moins un auteur fait partie de l'INSA. Ces articles seront appelés les articles "INSA" dans la suite du rapport. Pour cela nous avons identifié très précisément le personnel rattaché à l'INSA.

2.2.3.1 Identification du personnel INSA dans la base LabINSA

Dans la table PERSONNEL de LabINSA, deux champs peuvent nous renseigner sur le personnel des laboratoires.

1. Le code Etablissement : ce code désigne l'établissement de rattachement d'une personne. Cependant **ce code est facultatif** lors de la saisie. Aussi nombreuses sont les personnes dont le code n'est pas renseigné (ou "Null"). Les codes établissement de l'INSA sont 27 ou 501. Ces deux codes nous serviront lors de la rédaction de nos requêtes SQL.

Par convention lors de la mise en place de la base LabINSA, il a été convenu que lorsque le code établissement n'était pas renseigné, la personne fait partie de l'INSA par défaut. Or nous n'avons eu aucun moyen de vérifier cette convention. Nous n'avons donc retenu celle-ci que partiellement (voir ci-après)

et nous avons essayé de déterminer autant que possible l'établissement de rattachement du personnel. Pour cela nous avons consulté le site Web de l'INSA (www.insa-lyon.fr/), ainsi que les sites Web des laboratoires (les liens vers ces sites sont sur le site de l'INSA : www.insa-lyon.fr/pg/index.php?Rub=63). Pour toutes les personnes identifiées INSA nous avons saisi le code établissement 27 dans la table PERSONNEL.

2. Le code Qualité : ce code, en revanche, **est obligatoire** lors de la saisie. Il désigne la qualité de la personne, à savoir si elle est Chercheur, Enseignant, Doctorants etc. Les codes qualité sont au nombre de quatorze (voir tableau 3).

Tableau 3 : table QUALITE de LabINSA

CD_QUALITE	LIBELLE_QUALITE
AU	AUTRE PERSONNEL
CE	C.E.S.
CH	CHERCHEUR
CN	MEMOIRE CNAM
DE	DEA
DI	IATOS DEA
DO	DOCTORANT
ES	ENSEIGNEMENT SUPERIEUR
EX	INTERVENANT EXTERIEUR
PO	POST-DOCTORANT (labo)
SD	SECOND DEGRE
ST	STAGIAIRE
TA	IATOS
TD	IATOS DOCTORANTS

Nous avons exclu d'emblée les Intervenants Extérieurs, C.E.S., "Autre Personnel", "Mémoire CNAM" et Stagiaires ; soit les codes qualité AU, CE, CN, EX, ST.

Pour le personnel DEA, IATOS DEA, Doctorant et IATOS Doctorant dont le code établissement n'est pas renseigné, le statut des laboratoires intervient alors, selon que les laboratoires ont un ou plusieurs organismes de tutelle :

- Si le laboratoire a comme seule tutelle l'INSA, les Doctorants et DEA sont naturellement rattaché à l'INSA. Aussi, pour ce type de laboratoire, nous avons conservé les codes qualité : DE, DI, DO et TD. Pour la même raison, une personne dont le code établissement n'est pas renseigné et dont le code qualité n'est pas AU, CE, CN, EX ou ST, fait également partie de l'INSA. Aussi pour ces laboratoires nous avons conservé la convention signalée ci-dessus ; à savoir

qu'un code établissement "NULL" signifie que la personne fait partie de l'INSA.

- Si le laboratoire a plusieurs organismes de tutelle, nous ne pouvions savoir à quel établissement sont rattachés les DEA et Doctorants, si leur code établissement n'était pas renseigné. Nous avons donc exclu les codes qualité : DE, DI, DO et TD. Pour la même raison, la convention du code établissement "NULL" est exclue dans ce cas.

Remarque : il est à noter que l'exclusion des DEA et Doctorants n'engendre pas un biais très important quant au nombre d'articles "INSA" qui ont été trouvés. En effet, renseignements pris auprès de Thierry Lafouge et Monique Joly, les DEA et Doctorants ne publient qu'exceptionnellement seuls des articles de périodiques scientifiques. Dans la grande majorité des cas ils sont associés à leur tuteur qui est un chercheur du laboratoire. Nous retrouverons donc les publications des Doctorants "INSA" grâce au nom du tuteur, si celui-ci fait parti de l'INSA.

Il nous reste à traiter le personnel rattaché aux EPST. Là encore le statut des laboratoires entre en jeu. Si l'INSA est seule tutelle du laboratoire le personnel EPST fait obligatoirement partie de l'INSA. Dans ce cas nous avons ajouté aux codes établissement déjà retenus pour nos requêtes, les codes 28 du CNRS et 124 de l'INSERM. En revanche si l'INSA n'est pas seule tutelle du laboratoire, nous pouvons connaître le personnel EPST de l'INSA pour 2003/2004 grâce à la liste que nous a fournie la Direction de la Recherche. Pour ces personnes nous remplaçons le code EPST par le code INSA 27.

En conclusion:

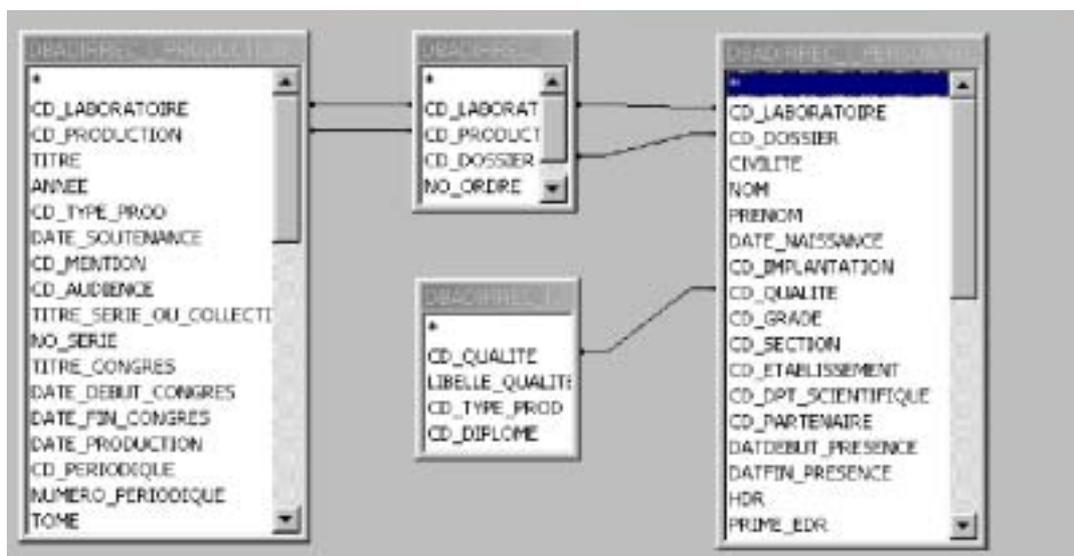
- Pour les laboratoires inter-établissements, nous avons exclu les codes qualité : DE, DI, DO, TD, AU, EX, CN, CE, ST, PO ainsi que les personnels de code établissement "Null". Par ailleurs nous n'avons conservé que les personnels EPST 2003/2004.
- Pour les laboratoires exclusivement sous tutelle de l'INSA, nous avons exclu les code qualité : AU, EX, CN, CE, ST. En revanche, nous avons gardé également le critère "Null" pour le code établissement ; nous avons ajouté les codes 28 (CNRS) et 124 (INSERM) ; et tout le personnel EPST a été conservé.

2.2.3.2 Rédaction des requêtes SQL

Après avoir défini les conditions sur les codes Etablissement et Qualité, nous avons rédigé deux requêtes, selon le statut du laboratoire, pour obtenir les articles "INSA". Pour cela nous avons eu besoin des tables PRODUCTION, PRODUCTION-AUTEUR, PERSONNEL et QUALITE de la base LabINSA.

- La table PRODUCTION contient les productions scientifiques des laboratoires.
- La table PERSONNEL contient le personnel des laboratoires.
- La table PRODUCTION-AUTEUR fait le lien entre les deux tables précédentes grâce au code Laboratoire, au code production et au code Dossier.
- La table QUALITE donne la "Qualité" de chaque personne.

L'organisation des tables dans LabINSA est présentée ci-dessous. Les requêtes en mode SQL sont en annexe 10.



2.2.4 Comparaison des corpus LabINSA et Epsilon2004

2.2.4.1 Les articles "INSA" présents dans le corpus Epsilon2004

Tout d'abord, nous avons créé une table contenant les références bibliographiques des articles "INSA" à partir de la table PRODUCTION de la base LabINSA. Nous nous sommes limités aux années 2001 et 2002 pour les raisons évoquées auparavant (page 30, paragraphe 2). Puis nous avons dédoublonné **manuellement** nos résultats en utilisant le titre. Nous n'avons pas fait cela de façon automatique car nous craignons les erreurs de saisie. Par la suite nous verrons que nos craintes se sont confirmées. La requête SQL rédigée pour réaliser la sélection des références INSA est en annexe 10.

En parallèle, nous avons créé un deuxième corpus grâce aux filtres mis au point par Camille Prime-Claverie. A partir du corpus global d'Epsilon2004, nous avons sélectionné les articles de l'INSA, pour les années 2001 et 2002. Un problème technique s'est posé. Les champs "*authors*" et "*institutions*" des références bibliographiques sont du type "Mémo". Or ce type ne permet pas d'utiliser le prédicat "SELECT DISTINCT" dans la requête ; prédicat qui permet de dédoublonner les résultats. Aussi avons nous dû sélectionner nos références bibliographiques (voir la requête SQL en annexe 10); puis nous avons transformé le champs "*title*" en clef primaire pour dédoublonner nos résultats. Puis nous avons trié les références bibliographiques pour ne garder que les références des articles "INSA".

Nous avons, par la suite, comparé ces 2 corpus. Pour cela nous avons utilisé le titre et l'année. Plusieurs essais nous ont montré que traiter séparément les années 2001 et 2002 ne change rien, les résultats sont strictement identiques. En revanche une comparaison automatique stricte, au moyen d'une requête SQL, des titres fait perdre des références car le titre de l'article peut avoir été saisi différemment dans la base LabINSA et dans les *Current Contents* : un exemple des plus courant est la présence d'un point à la fin du titre dans la base LabINSA. Nous avons donc décidé de faire une comparaison automatique sur les 25 premiers caractères des titres. Puis nous avons vérifié manuellement nos résultats pour éliminer les doublons dus à des titres débutant par les mêmes termes mais finissant différemment.

Nous nous sommes très vite aperçus que la comparaison automatique ne serait pas valable. Les fautes d'orthographe, les pluriels oubliés ou rajoutés, les mots inversés, les espaces omis ou rajoutés, les molécules chimiques écrites en toutes lettres ou avec les symboles etc. rendaient la comparaison automatique quasiment impossible. Le taux d'erreur était beaucoup trop élevé. Cela a confirmé nos craintes exprimées ci-dessus. Nous avons donc réalisé une comparaison manuelle des titres entre les deux corpus. Nous obtenons un troisième corpus que nous avons appelé le corpus C contenant les références bibliographiques communes à la base LabINSA et à Epsilon2004. Ce corpus C est construit avec les références bibliographiques de la base LabINSA pour permettre une soustraction automatique avec le corpus des articles "INSA" de cette même base LabINSA.

Remarque : nous présentons tout de même la requête SQL utilisée pour comparer les références bibliographiques des deux corpus sur les années considérées et les 25 premiers caractères du titre en annexe 10.

2.2.4.2 Les articles "INSA" absents du corpus Epsilon2004

Nous avons recherché les références bibliographiques des articles "INSA" absents du corpus d'Epsilon2004. Pour cela nous avons réalisé une différence symétrique entre notre corpus C et le corpus de la base LabINSA (voir la requête SQL en annexe 10). Les références bibliographiques obtenues ont été recherchées dans les *Current Contents*, pour vérifier qu'effectivement ces articles n'y étaient pas. Nous avons obtenu un corpus D de références bibliographiques correspondant aux articles "INSA" absents du corpus d'Epsilon2004 **mais aussi** des *Current Contents*. Nous avons donc changé d'échelle ici puisque nous travaillons avec un corpus d'articles "INSA" absents de la base de l'ISI.

2.2.5 Analyse du corpus D

2.2.5.1 Répartition des articles "INSA" selon les laboratoires et les pôles de compétences

Nous avons réparti, selon les laboratoires et selon les 5 pôles de compétences définis par l'INSA, d'une part les articles "INSA" du corpus de la base LabINSA et d'autres part les articles "INSA" du corpus D. Cela nous permet de savoir la

proportion d'articles absents des *Current Contents* pour chaque laboratoire mais également par pôle de compétences.

De plus nous avons réalisé le test du Khi-deux pour vérifier les hypothèses suivantes :

- Est-ce que le statut inter-établissement des laboratoires influe sur le nombre d'articles présents dans les *Current Contents* ?
- Est-ce que le statut EPST des laboratoires EPST influe sur le nombre d'articles présents dans les *Current Contents* ?

2.2.5.2 Analyse de la revue des articles du corpus D

Nous avons souhaité ici vérifier l'une de nos hypothèses, à savoir que l'ISI ne dépouillerait qu'une partie des revues dans lesquelles publient les laboratoires de l'INSA. Pour cela nous avons repris notre table des revues de l'ISI créé ci-dessus (voir page 33, paragraphe 2.2.2).

En parallèle, nous avons recherché le périodique des articles du corpus D (voir la requête SQL en annexe 10). Ceci grâce au code périodique de la table PRODUCTION qui renvoie à la table PERIODIQUE. Cependant nous avons rencontré deux difficultés. Tout d'abord le code périodique de la table PRODUCTION n'est pas obligatoire. Nous n'avons donc pas pu retrouver la revue des articles dont le code n'était pas renseigné. De plus, pour comparer les revues, nous utilisons l'International Standard Serial Number (ou ISSN). Ainsi la comparaison peut être automatique et les erreurs de saisie des libellés des périodiques sont évitées. Or dans la base LabINSA, le champs "ISSN" de la table PERIODIQUE n'est pas obligatoire. Aussi n'était-il presque jamais renseigné. Nous avons donc complété les ISSN des revues grâce au catalogue du SUDOC (<http://www.sudoc.abes.fr/>) ; mais aussi grâce au site de l'ISI (<http://www.isinet.com/cgi-bin/jrnlst/jlsearch.cgi?PC=MASTER>). Puis nous avons pu comparer (voir la requête SQL en annexe 10) les revues dépouillées par l'ISI et celles des articles "INSA".

Résultats et Discussion

1. Trois vues possibles de la production scientifique de l'INSA

Les tableaux suivants présentent 3 vues différentes des productions scientifiques des laboratoires de l'INSA. Une selon les rapports d'activités ; une selon le contenu de la base LabINSA avant le transfert des données prévu au mois de juillet, et une selon les données d'Epsilon2004. Nous rappelons que seuls les articles de périodiques scientifiques ont été comptabilisés.

1.1. Articles présentés par les rapports d'activité

Après avoir réalisé le dépouillement manuel des rapports quadriennaux, nous avons réparti les articles par année et par laboratoire (voir le tableau 4). Pour la période 1998/2001, **le nombre moyen est de 376 articles par an.**

Nous tenons à préciser ici qu'une comparaison entre laboratoires ne signifierait absolument rien compte tenu du statut différent de la notion de revue selon les disciplines. La seule valeur comparable avec les autres corpus est le total par année.

Pour les laboratoires ayant plusieurs organismes de tutelle (en *italique* dans le tableau 4), seuls les articles publiés par les chercheurs INSA ont été comptabilisés (grâce au concours de Jean Michel Jolion).

Tableau 4 : Nombres d'articles par année et par laboratoires selon les rapports quadriennaux

Laboratoire	1997	1998	1999	2000	2001	2002	2003
BF2I	9	14	14	7	11	1	
CEGELY (INSA)		10	13	5	7	1	
CETHIL (INSA)		10	16	21	19	17	18
CITI		7	7	4		6	
CNDRI		1	5	4	2	1	
CREATIS (INSA)		24	19	19	18	13	
EDU/EVS		2		2	2		
ERM 107 (INSA)	2	1	0	2	2		
GEMPPM		58	57	46	81		
ICTT (INSA)		5	3	1	3		
IMP(LMM-INSA)		31	37	29	23		
LAEPSI		8	8	14	6		
LAI		6	3	1	7		
LaMCoS		32	32	27	39	2	
LCO (INSA)		5	5	6	4		
LDMS/CASM		12	10	6	2	5	
LGEF		4	5	9	6		
LIRIS (INSA)		13	10	8	13		
LMFA		9	12	6	8		
LPCI		2	9	8	12	1	
LPM		53	65	33	32		
LVA		5	3	6	2	2	
MAPLY (INSA)		11	9	15	10		
PLM		20	25	23	21	3	
PRISMa (INSA)		7	3	12	8		
UMG		13	16	7	8	3	
URGC		18	29	27	16		
Totaux/année	11	381	415	348	362	55	18
TOTAL	1590						

Le laboratoire ESCHIL n'a pas de rapport quadriennal. Ce laboratoire a un statut très particulier (selon Mr Reynouard Directeur de la Direction de la Recherche), et a été écarté de l'étude par la suite.

Le laboratoire IMP est composé de deux équipes : une exclusivement Université Claude Bernard ; une autre, l'équipe LMM, affiliée à l'INSA. Seules les publications des chercheurs de l'équipe LMM ont été comptabilisées.

Jean Michel Jolion avait comptabilisé l'unité ERM 107. Cette unité n'existe pas sur le site de l'INSA. Après informations prises auprès de Monsieur Reynouard, cette équipe est rattachée à l'INSERM et est en liaison avec le laboratoire LIRIS. Après accord avec Jean-Michel Jolion, l'unité ERM 107 a été écartée de l'étude.

1.2. Articles "INSA" présents dans LabINSA

Le tableau ci-dessous présente le nombre d'articles "INSA" présents dans LabINSA fin juillet 2004.

Tableau 5 : Nombres d'articles "INSA" dans la base LabINSA par année et par laboratoire

Laboratoire	2000	2001	2002	2003	2004
BF2I	3	5	2	1	
CEGELY				3	0
CETHIL	26	21	23	19	
CNDRI	4	2	1		
CREATIS	36	45	4		
ESCHIL			2	4	
EVS/EDU			1	1	
GEMPPM	55	81	44	23	0
LAEPSI				15	
LAI	1	8	3	2	1
LaMCoS	16	35	55	35	3
LCO	6	4	6	1	
LDMS	5	2	7	2	0
LGEF	8	5	3	2	
LIRIS	7	23	4	9	
LPCI	6	15	2	2	
LPM		0	36	30	
LVA	7	2	6	5	
MAPLY-INSA			7	4	0
PLM	7	9	14	32	2
PRISMa				4	
UMG	5	13	6		
URGC	29	21	17	8	
Totaux/année	221	291	243	202	6
TOTAUX	963				

Pour les laboratoires CITI, ICTT, IMP(LMM) et LMFA, aucune donnée n'avait été saisie dans la base LabINSA fin juin 2004.

Pour les laboratoires PRiSMa et LAEPSI, les données n'ont été saisies qu'à partir de 2003. Nous ne retrouverons pas ces laboratoires dans la suite de notre étude.

Notre choix de comparer les corpus de la base LabINSA et d'Epsilon2004 sur les années 2001 et 2002, se trouve confirmé ici. L'année 2001 est l'année où le plus grand nombre d'articles ont été saisis dans la base LabINSA, soit 291 articles ; et l'année 2002 est l'année où le plus grand nombre de laboratoires, soit 20 laboratoires, ont saisi leurs données. **Nous avons donc comparé les deux corpus sur les années 2001 et 2002.**

1.3. Articles "INSA" par l'étude d'Epsilon2004

Le tableau ci-dessous présente le nombre d'articles de périodiques attribués à l'INSA par l'étude Epsilon2004, pour la période de 2000 à 2003.

Tableau 6 : nombres d'articles attribués à l'INSA

	2000	2001	2002	2003
INSA	252	255	287	325
TOTAL	1119			

1.4. Résultats Globaux et Conclusion

Les résultats globaux pour la période 2000/2003 sont les suivants.

Tableau 7 : répartition des articles "INSA" par année, selon les 3 sources d'informations

TOTAUX	2000	2001	2002	2003	Total
Rapports	346	360			
LabINSA	221	291	243	202	957
Epsilon2004	252	255	287	325	1133

Remarque: pour les rapports quadriennaux, le nombre de publications est incomplet pour 2002 et 2003. La somme n'a donc pas été calculée.

On peut noter que l'on retrouve globalement plus d'articles dans le corpus d'Epsilon2004 que dans celui de LabINSA (15.5% d'écart). Cette constatation est à relativiser en fonction des années. Pour 2000 et 2001 les écarts entre Epsilon2004 et LabINSA sont plus faibles (LabINSA présente même plus d'articles en 2001), que pour 2002 et 2003. L'hypothèse la plus vraisemblable est la non saisie par les laboratoires de toutes leurs publications pour 2002 et 2003.

En revanche, pour 2000 et 2001, années où les rapports d'activités sont "complets" pour les articles, les écarts entre les rapports quadriennaux d'un côté et les corpus LabINSA et Epsilon2004 de l'autre sont importants.

Le corpus de LabINSA ne représente que 64% du corpus des rapports quadriennaux en 2000 et 81% en 2001. Cela traduit un manque de rapatriement des données dans la base LabINSA confirmant ainsi l'hypothèse émise ci-dessus.

2. Comparaison des corpus de LabINSA et d'Epsilon2004

2.1. La Clef de dédoublement

Notre clef de dédoublement porte sur l'intégralité du titre, l'année (limitée à 2001 et 2002) et sur les articles de périodiques. Nous avons obtenu 3 doublons et 1 triplé. Leurs codes laboratoires associés à leurs codes production sont dans le tableau 8.

Tableau 8 : articles en doublons dans la base LabINSA

LPM 119	CEGELY 56	
LAMCOS 1278	MAPLY-INSA 7	
CNDRI 15	GEMPPM 400	
GEMPPM 480	LAMCOS 1102	CREATIS 355

Ainsi il existe bien des doublons dans la base LabINSA. Notre clef ne les fait pas tous apparaître car elle est limitée sur l'année et sur le type de production. On peut donc légitimement penser que d'autres doublons existent.

Par la suite, compte tenu des problèmes évoqués dans la partie précédente concernant les fautes d'orthographe, les pluriels oubliés, les mots inversés etc., nous avons dédoublonné **manuellement** la table contenant les articles "INSA" pour 2001 et 2002. Nous avons trouvé **21 doublons de plus**. L'ensemble des doublons est dans le tableau 9 page 47.

On constate donc que le fait de dédoubler manuellement fait ressortir beaucoup plus de doublons. Aussi la réalisation d'une clef de dédoublement pour rechercher de façon automatique les doublons, s'avère inefficace. Les saisies dans LabINSA comportent énormément d'erreurs.

En conséquence, une vérification de tout ce qui a été saisi dans les bases LabINFO est nécessaire. Ceci permettra une exploitation automatique des données de la base LabINSA.

2.2. Les périodiques de la base LabINSA

La table PERIODIQUE de la base LabINSA comprenait 1321 libellés de périodiques. Les libellés étaient très mal renseignés : fautes d'orthographe, abrégés erronés etc. Aussi avons nous eu beaucoup de mal à les dédoublonner mais aussi à rechercher les ISSN correspondants. Par manque de temps nous n'avons pas pu finir ce travail. Au final nous avons obtenu 1011 libellés de périodiques dont 451 ont un ISSN renseigné.

Puis nous avons croisé notre table ci-dessus avec la table comprenant les libellés et ISSN des revues dépouillées par l'ISI. Nous avons obtenu 308 revues communes soit 68.3% du corpus des revues que nous avons identifiées (ISSN connu).

2.3. Les articles "INSA" de la base LabINSA

Nous avons rédigé deux requêtes en mode SQL selon le type de laboratoire étudié. Nous obtenons :

- 471 articles "INSA" pour les laboratoires inter-établissements.
- 1405 articles "INSA" pour les laboratoires dépendant uniquement de l'INSA.

La base LabINSA contient 2168 articles de périodiques scientifiques. **Ainsi, sur 2168 articles présents dans la base LabINSA, 1876 sont des articles "INSA".**

Cependant, il faut remarquer que nos choix concernant le code établissement et le code qualité ne résolvent pas toutes les difficultés. En parcourant le corpus de LabINSA et en effectuant différentes recherches, nous avons trouvé des références bibliographiques assez atypiques.

Tout d'abord certains articles sont le résultat du travail de personnes toutes qualifiées "d'extérieures" au laboratoire (code qualité "EX"), telle la publication n°1348 du laboratoire LaMCoS. Cela paraît étrange et pose problème pour les laboratoires rattachés seulement à l'INSA, puisqu'au moins un auteur est sensé appartenir au laboratoire donc à l'INSA. Compte tenu du code qualité, nous avons tout de même exclu ces publications.

Pour les laboratoires inter-établissements, certains articles ont été rédigés par des personnes "extérieures" et des chercheurs du laboratoire. Or si le code

établissement des chercheurs n'était pas renseigné, nous n'avons pas pu rattacher l'article à un établissement (LIRIS n°133). Ces publications ont donc été exclues. Certains articles ont été publiés par un unique doctorant (LIRIS n°62). Or pour les laboratoires inter-établissements, nous n'avons pas pu rattacher le doctorant à son établissement, si son code n'était pas renseigné. La publication a donc été exclue. Enfin, nous n'avons travaillé qu'avec la liste du personnel EPST pour l'année 2003/2004. N'ayant pas les listes des années précédentes, nous n'avons pas pu changer le code établissement pour le personnel des laboratoires inter-établissements.

Nous avons recherché les publications des auteurs dont nous n'avions pas pu déterminer l'établissement de rattachement, pour estimer le biais engendré par ces publications atypiques. Après avoir effectué ces recherches et complété la table PERSONNEL de la base LabINSA nous avons acquis la certitude que le nombre de publications telles que celles présentées ci-dessus, est tout à fait négligeable.

2.4. Les articles "INSA" présents dans le corpus d'Epsilon2004

Suite à nos interrogations des corpus de la base LabINSA et d'Epsilon2004, nous avons trouvé, pour les deux années 2001 et 2002, 530 articles "INSA" dans la base LabINSA et 575 articles attribués à l'INSA dans le corpus d'Epsilon2004. nous faisons la distinction ici entre un article "INSA" dont nous sommes sûr qu'au moins un auteur fait partie de l'INSA, et un article attribué à l'INSA par les filtres de Camille Prime-Claverie dont nous n'avons pas vérifié les auteurs.

Enfin comme signalé ci-dessus (page 44 paragraphe 2.1), nous avons dédoublonné manuellement nos 530 articles "INSA" et obtenu 25 doublons et triplé.

Tableau 9 : liste des laboratoires ayant généré un article "INSA" en doublon ou en triplé pour 2001 et 2002

GEMPPM 480	LaMCos 1102	CREATIS 355
GEMPPM 871	CNDRI 46	
LAMCOS 661	MAPLY-INSA 19	
CNDRI 16	CREATIS 373	
CEGELY 56	LPM 119	
LAMCOS 721	MAPLY-INSA 23	
UMG 8	LCO 34	
LAMCOS 880	URGC 1161	
LAMCOS 683	MAPLY-INSA 4	
LCO 39	PLM 91	
LAMCOS 94	URGC 1354	
CETHIL 444	IMP-INSA 71	
LAMCOS 1278	MAPLY-INSA 7	
CNDRI 15	GEMPPM 400	
URGC 942	LAMCOS 1379	
LPCI 108	LPCI 129	
GEMPPM 482	GEMPPM 403	
UMG 18	UMG 24	
UMG 21	UMG 27	
GEMPPM 399	GEMPPM 398	
UMG 25	UMG 19	
UMG 23	UMG 17	
LPM 122	LPM 11	
LPCI 136	LPCI 131	
UMG 26	UMG 20	

Sur ces 25 doublons, 10 (en gras dans le tableau 9) sont des doublons propres à un laboratoire ; ils ont donc été supprimés. En revanche, les 15 autres sont des doublons entre laboratoires. Nous désirions faire une répartition des articles "INSA" entre laboratoires et selon les 5 pôles de compétences définis par l'INSA. Conséquence nous avons conservé ces 15 doublons puisque aucune raison ne justifiait d'enlever une publication à un laboratoire plutôt qu'à un autre.

Par ailleurs comme nous l'avons expliqué page 21, la méthodologie utilisée par Camille Prime-Claverie peut avoir surévalué le nombre d'articles de l'INSA dans le corpus initial d'Epsilon2004. Nous avons donc analysé les 575 articles du corpus d'Epsilon2004. Sur ces 575 articles, 272 articles étaient absents de la base LabINSA. Sur ces 272 articles, nous avons recherché uniquement les articles des laboratoires inter-établissements, grâce au champs "institution" des notices bibliographiques. Nous avons obtenu 158 articles. Cependant cette liste est loin d'être exhaustive. Nous avons rencontré des articles dont le champs "institution" contenait seulement une appellation de l'INSA (INSA ou Inst Natl Sci Appl etc.)

sans précision quant au laboratoire. Dans ces cas là nous n'avons pu attribuer les articles à un établissement précis.

Ainsi sur les 158 articles, 33 articles n'ont aucun auteur identifié INSA. Par conséquent nous pouvons dire que le corpus d'Epsilon2004 des articles attribués à l'INSA se "réduit" à 542 articles "INSA".

Nous pouvons également estimer le nombre d'articles "INSA" (520 articles de la base LabINSA et 239 articles du corpus d'Epsilon2004 absents de la base LabINSA), soit un total de **759 articles "INSA" pour la période de 2001 à 2002**. Ceci correspond à un nombre moyen de 379 articles par année. Ce résultat est en accord avec les rapports quadriennaux qui indiquent une moyenne de 376 articles par an (voir page 39).

En conclusion nous avons obtenu:

- **520 articles "INSA", pour 2001/2002, issus du corpus de LabINSA.**
- **542 articles "INSA", pour 2001/2002, dans le corpus d'Epsilon2004.**

Nous constatons que le corpus d'Epsilon2004 contient plus de références bibliographiques d'articles que la base LabINSA elle-même. Comme évoquées ci-dessus, plusieurs hypothèses peuvent expliquer cela :

- Les articles n'ont pas été saisis dans la base LabINSA. Cette hypothèse est confirmée avec force, par le fait suivant. Nous avons constaté qu'aucune publication de 2001 et 2002, pour les laboratoires CITI, LAEPSI et PRISMA, n'avait été saisie dans la base LabINSA. Or nous avons retrouvé des publications de 2001 et 2002, de ces laboratoires dans le corpus d'Epsilon2004.
- Certains articles du corpus d'Epsilon2004 peuvent ne pas être des articles de l'INSA, si aucun auteur ne fait partie de l'INSA.

Puis nous avons comparé **manuellement** ces deux corpus en utilisant le titre et l'année. Nous avons construit une nouvelle table contenant les articles communs. Nous avons obtenu 303 articles (**notre corpus C**) communs aux deux corpus. **Les 217 articles "INSA" restant sont absents du corpus d'Epsilon2004.**

2.5. Les articles "INSA" de la base LabINSA absents du corpus Epsilon2004

Nous avons réalisé une différence symétrique, grâce à une requête SQL (voir annexe 10) entre notre corpus de 520 articles et notre corpus de 303 articles. Nous avons obtenu 217 articles

Par la suite nous avons repris ce corpus et vérifié qu'effectivement ces 217 articles n'étaient pas dans les *Current Contents*. Nous avons donc changé ici d'échelle puisque nous travaillons avec la base des *Current Contents* et non plus avec le corpus d'Epsilon2004.

Nos résultats ont montré que 52 articles, absents du corpus d'Epsilon2004, étaient **présents** dans les *Current Contents*. Nous avons créé une table spécifique pour ces 52 références bibliographiques, afin d'identifier savoir pourquoi elles ne figurent pas dans le corpus d'Epsilon2004.

Le détail des 52 articles est le suivant :

- 23 articles ne font pas partie du corpus d'Epsilon2004 à cause de **l'adressage** (voir l'exemple 1 en annexe 11). Le champs "Institutions" ne contient aucun des noms de ville utilisés dans l'équation 2 par Camille Prime-Claverie.
- 18 sont dans le corpus d'Epsilon2004 mais ne sont pas retenus par notre requête portant sur les articles de l'INSA en 2001 et 2002. Notre requête utilise à la fois les filtres informatiques de Camille Prime-Claverie et deux conditions sur l'année (2001 et 2002) et le type de production (article).
7 articles ne sont pas retenus par les filtres de Camille Prime-Claverie à cause d'un **adressage** incomplet (voir l'exemple 2 en annexe 11). Rien dans le champs "Institutions" n'évoque l'INSA (abréviation, sigle etc.) et le numéro d'UMR est absent. 6 articles sont signalés "*article*" dans LabINSA et "*Review*" dans les *Current Contents*. 2 articles ont une année de publication différente entre la base LabINSA et les *Current Contents*. Enfin nous n'avons pas pu identifier la cause pour 3 articles.
- 11 articles sont présents dans le corpus et n'ont pas été retrouvés lors de notre premier tri manuel. La cause en est **le titre** (voir l'exemple 3 en annexe 11).

Soit le titre est en français dans la base LabINSA, en anglais dans les *Current Contents* ; soit il est incomplet dans la base LabINSA.

Dans les deux premiers cas ci-dessus, nous constatons que nous touchons à la limite de la méthodologie employée par Camille Prime-Claverie et à la limite de qualité des données de l'INSA. Cependant cela est due moins à une faiblesse de la méthodologie utilisée, que nous estimons par ailleurs très pertinente, qu'à un adressage extrêmement laxiste des publications par leurs auteurs. L'adressage joue un rôle essentiel dans la visibilité des publications. Une normalisation doit être envisagée et les auteurs doivent être sensibilisés à ce problème.

De façon globale, nous avons donc :

- **355 articles "INSA" présents dans les *Current Contents*.**
- **165 articles "INSA" absents des *Current Contents*, soit 31.73%.**

Ces 165 articles constituent notre corpus D.

2.6. Analyse du Corpus D

2.6.1 Répartition des articles "INSA" selon les laboratoires et les pôles de compétences

Le tableau 10 présente la répartition.

Tableau 10 : répartition des articles "INSA"

Sigle du Laboratoire	Nombre d'articles	
	dans LabINSA	absents des Current Contents
BF2I	7	1
CEGELY	13	8
CETHIL	35	13
CNDRI	5	0
CREATIS	27	8
EVS	1	1
GEMPPM	123	32
ICTT	3	1
IMP-INSA	19	1
LAI	11	5
LaMCoS	87	35
LCO	10	1
LDMS	10	3
LGEF	8	4
LIRIS	17	11
LMFA-INSA	6	3
LPCI	16	10
LPM	34	5
LVA	8	1
MAPLY-INSA	9	3
PLM	21	0
UMG	14	1
URGC	36	18
Total	520	165

Sur 520 références bibliographiques, 165 sont absentes des *Current Contents*. Ainsi **68,27 % des articles de l'INSA sont présents dans la base de l'ISI.**

Nous avons fait la même répartition selon les 5 pôles de compétences de l'INSA (voir présentation des pôles page 13 et 14). Nous avons indiqué dans la dernière colonne du tableau 11, la proportion d'articles absents des *Current Contents* par rapport au nombre d'articles "INSA" total.

Tableau 11 : répartition des articles "INSA" selon les 5 pôles de compétences.

Pôles de compétences	Nombre d'articles		Proportion (%)
	dans LabINSA	absents des <i>Current Contents</i>	
Energie-Environnement	36	14	38,89
STIC	118	42	35,59
Mécanique	156	63	40,38
Matériaux	158	43	27,22
Biologie-Santé	52	3	5,77
Total	520	165	31,73

Nous constatons que le pôle Biologie-Santé est particulièrement bien représenté dans les *Current Contents* avec moins de 6% des articles absents de cette base. La Biologie-Santé fait partie des sciences dites "fondamentales", au même titre que la physique, la chimie. Ce résultat n'est donc pas surprenant car l'ISI cible, à travers le dépouillement des revues, ces sciences fondamentales.

Les domaines plus récents comme l'Energie-Environnement et les STIC sont moins bien représentés. Ces domaines sont appelés à se développer de façon importante dans l'avenir. Ils pourraient alors être plus présents dans les *Current Contents*.

Le pôle matériaux est assez bien représenté avec 27% des articles absents de la base LabINSA.

Enfin la mécanique est la moins bien représentée, avec 40 % d'articles absents de la base de l'ISI. La mécanique est un domaine hétérogène qui peut aller de la physique jusqu'aux sciences appliquées. Ces dernières ne sont pas considérées comme des sciences fondamentales par l'ISI.

Nos résultats sont également en accord avec ceux d'Adel Larouci (LAROUCI 2002). Ce-dernier a sélectionné un laboratoire de physique "pure" et un laboratoire de Sciences et Techniques pour l'Ingénieur (le nom des laboratoires n'est pas précisé). Se basant sur leurs rapports d'activités, il a recherché dans les *Current Contents* leurs publications pour la période 1999/2001. Il a retrouvé 94,5% des publications du laboratoire de physique et 87,5% des publications du laboratoire des Sciences et Techniques pour l'Ingénieur. Cela va dans le sens indiqué ci-

dessus ; l'ISI privilégierait les sciences fondamentales, comme la physique, au détriment des Sciences pour l'Ingénieur.

Concernant les test du Khi-deux, le test est non significatif pour la première hypothèse. Le statut inter-établissement d'un laboratoire n'influe pas sur le nombre d'articles présents dans les *Current Contents* ($\chi^2_{\text{cal}} = 1.39 < \chi_{\text{th}} = 6.63$ à 1%). En revanche le test est significatif pour la deuxième hypothèse. **Le statut EPST d'un laboratoire ESPT influe sur le nombre d'articles présents dans les *Current Contents* ($\chi^2_{\text{cal}} = 10.80 < \chi_{\text{th}} = 6.63$ à 1%).** Le détail des tests est présenté en annexe 12.

L'écart à l'indépendance et le Khi-deux par case (voir annexe 12) vont dans le sens selon lequel les laboratoires EPST publieraient plus dans les revues dépouillées par l'ISI, que les autres laboratoires. Plusieurs hypothèses peuvent expliquer cela :

- Si l'on part du principe que les revues dépouillées par l'ISI représentent le "cœur" de la science mondiale, on peut estimer que les laboratoires EPST sont "meilleurs" que les autres. Nous pensons que cette affirmation triviale reste à démontrer (comment estimer qu'un chercheur est meilleur qu'un autre ?) et peut provoquer rapidement une vive polémique. Cependant elle n'est totalement pas dénuée de sens puisque que les EPST (CNRS, INSERM, INRA) sont sensés être le "fleuron" de la recherche française.
- Les laboratoires EPST peuvent avoir une politique directive plus forte envers leurs chercheurs pour publier dans les revues de l'ISI, puisque ces revues sont sensées être les plus prestigieuses.
- Enfin les domaines d'activités des EPST peuvent être mieux couverts par les revues dépouillées par l'ISI que les domaines des autres laboratoires. Par exemple pour des domaines comme l'informatique, il peut exister un décalage entre les revues faisant autorité dans le domaine et les revues dépouillées par l'ISI.

2.6.2 Analyse de la revue des articles du corpus D

Nous nous sommes tout d'abord intéressé aux revues dans lesquelles sont publiés les articles pour savoir si elles étaient dépouillées ou non par l'ISI. Nous avons recherché les articles dont le code périodique était renseigné. Sur les 165 articles,

nous avons obtenu 135 articles avec un code renseigné. Puis nous avons complété les ISSN manquants. Nous avons pu identifier la revue de 115 articles. Cela correspondait à 73 revues que nous avons regroupées dans une table.

Puis nous avons réalisé une différence symétrique entre la table des revues de l'ISI et celle des revues des articles "INSA" (voir requête SQL en annexe 10).

Sur les 73 revues, **44 seulement ne sont pas dépouillées par l'ISI**. Cela signifie que 29 revues obtenues à partir d'articles **absents** des *Current Contents* sont dépouillées par l'ISI. Ce résultat est surprenant.

Deux hypothèses sont possibles:

- Soit nous n'avons pas pu retrouver ces articles dans les *Current Contents*, pour une raison que nous ignorons.
- Soit l'ISI ne dépouille pas de façon exhaustive les revues, comme cela est pourtant indiqué.

Nous présentons en annexe 13 les 44 revues non dépouillées par l'ISI, et les 29 revues dépouillées mais pour lesquelles les articles du corpus D ne sont pas signalés par l'ISI.

Puis nous avons déterminé à combien d'articles correspondaient ces 44 revues. Nous avons obtenu 76 articles.

En conclusion, les 165 articles se répartissent comme suit :

- 30 articles n'ont pas de code périodique dans la base LabINSA et sont absents des *Current Contents* (vérification faite sur le titre et les auteurs).
- 20 articles dont la revue n'est pas identifiée (ISSN inconnu) sont absents des *Current Contents* (vérification faite sur le titre, les auteurs et le libellé du périodique saisi dans LabINSA).
- 76 articles ont une revue identifiée (ISSN connu), mais celle-ci n'est pas dépouillée par l'ISI.
- 39 articles ont une revue identifiée (ISSN connu) et **dépouillée par l'ISI** ; mais ces articles sont **absents** des *Current Contents* (vérification faite sur le titre et les auteurs). Nous avons présenté les références de ces 39 articles en annexe 14.

Ainsi 76% des articles (126 articles) sont absents des *Current Contents* de façon **justifiée** et 23% des articles (39 articles) le sont pour **une raison inconnue**.

Par ailleurs, nous avons réparti les 76 articles, correspondant aux 44 revues non dépouillées par l'ISI, selon les 5 pôles de compétences de l'INSA (voir tableau 12).

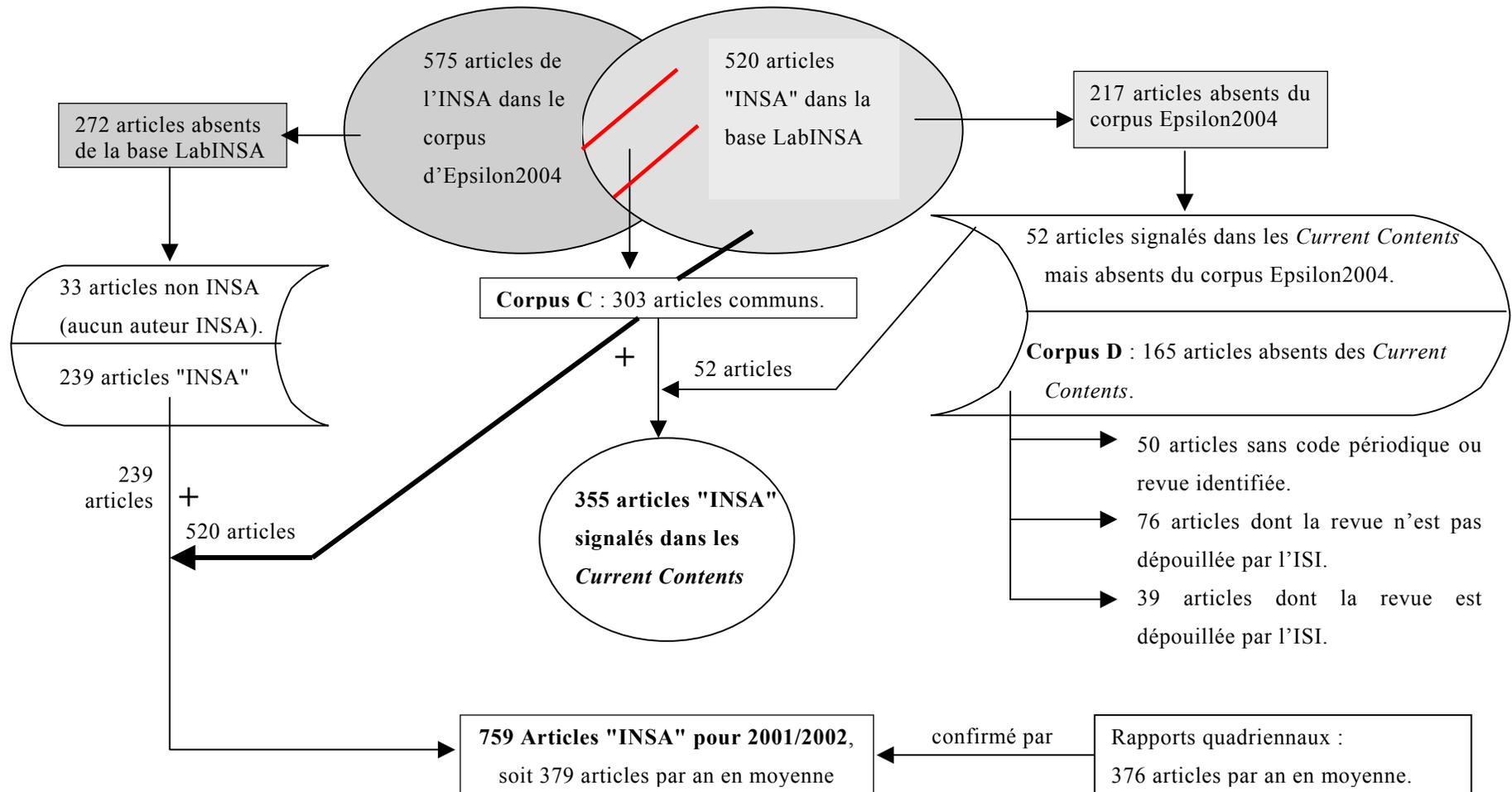
Tableau 12 : répartition des articles issus des revues non dépouillées par l'ISI.

Pôles de compétence	Nombre d'articles	Proportion %
Energie-Environnement	8	10,81
STIC	16	21,62
Mécanique	30	40,54
Matériaux	22	29,73
Biologie-Santé	0	0
Total	76	100

Nous constatons que deux tendances, mises en évidence lors de la répartition des articles "INSA" selon les pôles de compétences (voir tableau 11, page 52) sont confirmées. Le pôle Mécanique est le pôle dont le plus grand nombre d'articles (41%) sont absents des *Current Contents*. Le pôle Biologie-Santé n'en a aucun. En revanche le pôle Matériaux présente plus d'articles absents des *Current Contents* que les pôles Energie-environnement et STIC. Ce résultat est contraire au résultat obtenu dans le tableau 11.

Nous avons présenté ci-après un schéma récapitulatif de la comparaison des corpus de la base LabINSA et d'Epsilon2004 pour 2001/2002.

Schéma récapitulatif de la comparaison des corpus d'articles de périodiques scientifiques de la base LabINSA et d'Epsilon2004 pour 2001/2002



3. Propositions de solutions aux problèmes posés par la base LabINSA

Tout au long de notre étude, nous avons rencontré des difficultés avec la base LabINSA. Nous tenions à présenter quelques solutions pour améliorer l'utilisation de cette base. Il ne s'agit nullement de remettre en cause le travail effectué lors de la mise en place de cette base, mais plutôt d'en faciliter l'exploitation. De plus, nous ne prétendons pas avoir réponse à tous les problèmes, ni même avoir mis en lumière toutes les "imperfections". Nous entendons juste signaler les difficultés rencontrées et proposer des réponses.

Tout d'abord à propos de la base elle-même, la possibilité de doublons nous est apparue comme un problème majeur. Une même production scientifique peut être saisie plusieurs fois. Une clef de dédoublonnage basée sur le titre, l'année et le type de production pourrait être mise en place pour filtrer les productions au moment du transfert des données dans la base LabINSA.

Certaines tables, propres à chaque laboratoire pourraient devenir communes à tous. Nous pensons particulièrement à la table PERIODIQUE. Une table commune regrouperait tous les périodiques dans lesquels publient les laboratoires, avec la possibilité d'en saisir de nouveaux. Le champs de l'ISSN devrait être obligatoire, et ainsi, une clef de dédoublonnage sur l'ISSN permettrait d'éviter la saisie de doublons. Nous avons créé une table qui contient tous les libellés des revues indiquées dans la base LabINSA ; et nous avons commencé à compléter les ISSN. Cette table pourrait servir de base à la table PERIODIQUE de la base LabINSA.

Nous avons également rencontré des difficultés avec la table PERSONNEL, pour le personnel EPST. Le personnel EPST (CNRS, INSERM etc.) est bien désigné comme tel dans la table PERSONNEL. Mais pour un laboratoire inter-établissement, il est très difficile de savoir à quel établissement ce personnel est rattaché. Si ce laboratoire est affilié à deux établissements, il n'est pratiquement jamais fait mention, dans la base LabINSA, de l'établissement de rattachement de telle ou telle personne EPSTP. Certains chercheurs ont un code établissement 28 (le CNRS) mais rien n'indique leur rattachement à l'INSA ou à un autre

établissement. Pour palier à cela nous avons dû utiliser la liste du personnel EPST de l'INSA pour 2003/2004.

Dans la table PRODUCTION, le code périodique, permettant de faire le lien avec la revue est facultatif. Aussi, il est impossible de savoir dans quel périodique certaines productions ont été publiées. Nous pensons que ce code pourrait être rendu obligatoire pour rendre plus cohérent l'ensemble de la base LabINSA.

Concernant la saisie des données dans la base LabINSA, nous souhaitons aussi faire quelques remarques. L'INSA de Lyon compte 27 laboratoires ayant chacun une base de données LabINFO dans laquelle une personne du laboratoire saisit des données. Ces dernières sont ensuite rapatriées dans la base LabINSA. Au total 27 personnes différentes saisissent des données dans LabINSA ce qui nous semble beaucoup trop. A cela vient s'ajouter au fait que ces personnes ne sont pas des professionnels de la documentation.

D'où un manque de cohérence important de la base LabINSA en ce qui concerne plus particulièrement la production scientifique. Certains champs, tels le code périodique ou le code établissement, sont saisis par certains laboratoires, par d'autres non, par d'autres encore une fois sur deux etc. A cela s'ajoute les nombreuses erreurs de saisies : fautes d'orthographe, titres incomplets, libellés de périodique obsolètes etc. De plus le base LabINSA est très incomplète ; de nombreuses publications manquent dans la table PRODUCTION par exemple. Tout ceci concourt à faire de LabINSA une base très difficilement exploitable. Pour notre étude nous avons fait une comparaison manuelle des titres car automatiser la chose était impossible. Il nous semble presque indispensable de détacher un professionnel pour reprendre les données saisies et corriger les erreurs, si l'on veut aboutir, en exploitant cette base de données, à des résultats justes et publiables sur le Web ; ce que semblent souhaiter les laboratoires.

Enfin, nous avons eu le sentiment, déjà exprimé dans ce rapport, que les bases LabINFO et LabINSA **souffraient d'un manque d'image auprès des chercheurs**. Ceux-ci ont beaucoup de difficultés à apprécier leur utilité. Ils semblent plus envisager ces bases comme un moyen de contrôler et d'évaluer leurs travaux que comme un moyen de les valoriser. C'est peut être à ce niveau que se situe la clef du

problème. En effet si une communication plus large pouvait être faite, par exemple par la Direction de la Recherche et/ou DOC'INSA, auprès des laboratoires sur les aspects positifs de LabINSA, comme par exemple la possibilité de mettre en ligne les publications des laboratoires pour valoriser leurs recherches, les chercheurs pourraient y voir un intérêt. Ceci pourrait engendrer des moyens supplémentaires ainsi qu'une prise de conscience de la nécessité d'une saisie complète et juste dans les bases de données.

4. Les causes du manque de visibilité internationale de l'INSA de Lyon

Les deux causes principales qui nous sont apparues au cours de notre étude sont l'adressage des publications scientifiques et les domaines d'activités des laboratoires de l'INSA. A cette dernière cause est corrélée la sélection des revues que dépouille l'ISI.

Le problème lié à l'adressage est un problème récurrent qui a été soulevé par toutes les études bibliométriques antérieures. En 1999, Isabelle Marchand (MARCHAND 1999), qui travaillait sur l'évaluation de la production scientifique de l'INSA et de l'Université Claude Bernard, indiquait que la non normalisation des adresses posait beaucoup de problèmes pour collecter les références bibliographiques et les classer par laboratoire. Elle préconisait une sensibilisation des chercheurs à ce problème. Cette conclusion a été largement reprise dans les rapports d'Isabelle Bedague, Adel Larouci et Camille Prime-Claverie.

Isabelle Bedague (BEDAGUE 2001) a recensé pour l'année 1998, 13 manières d'affilier un article de périodique scientifique au laboratoire CREATIS. Par ailleurs, elle a signalé le problème de l'ordre d'écriture des adresses des établissements. Si l'auteur écrit le nom de l'EPST en premier, ce qui est souvent le cas, la publication est attribuée à l'EPST et non à l'établissement d'appartenance de l'auteur. Ainsi les publications sont attribués aux EPST au détriment des établissements d'appartenance. Parfois les chercheurs ne mentionnent même pas l'établissement d'appartenance mais uniquement l'EPST. Ainsi le poids du CNRS, par exemple, dans les publications françaises, est largement surévalué.

De son côté, Adel Larouci (LAROUCI 2002) a dénombré 101 dénominations de l'UCB dans les références bibliographiques. Camille Prime-Claverie (PRIME-CLAVERIE 2004) a dû utiliser 30 expressions régulières de l'UCB pour construire ses filtres informatiques. Camille Prime-Claverie précise également que ce genre de filtrage n'est réalisable qu'avec une très bonne connaissance de terrain. Une analyse de ce type n'est pas envisageable à l'échelle nationale ou internationale.

Ainsi une meilleure visibilité sur le plan international des établissements d'enseignements supérieurs (EES), tel l'INSA de Lyon, passe par une normalisation des adresses de ces établissements. En nous basant sur le travail de Camille Prime-Claverie et notre propre connaissance des bases de données, nous proposons ci-dessous une normalisation pour l'affiliation des auteurs avec dans l'ordre :

1. Le nom de l'établissement de rattachement : ne pas utiliser de sigle mais un libellé abrégé qu'il faut normaliser.
2. Le nom du laboratoire : utiliser un sigle caractéristique, sinon un libellé abrégé.
3. Le sigle de l'EPST et son numéro d'UMR, si le laboratoire est affilié EPST.
4. Le code postal, le nom de la ville et le nom du pays.

Il convient d'éviter les noms des établissements où se situe physiquement le laboratoire (exemple un hôpital abritant un laboratoire de l'UCB), pour ne préciser que l'établissement de rattachement.

Exemple 1 : Inst Natl Sci Appl, CREATIS, UMR CNRS 5515, F-69621 Villeurbanne France;

Exemple 2 : Inst Natl Sci Appl, Gen Elect Ferroelect Lab, F-69621 Villeurbanne, France.

La deuxième cause du manque de visibilité de l'INSA est le domaine d'activités de ses laboratoires. Pour bien comprendre cela, il faut tout d'abord connaître les biais de la base commerciale des *Current Contents*.

- Le principal indicateur de production scientifique est l'article de périodique scientifique, d'où notre choix et celui de Camille Prime-Claverie pour réaliser nos travaux. Mais les publications scientifiques de l'INSA ne se limitent pas aux articles. Les laboratoires publient des brevets, des actes de congrès, des ouvrages, des logiciels etc.

- L'ISI affirme indexer les revues jugées comme représentatives de l'évolution d'un domaine de recherche donné. Or il semble que les revues anglophones en général et américaines en particulier soient privilégiées. Une étude récente de Pascal Bador (BADOR & al. 2003) montre que sur un corpus de 289 revues européennes relatives à la pharmacie, seules 60 sont dépouillées par l'ISI. Par ailleurs, nous confirmons cela avec notre table des revues non dépouillées par l'ISI (voir en annexe 13). Sur 44 revues non dépouillées, 35 sont européennes.
- Enfin l'ISI semble privilégier les sciences définies comme "fondamentales" : la Physique, la Chimie, la Biologie, la Médecine etc. Par conséquent certains domaines sont favorisés par rapport à d'autres. La Mécanique et les Sciences pour l'Ingénieur ne sont pas jugées "fondamentales" par exemple.

Compte tenu des biais de l'ISI, le domaine d'activité des laboratoires de l'INSA peut expliquer son manque de visibilité. La mécanique, par exemple, est un domaine traditionnel de l'INSA de Lyon, très actif au niveau de la recherche. Pour 2001/2002, 30% des articles "INSA" se rapportent au pôle Mécanique ; or ce pôle a la plus forte proportion d'articles absents des *Current Contents* (40%). A l'inverse les articles du pôle Biologie-Santé, pôle qualifié "d'original" dans l'INSA, se retrouvent presque tous dans les *Current Contents*.

Il nous semble difficile de remédier à ce "problème". L'ISI ne semble pas considérer la mécanique comme une science fondamentale. Cela a pour corollaire un faible dépouillement des revues traitant de la mécanique. Sur les 74 articles absents des *Current Contents* à cause du non dépouillement de leur revue, 41% sont des articles se rapportant à la mécanique. A l'opposé, nous imaginons mal l'INSA modifier ses activités pour être plus visible dans les *Current Contents*. Il faut donc bien avoir conscience de ce biais, quasiment insoluble, lorsque l'on analyse la visibilité internationale de l'INSA via les *Current Contents*.

Le cas des pôles Energie-Environnement et Sciences et Technologies de l'Information et de la Communication (ou STIC) semble moins problématique. Ce sont des domaines récents qui sont appelés à se développer. Lorsqu'ils auront atteint une dimension internationale suffisante, on peut penser que l'ISI choisira alors de dépouiller plus de revues de ces domaines. Ainsi la visibilité de l'INSA augmentera dans les *Current Contents*.

Conclusion

L'objectif principal de notre travail était de déterminer les causes du manque de visibilité internationale des productions scientifiques de l'INSA de Lyon. Pour cela nous avons comparé les données interne de l'INSA avec les données recueillies sur les *Current Contents*.

Deux causes ont été clairement identifiées. La première cause est la non normalisation des adresses utilisées par les chercheurs pour affilier leurs publications. Ce problème peut être résolu en formulant des conseils de rédaction des adresses voire en les imposant quand cela est possible. Nous avons proposé un schéma de normalisation. Ce point fait d'ailleurs partie du projet Epsilon. Des propositions d'adressages seront faites établissements et une validation leur sera demandée. Cependant, ces consignes doivent être clairement expliquées. Trop de chercheurs n'ont pas encore conscience de l'intérêt de bien rédiger l'adressage et sont par la suite extrêmement surpris que l'on ne retrouve pas leurs publications. Nous pensons qu'un adressage normalisé et adopté par tous, améliorera grandement la visibilité de l'INSA de Lyon dans les bases de données internationales.

La deuxième cause est le domaine d'activités des laboratoires de l'INSA de Lyon. La Mécanique, très présente à l'INSA de Lyon, n'est pas considérée comme une science "fondamentale" par l'ISI. Cela a pour conséquence le dépouillement de peu de revues de mécanique. En revanche, la Biologie et la Santé, sciences fondamentales pour l'ISI, sont particulièrement bien représentées dans les *Current Contents*. Enfin certains domaines comme l'énergie, l'environnement et les Sciences et Technologies de l'Information de la Communication, encore récents, peuvent être appelés à être mieux représentés dans les *Current Contents*. Aussi nous pensons que la visibilité de l'INSA de Lyon s'améliorera aussi dans le futur, si l'ISI prend plus en considération les domaines Energie, Environnement et Sciences de l'Information. Cependant le biais dû à un domaine comme la mécanique doit être connu et accepté pour bien interpréter les résultats des études bibliométriques.

D'un point de vue plus global, les études bibliométriques menées sur l'INSA de Lyon constituent une excellente illustration des difficultés à analyser la production scientifique d'un établissement.

Nous avons pris conscience à travers notre propre étude mais aussi à la lecture des rapports antérieurs, de la difficulté de rendre compte, à l'aide d'un seul outil particulier (base de données, rapports d'activité), de la production scientifique d'un établissement de taille "moyenne" ; l'INSA de Lyon ne compte "que" 27 laboratoires. Un outil, telle qu'une base de données, doit être très bien conçue au départ sous peine de ne pas pouvoir être correctement utilisée. Nous avons vu que certaines tables de la base LabINSA, posait quelques difficultés d'exploitation.

De plus les personnes travaillant directement ou indirectement, avec cet outil doivent être convaincues de son utilité et conscientes de la nécessité de saisir une information juste. Les professionnels de la documentation ont un rôle à jouer à ce niveau. L'objectif des études bibliométriques n'est pas de juger ou de sanctionner de façon négative l'activité scientifique mais au contraire de décrire l'état de la science, d'aider à sa compréhension, de façon à la rendre plus efficace. L'exemple le plus flagrant est donné ici par l'adressage des publications. Publier un article très pertinent mais que l'on ne peut pas attribuer à l'établissement qui l'a produit ne sert pas son image.

Enfin si les études bibliométriques menées jusque là ont concerné au maximum 5 établissements du Pôle Universitaire Lyonnais (PRIME-CLAVÉRIE 2004), l'objectif est bien d'arriver à une vision plus large de l'activité scientifique. Un point de vue régional par exemple permettrait de replacer l'activité scientifique des établissements lyonnais dans un contexte mondial. Et pourquoi ne pas envisager une vision à l'échelle nationale de la Recherche. Ce sont les défis que devra relever la bibliométrie de demain, défis qui s'annoncent difficiles mais néanmoins passionnants.

Bibliographie par ordre alphabétique d'auteurs

Cette bibliographie a été rédigée à l'aide du guide de rédaction des références bibliographiques de DOC'INSA.

(<http://docinsa.insa-lyon.fr/docinsa/docs/refdocs.html>)

BADOR P., BEN ROMDHAME M., LAFOUGE T. *European pharmaceutical journals: Relationship between demand and indexation - the example of the main French document supplier.* The canadian journal of information and library Science, 2003, Vol 27 (2), pp. 33-55.

Bibliométrie : mode d'emploi pour un enjeu majeur. Vie Universitaire, avril 2004, n°71, pp. 26-28.

BEDAGUE Isabelle. *Essai d'une construction de méthodologie d'évaluation de la production scientifique du PUL rapport de stage de DESS,* DESS Ingénierie Documentaire. Lyon : Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, 2001, 65 p.

INSA DE LYON. *INSA de Lyon* [en ligne]. Disponible sur <<http://www.insa-lyon.fr/>>. (Consulté le 06.09.2004).

INSA DE LYON. *INSAVALOR vous ouvre les portes de la Recherche à l'INSA de Lyon* [en ligne]. Disponible sur <<http://www.insavalor.fr/p/propos.php>>. (Consulté le 06.09.2004).

INSTITUTE OF HIGHER EDUCATION? SHANGHAI JIAO TONG UNIVERSITY. *Academic Ranking of World Universities-2004*[en ligne]. Disponible sur <<http://ed.sjtu.edu.cn/ranking.htm>>. (Consulté le 16/09/2004).

LAROUCI Adel. *Identification de la production scientifique du pôle universitaire lyonnais : sciences physiques : rapport de stage,* licence. Lyon : I.U.P. de Documentation et Information Scientifiques et Techniques, 2002, 36 p.

MARCHAND Séverine. *Evaluation de la production scientifique de UCBL/INSA : rapport de stage de DESS,* DESS Ingénierie Documentaire. Lyon : Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, 1999, 70 p.

OBSERVATOIRE DES SCIENCES ET DES TECHNOLOGIES. *OST* [en ligne]. Disponible <<http://www.obs-ost.fr/fr/>> (Consulté le 15/09/2004).

PRIME-CLAVERIE Camille. *Indicateurs bibliométriques de la production des établissements scientifiques du PUL : rapport confidentiel Epsilon 4.* Lyon, 2004 36 p.

ROSTAING Hervé. *La bibliométrie et ses techniques.* Toulouse, Marseille : Sciences de la Société (Toulouse), Centre de Recherche Rétrospective de Marseille, 1996, 131 p.

Bibliographie par rubrique

Cette bibliographie a été rédigée à l'aide du guide de rédaction des références bibliographiques de DOC'INSA.

(<http://docinsa.insa-lyon.fr/docinsa/docs/refdocs.html>)

Articles de périodiques

BADOR P., BEN ROMDHAME M., LAFOUGE T. *European pharmaceutical journals: Relationship between demand and indexation - the example of the main French document supplier.* The canadian journal of information and library Science, 2003, Vol 27 (2), pp33-55.

Bibliométrie : mode d'emploi pour un enjeu majeur. Vie Universitaire, avril 2004, n°71, pp. 26-28.

Ouvrage

ROSTAING Hervé. *La bibliométrie et ses techniques.* Toulouse, Marseille : Sciences de la Société (Toulouse), Centre de Recherche Rétrospective de Marseille, 1996, 131 p.

Rapports

PRIME-CLAVERIE Camille. *Indicateurs bibliométriques de la production des établissements scientifiques du PUL : rapport confidentiel Epsilon 4.* Lyon, 2004 36 p.

Rapports de stage

BEDAGUE Isabelle. *Essai d'une construction de méthodologie d'évaluation de la production scientifique du PUL rapport de stage de DESS,* DESS Ingénierie Documentaire. Lyon : Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, 2001, 65 p.

LAROUCI Adel. *Identification de la production scientifique du pôle universitaire lyonnais : sciences physiques : rapport de stage, licence.* Lyon : I.U.P. de Documentation et Information Scientifiques et Techniques, 2002, 36 p.

MARCHAND Séverine. *Evaluation de la production scientifique de UCBL/INSA : rapport de stage de DESS,* DESS Ingénierie Documentaire. Lyon : Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, 1999, 70 p.

Sites Internet

INSA DE LYON. *INSA de Lyon* [en ligne]. Disponible sur <<http://www.insa-lyon.fr/>>. (Consulté le 06.09.2004).

INSA DE LYON. *INSAVALOR vous ouvre les portes de la Recherche à l'INSA de Lyon* [en ligne]. Disponible sur <<http://www.insavalor.fr/p/propos.php>>. (Consulté le 06.09.2004).

INSTITUTE OF HIGHER EDUCATION? SHANGHAI JIAO TONG UNIVERSITY. *Academic Ranking of World Universities-2004*[en ligne]. Disponible sur <<http://ed.sjtu.edu.cn/ranking.htm>>. (Consulté le 16/09/2004).

OBSERVATOIRE DES SCIENCES ET DES TECHNOLOGIES. *OST* [en ligne]. Disponible <<http://www.obs-ost.fr/fr/>> (Consulté le 15/09/2004).

Table des annexes

ANNEXE 1 PRÉSENTATION DU PÔLE UNIVERSITAIRE LYONNAIS	I
ANNEXE 2 DESCRIPTIF DES LABORATOIRES DE L'INSA	IV
ANNEXE 3 EXEMPLE DE NOTICE BIBLIOGRAPHIQUE ISSUE DES <i>CURRENT CONTENTS</i>	VI
ANNEXE 4 L'OBSERVATOIRE DES SCIENCES ET DES TECHNOLOGIES (OU OST)	VII
ANNEXE 5 LES ÉQUATIONS DE RECHERCHES ET LES FILTRES INFORMATIQUES DE L'ÉTUDE EPSILON2004	X
ANNEXE 6 MODÈLE CONCEPTUEL DE DONNÉES DES BASES LABINFO	XV
ANNEXE 7 LISTE DES TABLES DE LA BASE LABINSA	XVI
ANNEXE 8 MODIFICATIONS DE LA BASE LABINSA	XVII
ANNEXE 9 CRÉATION DE LA TABLE "REVUESISI"	XXV
ANNEXE 10 LES REQUÊTES EN MODE SQL	XXVI
ANNEXE 11 ARTICLES "INSA" PRÉSENTS DANS LES CURRENT CONTENTS, MAIS ABSENTS D'EPSILON2004.	XXX
ANNEXE 12 LE TEST DU KHI-DEUX	XXXII
ANNEXE 13 LES REVUES DÉPOUILLÉES OU NON PAR L'ISI	XXXV
ANNEXE 14 LES ARTICLES ABSENTS DE FAÇON INJUSTIFIÉE DES <i>CURRENTS CONTENTS</i>	XXXVIII

Annexe 1 présentation du Pôle Universitaire Lyonnais



GIP POLE UNIVERSITAIRE DE LYON (PUL)

MISSIONS DU POLE UNIVERSITAIRE DE LYON (www.pul.fr)

Le Pôle Universitaire de Lyon, créé en 1995, rassemble 16 établissements d'enseignement supérieur.

Il a pour objectif de promouvoir une politique de site et constitue une plate-forme de concertation qui initie et met en œuvre des projets communs pour :

Favoriser l'ouverture, la politique d'accueil et le rayonnement international de l'Université de Lyon:

- Centre de mobilité et accueil spécifique des chercheurs et étudiants étrangers
- Collège Doctoral International.
- Centre d'Information International.
- Promotion de l'Université de Lyon.

Développer la politique de la recherche à Lyon, valoriser et insérer ses réalisations dans le développement économique local

- Edition de la revue Isotopes, l'actualité de la recherche et de sa valorisation à Lyon
- Coordination des structures de valorisation.
- Animation du site-web sur la recherche à Lyon : <http://www.lyon-sciences.prd.fr>.

Renforcer l'information et la qualité de vie pour les étudiants

Com'Et : un lieu d'information étudiante ouvert toute l'année.

Espace Multiservices Etudiants : l'accueil des nouveaux arrivants pendant la rentrée universitaire.

Editions de guides d'information.

Animations, offres culturelles pour les étudiants.

Développer la communication et la culture scientifique

Centre de Culture Scientifique et Technique : coordination de la Fête de la Science dans le Rhône, expositions, spectacles, Nuit du Cinéma Scientifique, valorisation d'expériences, formations à la médiation scientifique, organisation de colloques, conférences.

Accompagner le développement des nouvelles technologies pour l'enseignement

Développement du réseau informatique à haut débit sur l'agglomération lyonnaise.

Coordination et valorisation des actions TICE * dans les établissements (ANCOLY:

Animation du Campus Ouvert de Lyon).

* (Technologies d'Information et de Communication pour l'Enseignement)

ETABLISSEMENTS MEMBRES

Université Claude Bernard Lyon 1

Université Lumière Lyon 2

Université Jean moulin Lyon 3

Ecole Centrale de Lyon

Ecole Normale Supérieure de Lyon

Ecole Normale Supérieure Lettres et Sciences Humaines

Institut National des Sciences Appliquées de Lyon

Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques

Ecole Nationale Vétérinaire de Lyon

EM Lyon

Université Catholique de Lyon

Institut Polytechnique de Lyon

Institut National de Recherche Pédagogique

Institut Universitaire de Formation des Maîtres de Lyon

Ecole Nationale Supérieure des Arts et Techniques du Théâtre

Ecole Nationale des Travaux Publics de l'État (membre associé)

PARTENAIRES

Ville de Lyon

Grand Lyon

Conseil Régional Rhône-Alpes

Conseil Général du Rhône

CROUS de Lyon Saint-Etienne

CHIFFRES-CLES

1 250 000 habitants dans la communauté urbaine de Lyon
95 000 étudiants dans le PUL (sur 115 000 au total) dont 15 000 en 3^{ème} cycle et 12 000 étudiants étrangers
9 500 personnels permanents dont 5 500 enseignants-chercheurs et enseignants
510 laboratoires
800 thèses soutenues chaque année

COORDONNEES

Président : Philippe GILLET
Directeur : Guy BERTHOLON
Adresse : « La Pagode », Domaine Scientifique de la Doua, 43, boulevard du 11 Novembre 1918,
69622 Villeurbanne Cedex
Adresse sur le Campus : 38, avenue Niels Bohr
Tel : 04 72 43 26 42
Fax : 04 72 43 26 40
E mail : pul@univ-lyon.fr

Toutes les informations concernant le PUL et ses membres sont disponibles sur le site
www.pul.fr

Annexe 2 Descriptif des laboratoires de l'INSA

Sigle	Libellé complet du Laboratoire	Type	laboratoires inter-établissements	Nom du directeur	Contacts LabINFO (04 72 43)
BF2I	Biologie Fonctionnelle, Insectes et Interactions	UMR INRA 0203		Gérard FEBVAY	Jacqueline JOLY (83 56)
CEGELY	Centre de génie électrique de Lyon	UMR CNRS 5005	OUI	Laurent NICOLAS	Sandrine VALLET (82 38)
CETHIL	Centre de Thermique de Lyon	UMR CNRS 5008	OUI	Dany ESCUDIE	Agnès BAILLY (88 10)
CITI	Centre d'Innovation en Télécommunications et Intégration de services			Stéphane UBEDA	Mallou BENGOUA (64 15)
CNDRI	Contrôle Non Destructif par Rayonnements Ionisants			Daniel BABOT	JM LETANG (82 13)
CREATIS	Centre de recherche et d'applications en traitement de l'image et du signal	UMR CNRS 5515	OUI	Isabelle MAGNIN	JP Roux (82 58 ou 87 85)
ESCHIL	Environnement, ville et société_ Eq. Développement Urbain	UMR CNRS 5600		Monique BOTTA-ZIMMERMANN	Joelle CHARNAY (63 09)
EVS/EDU	Équipe de Sciences Humaines de l'INSA de Lyon			Jean SOUBRIER	Françoise PICARD (84 71)
GEMPPM	Groupe d'études de métallurgie physique et de physique des matériaux	UMR CNRS 5510		Jean-Yves CAVAILLE	Antonia RICCOBENE (83 82)

ICTT	Interaction collaborative, Télé formation, Télé activités		OUI	Bertrand DAVID	Margarita RAIMBAUD (79 91)
LAEPSI	Analyse Environnementale des Procédés et des Systèmes Industriels			Pierre MOSZKOWICZ	Agnès BRUNELLA-THENET (84 30)
LAI	Automatique Industrielle			Hervé Tanneguy REDARCE	Marie-Guy MERCIER (Marie-Guy.Mercier@insa-lyon.fr)
LAMCOS (LMC + LMSo)	Laboratoire de Mécanique des Contacts et des Solides	UMR CNRS 5514		Alain COMBESCURE	Anne Marie COLIN (84 52)
LCO	Chimie Organique	UMR CNRS 5181	OUI	Alain DOUTHEAU	Christine MONTALTO
LDMS (LMSt + CASM)	Laboratoire de Dynamique des Machines et des Structures	UMR CNRS 5006		Regis DUFOUR	Pascale Payerne (82 48)
LGEF	Génie Électrique et Ferroélectricité			Daniel GUYOMAR	Patricia LANGELOT (79 53)
LIRIS (LISI + RFV)	Laboratoire d'Informatique en Images et Systèmes d'Information		OUI	Bernard PEROCHE	Verdier (64 29)
LMFA	Mécanique des Fluides et d'Acoustique		OUI	Michel LANCE	Laurence Lopez (87 92)
LMM/IMP	Laboratoire des Matériaux Macromoléculaires/Ingénierie des matériaux polymères	UMR CNRS 5627	OUI	Jean-François GERARD	Isabelle POLO (89 79)
LPCI	Physico-Chimie Industrielle			Jean-Pierre MILLET	Elisabeth ECCIDIO (83 78)
LPM	Laboratoire de physique de la matière	UMR CNRS 5511		Gerard GUILLOT	Celine VINCENT (60 79)
LVA	Vibrations et Acoustique			Jean-Louis GUYADER	Charles Pezerat (62 15)
MAPLY	Laboratoire de mathématiques appliquées de Lyon	UMR CNRS 5585	OUI	Guy BAYADA	Maud LOUAIL (79 10)
PLM	Physiopathologie des Lipides et Membranes	UMR INSERM 585		Michel LAGARDE	Véronique DESCHAMPS (85 70)
PRISMA	Productique et Informatique des Systèmes Manufacturiers		OUI	Joel FAVREL	Nadira MATAR (82 19)
UMG	Microbiologie et génétique	UMR CNRS 5122	OUI	Nicole COTTE-PATTAT	Nadia GOUNI (04/72/43/15/53)
URGC	Unité de recherche en Génie Civil			Bernard CHOCAT	Valérie ORHON (60 93)

Annexe 3 exemple de notice bibliographique issue des *Current Contents*

L'exemple donné ci-dessous correspond à la production n°368 du laboratoire GEMPPM. On peut constater que le champs "institution" (ou IN) est multi-valué et contient les adresses de tous les auteurs.

Notice 1 de 1 dans CC Search(R) 7 Editions 2002

AN: 0001765905-0002See Contents-Page
 PT: Journal
 TI: Nanostructure, and molecular dynamics in rodlike
 polyimide/flexible-chain polyimide molecular composites
 AU: Bershtein-VA; Egorova-LM; Yakushev-PN; Meszaros-O; Sysel-P;
 David-L; Kanapitsas-A; Pissis-P
**IN: Russian Acad Sci, AF Ioffe Physicotech Inst, St Petersburg
 194021, Russia; Inst Chem Technol, Dept Polymers, CR-16628 Prague 6,
 Czech Republic; Inst Natl Sci Appl, F-69621 Villeurbanne, France; Natl
 Tech Univ Athens, Dept Phys, GR-15773 Athens, Greece**
 SO: JOURNAL-OF-MACROMOLECULAR-SCIENCE-PHYSICS. 2002; B41 (3) : 419-
 450
 DT: Article
 PY: 2002
 IS: 0022-2348
 LA: English
 JS: ORGANIC-CHEMISTRY/POLYMER-SCIENCE
 CC: Physical-Chemical-and-Earth-Sciences

Annexe 4 l'Observatoire des Sciences et des Technologies (ou OST)

Présentation



L'Observatoire des Sciences et des Techniques (OST) est un Groupement d'Intérêt Public (GIP) créé pour une durée de six ans par arrêté du 28 mars 1990, renouvelé le 28 mai 1996 pour six ans. Il vient d'être renouvelé pour 12 ans à dater du 13 avril 2002.

Il est constitué d'un Conseil d'administration de 13 membres, présidé par M. Jean-Jacques Duby et d'un Comité scientifique et de prospective, présidé par M. Emmanuel Jolivet.

Laurence Esterle, sa Directrice, Ghislaine Filliatreau, sa Directrice adjointe et l'équipe de l'OST composée d'une quinzaine de personnes, mènent à bien les missions, projets et études qui lui sont confiés, notamment par ses membres, et qui constituent chaque année le programme de travail de l'OST.

En février 2002, l'OST a publié son rapport Science & Technologie - Indicateurs Edition 2002 (éditions Economica, Paris). Les publications de l'OST comprennent aussi des ouvrages, des études, une lettre d'information ainsi que des articles qui sont publiés dans des revues scientifiques, professionnelles ou grand public.

L'OST produit des indicateurs qui permettent aux acteurs de la science et de la technologie au sens large de débattre de la politique de recherche et d'élaborer leur stratégie. Ses travaux sont destinés aux responsables des institutions publiques et privées de recherche, et plus généralement à ceux qui s'intéressent à la recherche, à la technologie et à l'innovation, aux échelles nationales, régionales, locales, européennes et internationales.

Les travaux de l'OST contribuent à apporter un éclairage sur les politiques de la recherche. Ils concernent tous ceux, chercheurs ou non, qui s'intéressent à l'avenir des sciences et des techniques.

Les Instances de l'OST

L'Observatoire des Sciences et des Techniques, Groupement d'intérêt public (GIP), est constitué de 13 membres :

Les ministères :

- de la Recherche.
- de la Défense.
- de l'Industrie.
- des Affaires étrangères.
- de l'équipement.

Des organismes publics de recherche :

- le Centre National de la Recherche Scientifique (CNRS).
- le Commissariat à l'Energie Atomique (CEA).
- le Centre National d'Etudes Spatiales (CNES).
- l'Institut National de la Santé et de la Recherche Médicale (INSERM).
- l'Institut National de la Recherche Agronomique (INRA).
- l'Institut de Recherche pour le Développement (IRD).
- le Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD)

L'Association Nationale de la Recherche Technique (ANRT) représentant les entreprises.

Les institutions membres de l'OST définissent les orientations de ses travaux et lui fournissent les ressources humaines et financières nécessaires. Chaque institution est représentée au Conseil d'administration de l'OST. Le Président du Conseil d'administration est M. Jean-Jacques Duby, directeur général de l'Ecole Supérieure d'Electricité à Gif-sur-Yvette (91).

L'OST est orienté par un Conseil scientifique et de prospective (CSP), constitué de treize membres et présidé par M. Emmanuel Jolivet, directeur du Centre - Inra, de Jouy-en-Josas (78). Ce comité exprime un avis sur la qualité scientifique des projets du GIP ; il conduit une réflexion prospective dans le cadre de la préparation du programme de travail de l'OST sur la production, le développement et l'utilisation des indicateurs.

Les coordonnées de l'OST

Adresse : Observatoire des Sciences et des Techniques

93, rue Vaugirard 75006 Paris

Téléphone : 01 42 22 30 30

Fax : 01 45 48 63 94

Site Internet : <http://www.obs-ost.fr/fr/>

Annexe 5 les équations de recherches et les filtres informatiques de l'étude Epsilon2004

Les équation 1 et 2 ont permis à Camille Prime-Claverie de constituer deux corpus de références bibliographiques.

Equation 1 :

Interrogation des CC sur les années 2000-2003 avec l'équation 1.

```
( ( ( insa or (inst* nat* sci* appl*) ) and ( lyon or villeurbanne ) )
or
( ( ucbl or ucb or (univ* claudes bernard) or doua or univ* or (obs*
lyon) ) and (lyon or villeurbanne or bron or (pierre benite) or (bourg
en bresse) or ( genis laval) ) ) or
( ( ( ( ctr hosp) or chu or hop*) and (lyon or oullins) ) or (lyon
sud) ) or ( ( (fac and (med or pharm)) or (grange blanche) or
(rockefeller) or (laennec) or (med sch) ) and (lyon or villeurbanne or
oullins) ) ) or
( ( (ecole cent*) or ecl) and (lyon or ecully) ) or
( ((ecole norm* super*) or ens) and lyon ) or
( (ecole vet*) and lyon ) ) in ad
```

Résultat : 12 888 références trouvées.

Equation 2 :

Interrogation des CC sur les années 2000-2003 avec l'équation 2 (par origine géographique dans le champ institution)

```
( (Lyon or Villeurbanne or (Pierre Benite) or Ecully or Oullins or
(Genis laval) or (Marcy and Etoile) or (Bourg en Bresse) or Bron or
(Seyne and Tamaris) ) and France) in ad
```

Résultat : 14 966 références trouvées.

Puis le champs "institution" de chaque référence a été éclaté pour séparer les différentes affiliations et deux filtres "informatiques" ont été mis au point pour affecter les références à chaque établissement.

Filtre 1

```
#!/usr/local/bin/perl

#programme qui attribue (si filtrage positif) un établissement aux
libellés d'institution

open (SORTIE, '>relations_institution_etablissement_PUL1.txt');
open (INDET, '>table-et-ind.txt');
open (DATA, 'table-institution-PUL.txt'); #liste des "institutions
lyonnaises" id_institution \t libelle_institution
    $ligne = <DATA>;
    chop($ligne);
    $count=0;
    while ($ligne ne "")
    {
        @tab_ligne=split(/\t/, $ligne); #découpe la ligne suivant les
        tabulations

        #filtre ECL
        if (
            ($tab_ligne[1] =~ /ecole\s*cent\s*lyon/i)
            or ($tab_ligne[1] =~ /ECL/)
            or ($tab_ligne[1] =~ /centrale\s*lyon/i)
            or ($tab_ligne[1] =~ /ecole\s*cent/i)
        )
        {
            print SORTIE ( $tab_ligne[0], "\t", "ECL\n");
            $count++;
        }

        #filtre INSA

        if (
            ($tab_ligne[1] =~ /INSA/)
            or ($tab_ligne[1] =~ /Insa/)
            or ($tab_ligne[1] =~ /Inst.*Nat.*Sci.*Appl/i)
            or ($tab_ligne[1] =~ /Nat.*Inst.*Appl.*Sci/i)
            #or ($tab_ligne[1] =~ /Insavalor\s/i)
        )
        {
            print SORTIE ( $tab_ligne[0], "\t", "INSA\n");
            $count++;
        }

        #filtre ENS
        if (
            (($tab_ligne[1] =~ /ENS/) && !( $tab_ligne[1] =~
            /Lettres/i))
            or (( $tab_ligne[1] =~ /Ecole\s*norm.*Super.*i) && !(
            $tab_ligne[1] =~ /Lettres/i))
        )
        {
```

```

        print SORTIE ( $stab_ligne[0], "\t", "ENS\n");
        $count++;
    }

#filtre Veto
if (
    ($stab_ligne[1] =~ /Ecol.*Nat.*Vet./i)
or ($stab_ligne[1] =~ /ENVL/)
or ($stab_ligne[1] =~ /Ecol.*Vet./i)
)
{
    print SORTIE ( $stab_ligne[0], "\t", "ENVL\n");
    $count++;
}

#filtre Lyon1
if (
    ($stab_ligne[1] =~ /UCB/)
or ($stab_ligne[1] =~ /Lyon\s*1/i)
or ($stab_ligne[1] =~ /Univ\s*C.*Bernard/i)
or ($stab_ligne[1] =~ /Univ\s*Villeurbanne/i)
or ($stab_ligne[1] =~ /CRAL/i)
or ($stab_ligne[1] =~ /Ctr.*Rech.*Astron.*Lyon/i)
or ($stab_ligne[1] =~ /Observ/i)
or ($stab_ligne[1] =~ /ISTIL/i)
#Medecine
or ($stab_ligne[1] =~ /CHU/)
or ($stab_ligne[1] =~ /Lyon\s*sud/i)
or ($stab_ligne[1] =~ /fac\s*Med/i)
or ($stab_ligne[1] =~ /fac\s*Pharm/i)
or ($stab_ligne[1] =~ /Rockefeller/i)
or ($stab_ligne[1] =~ /Grange\s*Blanche/i)
or ($stab_ligne[1] =~ /Laenn./i)
or ($stab_ligne[1] =~ /fac\s*Med/i)
or ($stab_ligne[1] =~ /Inst.*Cognit.*Sci/i)
#Hopitaux
or (
    ( ($stab_ligne[1] =~ /Hop/) or ($stab_ligne[1] =~ /Hosp/) )
    and
    ( ($stab_ligne[1] =~ /Herriot/i)
      or ($stab_ligne[1] =~ /Pradel/i)
      or ($stab_ligne[1] =~ /Wertheimer/i)
      or ($stab_ligne[1] =~ /Lyon/)
      or ($stab_ligne[1] =~ /Antiquaille/)
      or ($stab_ligne[1] =~ /Debrousse/)
      or ($stab_ligne[1] =~ /Croix.*Rousse/i)
      or ($stab_ligne[1] =~ /Hotel\s*Dieu/i)
      or ($stab_ligne[1] =~ /Debrousse/)
      or ($stab_ligne[1] =~ /Neurol.*Cardiol./i)
      or ($stab_ligne[1] =~ /Cardio.*Vasc/i)
      or ($stab_ligne[1] =~ /Neurol.*Cardiol./i)
      or ($stab_ligne[1] =~ /civils/i)
    ) #fin and
    ) #fin or
) #fin if Lyon1
{
    print SORTIE ( $stab_ligne[0], "\t", "Lyon1\n");
    $count++;
}

```

```

    }

    if ($count == 0)
    {
        print INDET ( $ligne, "\n");
    }

$ligne=<DATA>;
$count=0;
chop($ligne);
}

close (DATA);
close (SORTIE);
close (SORTIE2);

```

Avec le corpus issu de l'équation 1 : affectation de 7560 libellés d'institution.

Avec le corpus issu de l'équation 2 : affectation de 7877 libellés d'institution.

Filtre 2 :

```

#!/usr/local/bin/perl

#programme qui attribue (si filtrage positif) un établissement aux
libelles d'institution lyonnaises non déterminées par le filtre 1

open (SORTIE, '>relations_institution_etablissement_PUL1.txt');
open (INDET, 'table-et-ind.txt') || die ( " fichier n'existe pas ");
open (INDET2, '>table-et-ind2.txt');

    $ligne_institution = <INDET>;
    chop($ligne_institution);
    while ($ligne_institution ne "")
    {
        $det=0;
        @tab_institution=split(/\t/, $ligne_institution); #decoupe la ligne
suivant les tabu
            #lire le fichier Labocnrs
            open (LABOCNRS, 'table-labocnrs.txt') || die ( "
fichier table-labo-cnrs n'existe pas ");
            $ligne_labocnrs = <LABOCNRS>;
            chop($ligne_labocnrs);
            while ($ligne_labocnrs ne "")
            {

@tab_labocnrs=split(/\t/, $ligne_labocnrs); #decoupe la ligne des
labos-cnrs suivant les tabu (etablissement ; type labo ; num ; nom ;
appellations)

@tab_appellation=split(/;/, $tab_labocnrs[4]);

                #tests sur les appellations
                $count=0;
                for ( $u=0; $u<=$#tab_appellation; $u++)
                {

```

```

                                if (
$stab_institution[1] =~ /$stab_appelation[$u]/)
                                {
                                    $count ++;
                                }
                                }

                                #filtre
                                if (
                                    ( $stab_institution[1] =~
/$stab_labocnrs[1]/i) and      ($stab_institution[1] =~
/$stab_labocnrs[2]/ ) )      or ( $stab_institution[1]
                                or ( $stab_institution[1]
                                    or ( $count > 0)
                                    )
                                    {
                                        print
SORTIE ( $stab_institution[0], "\t", "$stab_labocnrs[0]\n");
                                        $det=1;
                                    }

                                $ligne_labocnrs =
<LABOCNRS>;
                                chop($ligne_labocnrs);
                                } #--fin while----$ligne_labocnrs

                                close(LABOCNRS);
                                if ($det==0){
                                    print INDET2 ( "$ligne_institution\n");
                                }
                                $ligne_institution = <INDET>;
                                chop($ligne_institution);
                                }#-----fin while institution

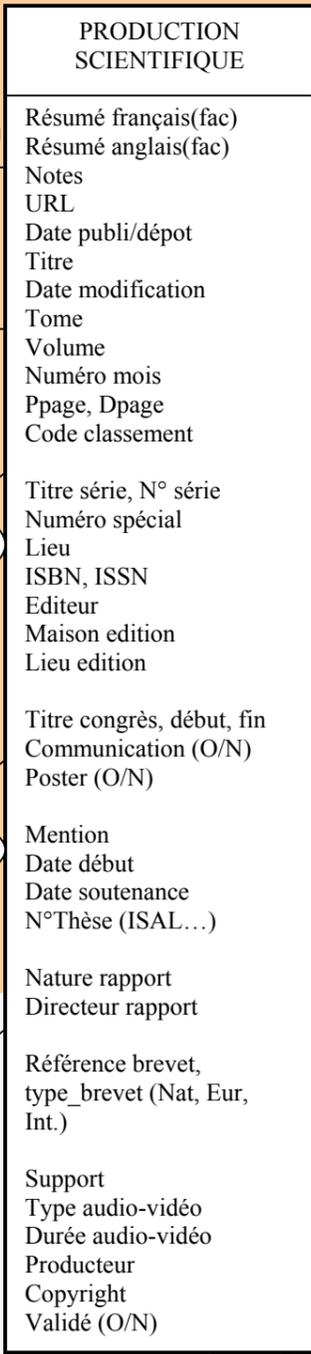
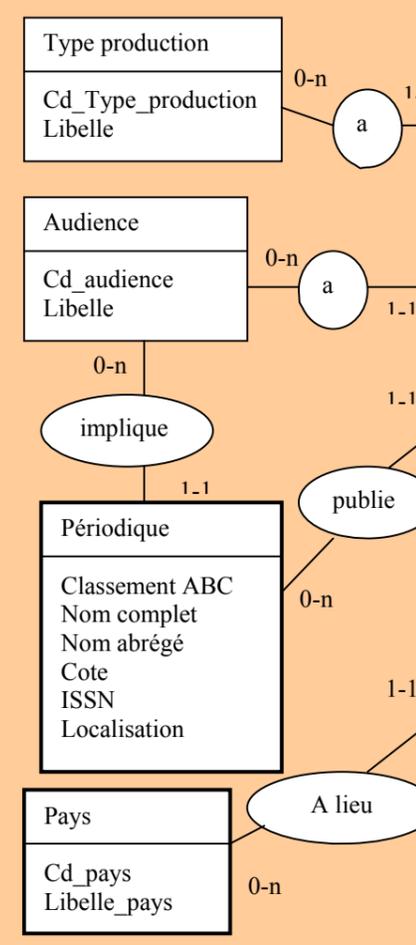
                                close (DATA);
                                close (INDET);
                                close (INDET2);

```

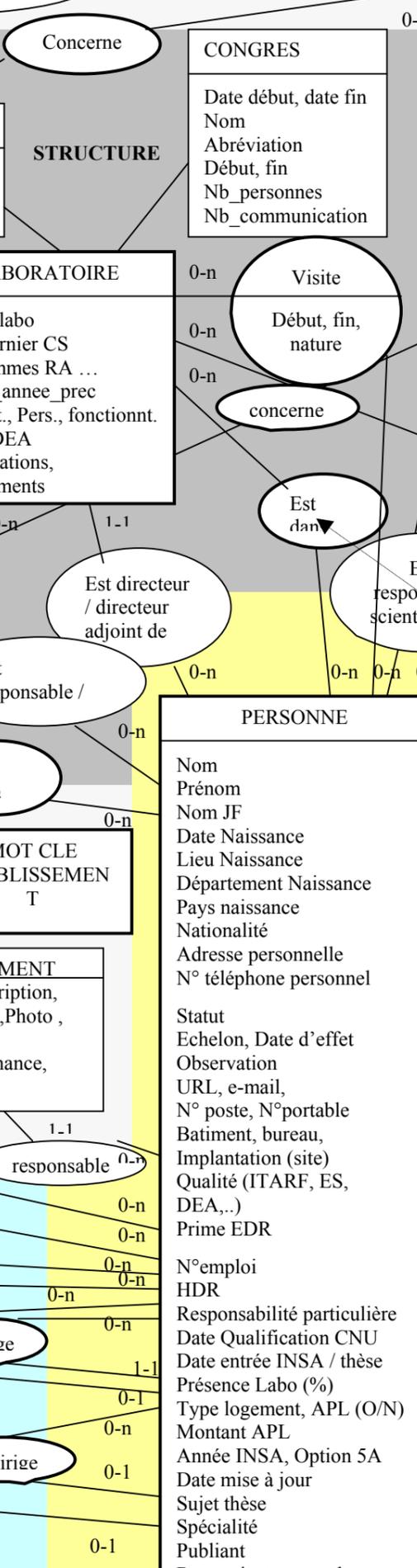
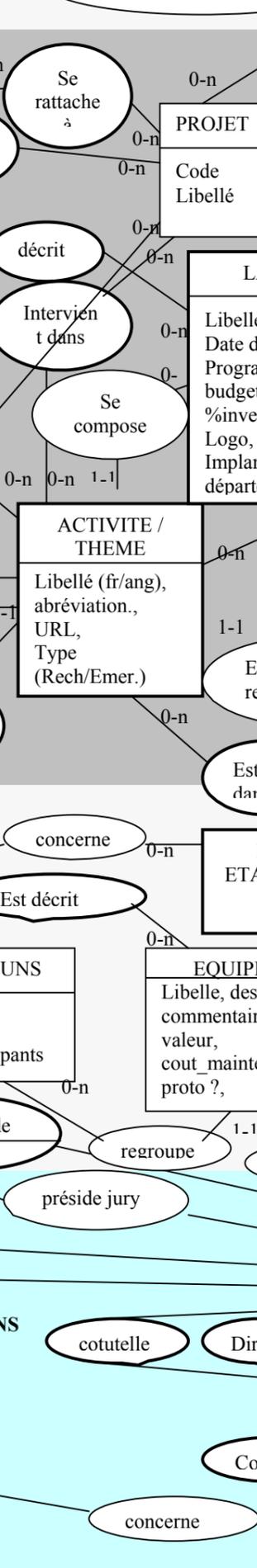
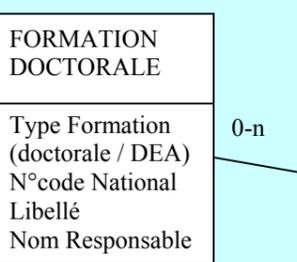
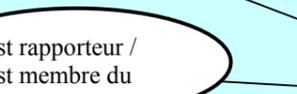
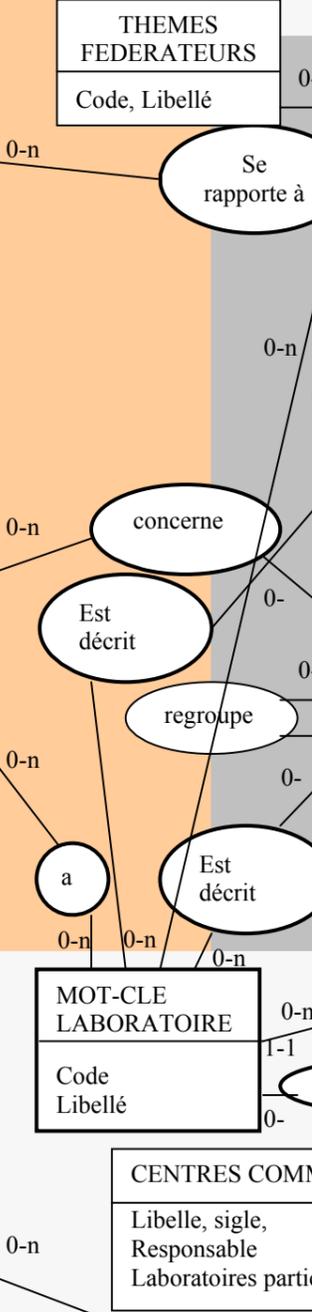
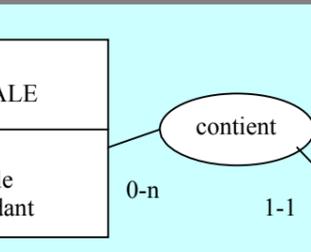
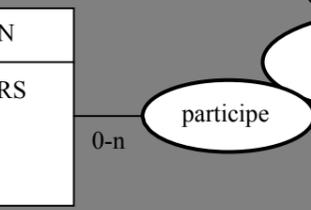
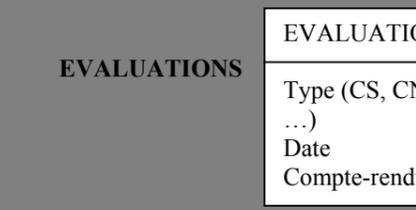
Avec le corpus issu de l'équation 1 : affectation de 387 libellés d'institution supplémentaires.
Avec le corpus issu de l'équation 2 : affectation de 477 libellés d'institution supplémentaires.

Annexe 6 Modèle conceptuel de données des bases LabINFO

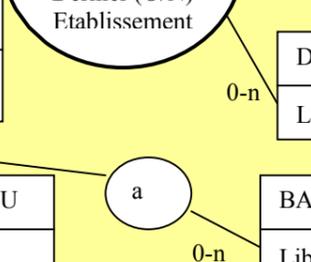
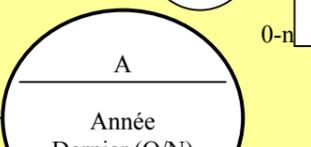
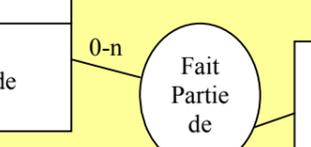
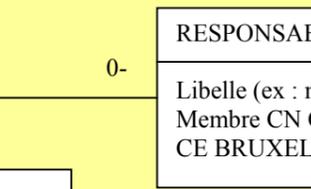
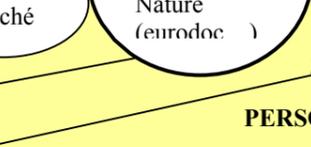
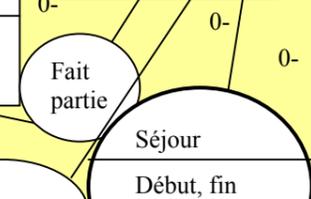
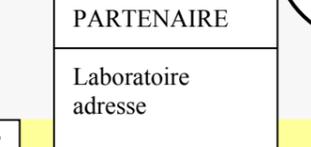
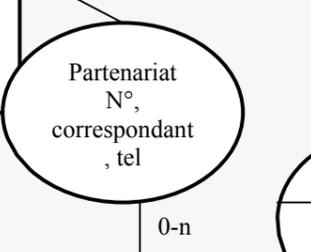
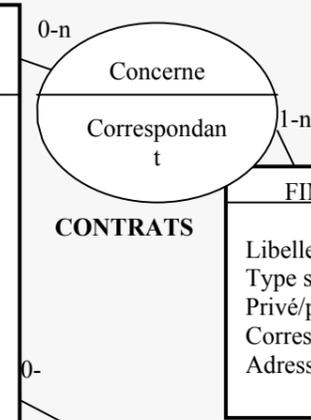
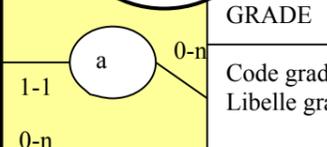
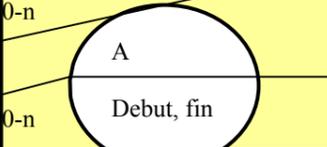
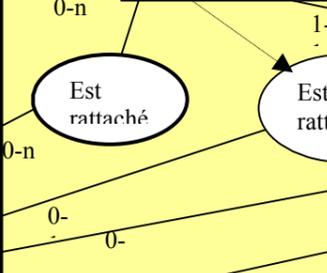
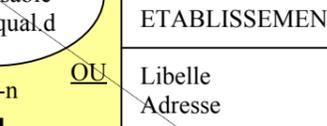
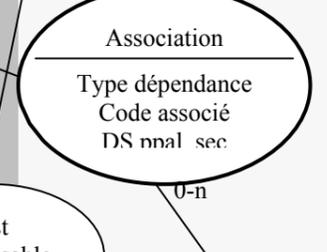
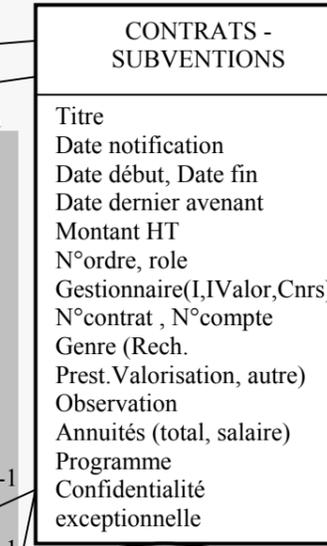
PRODUCTION SCIENTIFIQUE



MOTS-CLES



STRUCTURE



Annexe 7 liste des tables de la base LabINSA

Les tables dont le nom commence par I sont **Internes** au laboratoire les tables dont le nom commence par IC sont **Communes** à tous les laboratoires.

IC_AUDIENCE	IC_ROLE	I_FINANCE_PERSONNEL
IC_BAP	IC_SECTION	I_IMPLANTATION
IC_BAP_CNRS	IC_SUPPORT_AUDIO	I_LABORATOIRE
IC_CENTRE_EQUIPEMENT	IC_TYPE_ACTIVITE	I_LABORATOIRE_ECOLE
IC_CORPS	IC_TYPE_ETABLISSEMENT	I_LABORATOIRE_ETABLISSEMENT
IC_DIPLOME	IC_TYPE_FINANCEUR	I_LABORATOIRE_FORMATION
IC_DPT_SCIENTIFIQUE	IC_TYPE_PRODUCTION	I_LABORATOIRE_MOT_CLE_N2
IC_ECOLE_DOCTORALE	IC_TYPE_SEJOUR	I_MEMBRE_JURY
IC_ETABLISSEMENT		I_PARTENAIRE
IC_FORMATION_DOCTORALE	I_ACTIVITE	I_PERIODIQUE
IC_GENRE_CONTRAT	I_ACTIVITE_MOT_CLE_COMM	I_PERSHISTOQUALITE
IC_GRADE	I_ACTIVITE_MOT_CLE_N2	I_PERSONNEL
IC_GRAND_GROUPE	I_ACTIVITE_PROJET	I_PERSONNEL_DIPLOME
IC_MENTION	I_CONGRES	I_PERSONNEL_RESPONSABILITE
IC_MOT_CLE_COMM	I_CONTRAT	I_PERSONNEL_SEJOUR
IC_MOT_CLE_N1	I_CONTRAT_ANNUITE	I_PERSONNEL_VISITE
IC_MOT_CLE_N2	I_CONTRAT_FINANCEUR	I_PRODUCTION
IC_NATURE_VISITE	I_CONTRAT_PARTENAIRE	I_PRODUCTION2
IC_PAYS	I_CONTRAT_PROJET	I_PRODUCTION_AUTEUR
IC_QUALITE	I_DOCTORANT_ANNEE	I_PRODUCTION_PROJET
IC_RESPONSABILITE	I_EQUIPEMENT	I_PROJET
IC_RESSOURCE	I_FINANCEUR	I_RAPPORTEUR_JURY

Annexe 8 Modifications de la base LabINSA

Tables alimentées à partir des bases LabINFO des laboratoires

(documents fournis par le C.R.I.)

I_IMPLANTATION

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD IMPLANTATION</u>	NUMBER(7)	Numéro de l'implantation
ADRESSE1	VARCHAR2(50)	Première ligne de l'adresse postale de l'implantation
ADRESSE2	VARCHAR2(50)	Deuxième ligne de l'adresse postale de l'implantation
ADRESSE3	VARCHAR2(50)	Troisième ligne de l'adresse postale de l'implantation
CD_POSTAL	VARCHAR2(9)	Code postal
VILLE	VARCHAR2(50)	Ville
NO_TELEPHONE	VARCHAR2(20)	Numéro de téléphone
NO_FAX	VARCHAR2(20)	Numéro de fax
EMAIL	VARCHAR2(50)	E-mail
IMPLANTATION_PPAL	NUMBER(1)	Ce site est-il l'implantation officiel du laboratoire ? (Oui/Non)
CD_RESPONSABLE (*)	NUMBER(7)	Numéro de dossier du responsable du site

I_LABORATOIRE

Nom du champ	Type de données	Description
--------------	-----------------	-------------

<u>CD LABORATOIRE</u>	VARCHAR2(10)	Sigle du laboratoire
LIBELLE_LABORATOIRE	VARCHAR2(80)	Libellé du laboratoire
CD_DIRECTEUR (*)	NUMBER(7)	Numéro de dossier du directeur du laboratoire
CD_DIRECTEUR2 (*)	NUMBER(7)	Numéro de dossier du deuxième directeur du laboratoire
CD_DIRECTEUR_ADJOINT (*)	NUMBER(7)	Numéro de dossier du directeur adjoint
CD_CO_DIRECTEUR (*)	NUMBER(7)	Numéro de dossier du codirecteur
DATE_DERNIER_CS	DATE	Date de la réunion du dernier Comité Scientifique de laboratoire
CD_ETABLISSEMENT (*)	NUMBER(7)	Numéro de l'établissement
DATDEB_VALIDITE	DATE	Date de début de validité des informations de l'enregistrement
DATFIN_VALIDITE	DATE	Date de fin de validité des informations de l'enregistrement
CD_SECRETAIRE (*)	NUMBER(7)	Numéro de dossier de la secrétaire

I_LABORATOIRE_ETABLISSEMENT

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD ETABLISSEMENT (*)</u>	NUMBER(7)	Numéro de l'établissement
TYPE_DEPENDANCE	VARCHAR2(15)	Type d'association du laboratoire vis à vis de l'organisme de tutelle (Tutelle, Rattachement, Affiliation)
CODE_ASSOCIE_EPST	VARCHAR2(20)	Code du laboratoire dans l'organisme d'association (ex: code EPST UMR..., EA...)
CD_DPT_SCIENTIFIQUE1 (*)	NUMBER(7)	Département scientifique principal dans l'organisme
CD_DPT_SCIENTIFIQUE2 (*)	NUMBER(7)	Département scientifique secondaire dans l'organisme
CD_CORRESPONDANT (*)	NUMBER(7)	Numéro de dossier du correspondant de la composante

I_MEMBRE_JURY (membres du jury d'une thèse, d'une habilitation ...)

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD DOSSIER (*)</u>	NUMBER(7)	Identifiant du personnel dans le laboratoire
<u>CD PRODUCTION (*)</u>	NUMBER(7)	Numéro de la production

I_PARTENAIRE

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire

<u>CD_PARTENAIRE</u>	NUMBER(7)	Identifiant du partenaire dans le laboratoire
CD_ETABLISSEMENT (*)	NUMBER(7)	Numéro de l'établissement du partenaire
ENTITE	VARCHAR2(50)	Nom de l'entité, du laboratoire, etc.

I_PERIODIQUE (périodiques utilisés)

<i>Nom du champ</i>	<i>Type de données</i>	<i>Description</i>
<u>CD_LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD_PERIODIQUE</u>	NUMBER(7)	Identifiant du périodique dans le laboratoire
LIBELLE_PERIODIQUE	VARCHAR2(255)	
ABREGE_PERIODIQUE	VARCHAR2(255)	
ISSN	VARCHAR2(14)	International Standard Serial Number du périodique
FACTEUR_IMPACT	FLOAT	Facteur d'impact

I_PERSONNEL

Nom du champ	Type de données	Description
<u>CD_LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD_DOSSIER</u>	NUMBER(7)	Numéro de dossier du personnel du laboratoire
CIVILITE	VARCHAR2(4)	Civilité de la personne
NOM	VARCHAR2(30)	Nom de la personne
PRENOM	VARCHAR2(30)	Prénom de la personne
DATE_NAISSANCE	DATE	Date de naissance
CD_IMPLANTATION (*)	NUMBER(7)	Code de l'implantation de la personne
CD_QUALITE (*)	VARCHAR2(2)	Code de la qualité de la personne
CD_GRADE (*)	VARCHAR2(3)	Code du grade de la personne
CD_SECTION (*)	VARCHAR2(2)	Numéro de section CNU
CD_ETABLISSEMENT (*)	NUMBER(7)	Code de l'établissement de rattachement
CD_DPT_SCIENTIFIQUE (*)	NUMBER(7)	Code du département scientifique pour les EPST
CD_PARTENAIRE (*)	NUMBER(7)	Code du partenaire pour les intervenants extérieurs
DATDEBUT_PRESENCE	DATE	Date de début de présence de la personne dans le laboratoire
DATFIN_PRESENCE	DATE	Date de fin de présence de la personne dans le laboratoire
HDR	NUMBER(1)	Habilitation à diriger des recherches (Oui/Non)

PRIME_EDR	NUMBER(1)	Prime d'encadrement doctoral (Oui/Non)
CD_BAP (*)	NUMBER(4)	Code BAP
CD_BAP_CNRS1 (*)	NUMBER(4)	Code BAP 1 pour les IT du CNRS
CD_BAP_CNRS2 (*)	NUMBER(4)	Code BAP 2 pour les IT du CNRS
PUBLIANT	NUMBER(1)	Est-ce que la personne publie ? (Oui/Non)
CD_DIRECTEUR (*)	NUMBER(7)	Numéro de dossier du directeur de la thèse
CD_CO_DIRECTEUR (*)	NUMBER(7)	Numéro de dossier du codirecteur de la thèse
CD_CO_DIRECTEUR2 (*)	NUMBER(7)	Numéro de dossier du 2 ^{ème} codirecteur de la thèse
PCT_DIRECTEUR	NUMBER(3)	Pourcentage du directeur dans la direction de la thèse
PCT_CO_DIRECTEUR	NUMBER(3)	Pourcentage du 1 ^{er} codirecteur dans la direction de la thèse
PCT_CO_DIRECTEUR2	NUMBER(3)	Pourcentage du 2 ^{ème} codirecteur dans la direction de la thèse
CD_DIRECTEUR_COTUTELLE (*)	NUMBER(7)	Numéro de dossier du directeur de cotutelle
CD_FORMATION (*)	VARCHAR2(6)	Code de formation pour les thésards et DEA
DATE_FORMATION	DATE	Date de début de la formation
ANNEE_INSA	NUMBER(1)	Année d'étude à l'INSA (ex: 5 pour 5ème année) ou de thèse pour un doctorant (1/2/3)
DERNIERE_SITUATION	VARCHAR2(100)	Dernière situation connue
NO_POSTE	VARCHAR2(30)	N° de téléphone de la personne
NO_FAX	VARCHAR2(30)	N° de télécopie de la personne
EMAIL	VARCHAR2(100)	E-Mail de la personne
INITIALES_PRENOM	VARCHAR2(6)	Initiales du prénom

I_PRODUCTION

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD PRODUCTION</u>	NUMBER(7)	Identifiant de la production scientifique dans le laboratoire
TITRE	VARCHAR2(255)	Titre de la production scientifique
ANNEE	VARCHAR2(4)	Année de la production scientifique
CD_TYPE_PROD (*)	VARCHAR2(2)	Code du type de production scientifique
DATE_SOUTENANCE	DATE	Date de soutenance du DEA ou de la thèse
CD_MENTION (*)	CHAR	Code de la mention du DEA ou de la thèse

CD_AUDIENCE (*)	CHAR	Code de l'audience de la production
TITRE_SERIE_OU_COLLECTION	VARCHAR2(100)	Pour les articles, livres, chapitres de livre ...
NO_SERIE	VARCHAR2(5)	Pour les articles, livres, chapitres de livre ...
TITRE_CONGRES	VARCHAR2(255)	Pour les congrès, communications
DATE_DEBUT_CONGRES	DATE	Pour les congrès, communications
DATE_FIN_CONGRES	DATE	Pour les congrès, communications
DATE_PRODUCTION	DATE	Date de la production
CD_PERIODIQUE (*)	NUMBER(7)	Code du périodique dans le laboratoire (pour un article par ex.)
NUMERO_PERIODIQUE	VARCHAR2(10)	Numéro du périodique
TOME	VARCHAR2(5)	Pour un livre, chapitre ...
VOLUME	VARCHAR2(10)	
MOIS	NUMBER(3)	
PPAGE	VARCHAR2(5)	
DPAGE	VARCHAR2(5)	
ISSN	VARCHAR2(14)	The International Standard Serial Number
ISBN	VARCHAR2(16)	The International Standard Book Number
NUMERO_SPECIAL	VARCHAR2(255)	
LIEU_EDITION	VARCHAR2(100)	
LIEU_CONGRES	VARCHAR2(100)	
CD_PAYS (*)	NUMBER(4)	
EDITEUR_SCIENTIFIQUE	VARCHAR2(255)	
EDITEUR_COMMERCIAL	VARCHAR2(50)	
COMMUNICATION	NUMBER(1)	-1 : une communication (0 : n'est pas une communication)
POSTER	NUMBER(1)	-1 : un poster (0 : n'est pas un poster)
CD_ETAB_ORIGINE (*)	NUMBER(7)	
TYPE_BREVET	NUMBER(1)	1: National 2: Européen 3: International
REF_BREVET	VARCHAR2(20)	
CD_SUPPORT (*)	VARCHAR2(1)	
PRODUCTEUR	VARCHAR2(50)	
COPYRIGHT	NUMBER(1)	-1 : avec copyright (0 : sans copyright)

<i>NATURE_RAPPORT</i>	<i>VARCHAR2(100)</i>	<i>Rapport intermédiaire;Rapport final</i>
<i>URL</i>	<i>VARCHAR2(100)</i>	<i>Url d'accès à la production</i>
<i>NUM_THESE</i>	<i>VARCHAR2(8)</i>	<i>Numéro de la thèse</i>
<i>BREVET_LICENCIE</i>	<i>NUMBER(1)</i>	<i>-1 : brevet avec licenece (0 : sans licence)</i>
<i>SYSTEME_EXPLOITATION</i>	<i>VARCHAR2(50)</i>	
<i>DATE_VERSION</i>	<i>DATE</i>	
<i>DERNIERE_VERSION</i>	<i>VARCHAR2(50)</i>	
<i>RESUME_FRANCAIS</i>	<i>VARCHAR2(2000)</i>	
<i>RESUME_ANGLAIS</i>	<i>VARCHAR2(2000)</i>	
<i>CD_PRESIDENT_JURY</i>	<i>NUMBER(7)</i>	<i>Identifiant du personnel dans le laboratoire</i>
<i>DUREE_SUPPORT</i>	<i>VARCHAR2(10)</i>	
<i>NB_ANNEES_CONFIDENTIALITE</i>	<i>FLOAT</i>	

I_PRODUCTION_AUTEUR

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD PRODUCTION (*)</u>	NUMBER(7)	Identifiant de la production scientifique dans le laboratoire
<u>CD DOSSIER (*)</u>	NUMBER(7)	Numéro de dossier de l'auteur
NO_ORDRE	NUMBER(2)	Numéro d'ordre de l'auteur dans la référence bibliographique (1 par défaut)
CONF_INVITE	NUMBER(1)	La participation à la conférence est-elle invitée ? (Oui/Non)

I_RAPPORTEUR_JURY (rapporteurs d'un jury d'une thèse, d'une habilitation ...)

Nom du champ	Type de données	Description
<u>CD LABORATOIRE (*)</u>	VARCHAR2(10)	Sigle du laboratoire
<u>CD DOSSIER (*)</u>	NUMBER(7)	Identifiant du personnel dans le laboratoire
<u>CD PRODUCTION (*)</u>	NUMBER(7)	Numéro de la production

Tables alimentées à partir de la base des références

(documents fournis par le C.R.I.)

IC_AUDIENCE

Nom du champ	Type de données	Description
<u>CD_AUDIENCE</u>	CHAR	Code de l'audience
LIBELLE_AUDIENCE	VARCHAR2(50)	Libellé de l'audience
CONGRES	NUMBER(1)	L'audience est-elle spécifique pour les congrès ? (Oui/Non)

IC_ETABLISSEMENT

Nom du champ	Type de données	Description
<u>CD_ETABLISSEMENT</u>	NUMBER(7)	Numéro de l'établissement
LIB_ETABLISSEMENT	VARCHAR2(100)	Libellé de l'établissement
ADRESSE1	VARCHAR2(100)	Première ligne de l'adresse
ADRESSE2	VARCHAR2(100)	Deuxième ligne de l'adresse
CD_POSTAL	VARCHAR2(30)	Code postal
VILLE	VARCHAR2(100)	Ville
CD_PAYS (*)	NUMBER(4)	Code du pays
NO_TELEPHONE	VARCHAR2(50)	Numéro de téléphone
NO_FAX	VARCHAR2(50)	Numéro de fax
EMAIL	VARCHAR2(100)	E-Mail
ADRESSE_WEB	VARCHAR2(100)	URL du site Web de l'établissement
CD_TYPE_ETABLISSEMENT (*)	CHAR	Code du type de l'établissement

IC_PAYS

Nom du champ	Type de données	Description
<u>CD_PAYS</u>	NUMBER(4)	Code du pays
LIBELLE_PAYS	VARCHAR2(45)	Libellé du pays

IC_QUALITE

Nom du champ	Type de données	Description
<u>CD_QUALITE</u>	VARCHAR2(2)	Code de la qualité
LIBELLE_QUALITE	VARCHAR2(30)	Libellé de la qualité
CD_TYPE_PROD (*)	VARCHAR2(2)	Type de production réalisée pour une soutenance par les personnes qui ont cette qualité
CD_DIPLOME (*)	VARCHAR2(4)	Diplôme préparé par les personnes qui ont cette qualité

IC_SUPPORT_AUDIO

<i>Nom du champ</i>	<i>Type de données</i>	<i>Description</i>
<u>CD_SUPPORT</u>	VARCHAR2 (1)	Code
LIBELLE_SUPPORT	VARCHAR2(20)	Libellé

IC_TYPE_ETABLISSEMENT

<i>Nom du champ</i>	<i>Type de données</i>	<i>Description</i>
<u>CD_TYPE_ETAB</u>	CHAR	Code du type d'établissement
ABR_TYPE_ETAB	VARCHAR2(50)	Abréviation du type d'établissement
LIB_TYPE_ETAB	VARCHAR2(255)	Libellé du type d'établissement

IC_TYPE_PRODUCTION

<i>Nom du champ</i>	<i>Type de données</i>	<i>Description</i>
<u>CD_TYPE_PROD</u>	VARCHAR2(2)	Code du type de production
LIBELLE_TYPE_PROD	VARCHAR2(50)	Libellé du type de production
DOC_INTERNE	NUMBER(1)	Type de document à usage interne ? (Oui/Non)

Annexe 9 Création de la table "RevuesISI"

Procédure de création de la table :

- Sur le site de l'ISI (<http://www.isinet.com/cgi-bin/jrnlst/jlresults.cgi>), rapatrier, sous WORD, par des copier/coller, toutes les revues dépouillées par l'ISI.
- Nous avons choisi comme séparateur de champs le point virgule. Sous WORD, supprimer la numérotation des revues et le dernier retour chariot. Puis, grâce à la fonction "Remplacer", transformer les retours chariots par des points virgule et la séquence "retour à la ligne ; retour chariot" par un simple retour chariot.
- Enregistrer le fichier en format "texte seulement".
- Importer le fichier dans Access en choisissant le point virgule comme séparateur de champs.

Annexe 10 les requêtes en mode SQL

1. La clef de dédoublement

La requête SQL pour la clef de dédoublement est :

```
SELECT DISTINCT UCase(DBADIRREC_I_PRODUCTION.TITRE) AS TITRE,
DBADIRREC_I_PRODUCTION.ANNEE, Count(*) AS Nb
FROM DBADIRREC_I_PRODUCTION
GROUP BY UCase(DBADIRREC_I_PRODUCTION.TITRE),
DBADIRREC_I_PRODUCTION.ANNEE, DBADIRREC_I_PRODUCTION.CD_TYPE_PROD
HAVING ((DBADIRREC_I_PRODUCTION.ANNEE BETWEEN "2001" AND "2002") AND
(DBADIRREC_I_PRODUCTION.CD_TYPE_PROD = "AP") AND ((Count(>1));
```

2. Les Articles "INSA"

2.1. Les laboratoires inter-établissements

```
SELECT DISTINCT DBADIRREC_I_PRODUCTION.* INTO WPre1ArticlesINSA
FROM ((DBADIRREC_I_PRODUCTION INNER JOIN DBADIRREC_I_PRODUCTION_AUTEUR
ON (DBADIRREC_I_PRODUCTION.CD_PRODUCTION =
DBADIRREC_I_PRODUCTION_AUTEUR.CD_PRODUCTION) AND
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE =
DBADIRREC_I_PRODUCTION_AUTEUR.CD_LABORATOIRE)) INNER JOIN
DBADIRREC_I_PERSONNEL ON (DBADIRREC_I_PRODUCTION_AUTEUR.CD_DOSSIER =
DBADIRREC_I_PERSONNEL.CD_DOSSIER) AND
(DBADIRREC_I_PRODUCTION_AUTEUR.CD_LABORATOIRE =
DBADIRREC_I_PERSONNEL.CD_LABORATOIRE)) INNER JOIN DBADIRREC_IC_QUALITE
ON DBADIRREC_I_PERSONNEL.CD_QUALITE = DBADIRREC_IC_QUALITE.CD_QUALITE
WHERE (((DBADIRREC_I_PRODUCTION.CD_TYPE_PROD)="AP") AND
(((DBADIRREC_I_PERSONNEL].[CD_QUALITE])="AU" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="DE" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="DO" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="TD" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="CE" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="PO" Or
```

```

(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="CN" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="DI" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="EX" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="ST")=False) AND
((DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=27 Or
(DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=501) AND
((DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="CEGELY" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="CETHIL" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="CREATIS" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="ICTT" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LCO" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LIRIS" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LMFA-INSA" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="IMP-INSA" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="MAPLY-INSA" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="PRISMa" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="UMG"));

```

2.2. Les laboratoires ayant l'INSA comme seule tutelle

```

SELECT DISTINCT DBADIRREC_I_PRODUCTION.* INTO WPre2ArticlesINSA
FROM ((DBADIRREC_I_PRODUCTION INNER JOIN DBADIRREC_I_PRODUCTION_AUTEUR
ON (DBADIRREC_I_PRODUCTION.CD_PRODUCTION =
DBADIRREC_I_PRODUCTION_AUTEUR.CD_PRODUCTION) AND
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE =
DBADIRREC_I_PRODUCTION_AUTEUR.CD_LABORATOIRE)) INNER JOIN
DBADIRREC_I_PERSONNEL ON (DBADIRREC_I_PRODUCTION_AUTEUR.CD_DOSSIER =
DBADIRREC_I_PERSONNEL.CD_DOSSIER) AND
(DBADIRREC_I_PRODUCTION_AUTEUR.CD_LABORATOIRE =
DBADIRREC_I_PERSONNEL.CD_LABORATOIRE)) INNER JOIN DBADIRREC_IC_QUALITE
ON DBADIRREC_I_PERSONNEL.CD_QUALITE = DBADIRREC_IC_QUALITE.CD_QUALITE
WHERE (((DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="BF2I" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="CITI" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="CNDRI" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="EVS" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="GEMPPM" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LAI" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LAEPSI" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LAMCOS" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LDMS" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LGEF" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LPCI" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LPM" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="LVA" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="PLM" Or
(DBADIRREC_I_PRODUCTION.CD_LABORATOIRE)="URGC") AND
((DBADIRREC_I_PRODUCTION.CD_TYPE_PROD)="AP") AND
(((DBADIRREC_I_PERSONNEL].[CD_QUALITE])="AU" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="CE" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="CN" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="EX" Or
(DBADIRREC_I_PERSONNEL].[CD_QUALITE])="ST")=False) AND
((DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=27 Or
(DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=28 Or
(DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=124 Or
(DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT)=501 Or
(DBADIRREC_I_PERSONNEL.CD_ETABLISSEMENT) Is Null));

```

3. Les articles "INSA" de 2001 et 2002 issus de la base LabINSA

```
SELECT DISTINCT WArticlesINSA.* INTO [WArticlesINSA2001-2002LabINSA]
FROM WArticlesINSA
WHERE ((WArticlesINSA.ANNEE)="2001" Or (WArticlesINSA.ANNEE)="2002");
```

4. Les articles de l'INSA pour 2001 et 2002 issus du corpus d'Epsilon2004

```
SELECT referenceBIS.* INTO [WArticlesINSA2001-2002Epsilon]
FROM (institution INNER JOIN relations_institution_etablissement_PUL ON institution.id_institution
= relations_institution_etablissement_PUL.id_institution) INNER JOIN (referenceBIS INNER JOIN
relations_ref_institution ON referenceBIS.id_ref = relations_ref_institution.id_ref) ON
institution.id_institution = relations_ref_institution.id_institution
WHERE (((relations_institution_etablissement_PUL.id_etablissement)="INSA") AND
((referenceBIS.publication_year)="2001" Or (referenceBIS.publication_year)="2002") AND
((referenceBIS.document_type)="article"));
```

5. Comparaison sur les 25 premiers caractères du titre de LabINSA et d'Epsilon2004.

```
SELECT DISTINCT [WArticlesINSA2001-2002LabINSA].CD_LABORATOIRE,
[WArticlesINSA2001-2002LabINSA].CD_PRODUCTION, [WArticlesINSA2001-
2002DedoubEpsilon].id_ref, [WArticlesINSA2001-2002LabINSA].TITRE, [WArticlesINSA2001-
2002DedoubEpsilon].title, [WArticlesINSA2001-2002DedoubEpsilon].source,
WDBADIRREC_I_PERIODIQUE.LIBELLE_PERIODIQUE INTO WArticlesINSACommunFiltre
FROM ([WArticlesINSA2001-2002LabINSA] INNER JOIN [WArticlesINSA2001-
2002DedoubEpsilon] ON ( LEFT ( UCASE([WArticlesINSA2001-2002LabINSA].TITRE), 25) =
LEFT(UCASE ([WArticlesINSA2001-2002DedoubEpsilon].title), 25)) AND ( [WArticlesINSA2001-
2002LabINSA].ANNEE = LEFT([WArticlesINSA2001-2002DedoubEpsilon].publication_year, 4)))
LEFT JOIN WDBADIRREC_I_PERIODIQUE ON ([WArticlesINSA2001-
2002LabINSA].CD_PERIODIQUE = [WDBADIRREC_I_PERIODIQUE].CD_PERIODIQUE) AND
([WArticlesINSA2001-2002LabINSA].CD_LABORATOIRE =
[WDBADIRREC_I_PERIODIQUE].CD_LABORATOIRE);
```

Résultats : 233 articles, après dédoublement sur le titre.

6. Les articles "INSA" absents du corpus d'Epsilon2004

```
SELECT DISTINCTROW [WArticlesINSA2001-2002LabINSADedoub].* INTO
WArticlesINSAFinal
FROM WArticlesINSACommunManuel RIGHT JOIN [WArticlesINSA2001-2002LabINSADedoub]
ON (WArticlesINSACommunManuel.ANNEE = [WArticlesINSA2001-2002LabINSADedoub].ANNEE)
AND (WArticlesINSACommunManuel.TITRE = [WArticlesINSA2001-2002LabINSADedoub].TITRE)
WHERE WArticlesINSACommunManuel.TITRE = NULL;
```

7. Les articles "INSA" ayant un code périodique renseigné

```
SELECT DISTINCT WArticlesINSAFinalDedoub2.CD_LABORATOIRE,
WArticlesINSAFinalDedoub2.CD_PRODUCTION, WArticlesINSAFinalDedoub2.TITRE,
DBADIRREC_I_PERIODIQUE.LIBELLE_PERIODIQUE,
DBADIRREC_I_PERIODIQUE.ABREGE_PERIODIQUE, DBADIRREC_I_PERIODIQUE.ISSN
INTO WRevuesArticlesNonISI
FROM WArticlesINSAFinalDedoub2 INNER JOIN DBADIRREC_I_PERIODIQUE ON
(WArticlesINSAFinalDedoub2.CD_LABORATOIRE =
DBADIRREC_I_PERIODIQUE.CD_LABORATOIRE) AND
```

```
(WArticlesINSAFinalDedoub2.CD_PERIODIQUE =  
DBADIRREC_I_PERIODIQUE.CD_PERIODIQUE);
```

8. Les revues des articles "INSA" non dépouillés par l'ISI

```
SELECT DISTINCTROW WRevuesNonISIdedoub.LIBELLE_PERIODIQUE,  
WRevuesNonISIdedoub.ABREGES_PERIODIQUE, WRevuesNonISIdedoub.ISSN INTO  
WRevuesNonISI  
FROM WRevuesNonISIdedoub LEFT JOIN RevuesISI ON WRevuesNonISIdedoub.ISSN =  
RevueISI.ISSN  
WHERE (((RevueISI.ISSN) Is Null));
```

Annexe 11 Articles "INSA" présents dans les Current contents, mais absents d'Epsilon2004.

Exemple 1 : article absent du corpus d'Epsilon2004 à cause de l'adresse. Aucun nom de ville utilisé par Camille Prime-Claverie n'apparaît dans le champs IN.

PT: Journal
 TI: Strong similarity measures for ordered sets of documents in information retrieval
 AU: Egghe-L; Michel-C
IN: Limburgs Univ Ctr, B-3590 Diepenbeek, Belgium; Univ Instelling Antwerp, B-2610 Wilrijk, Belgium; DU Bordeaux 3, MSHA, CEM GRESIC, F-33607 Pessac, France
 SO: INFORMATION-PROCESSING-AND-MANAGEMENT. NOV 2002; 38 (6) : 823-848
 DT: Article
 PY: 2002
 IS: 0306-4573

Le titre, les auteurs, l'année et la revue correspondent à la production : ICTT n°94.

Exemple 2 : article non retenu par le filtre à cause de l'adresse. Aucune appellation de l'INSA ni de numéro d'UMR n'apparaissent dans le champs IN.

PT: Journal
 TI: Polyethylene as an additive for mineral oils - Part II: EHL traction behavior
 AU: Bercea-M; Bercea-T; Nelias-D; Flamand-L
IN: Petru Poni Inst Macromol Chem, R-6600 Iasi, Romania; GHEORGHE ASACHI Tech Univ, Dept Machine Design & Tribol, R-6600 Iasi, Romania; Lab Mecan Contacts, F-69621 Villeurbanne, France
 SO: TRIBOLOGY-TRANSACTIONS. APR 2002; 45 (2) : 145-152
 DT: Article
 PY: 2002
 IS: 1040-2004

Le titre, les auteurs, l'année et la revue correspondent à la production : LaMCoS n°745.

Exemple 3 : article non retenu à cause de la langue du titre titre.

PT: Journal

TI: Soils reinforced by bolting - Numerical analysis and applications to the face of a deep tunnel

AU: Dias-D; Kastner-R; Jassionnesse-C

IN: Inst Natl Sci Appl, URGC Geotech, F-69621 Villeurbanne, France;
Geos Ingn Conseils SA, F-74160 Archamps, France

SO: GEOTECHNIQUE-. FEB 2002; 52 (1) : 15-27

DT: Article

PY: 2002

IS: 0016-8505

Le titre (en français tel que dans LabINSA), les auteurs, l'année et la revue correspondent à la production : URGC n°1242.

Annexe 12 le test du Khi-deux

Le test du Khi-deux nous a permis de vérifier les hypothèses suivantes :

- Est-ce que le statut inter-établissement des laboratoires influe sur le nombre d'articles présents dans les *Current Contents* ?
- Est-ce que le statut EPST des laboratoires EPST influe sur le nombre d'articles présents dans les *Current Contents* ?

Nous avons défini comme articles "ISI" les articles signalés dans les *Current Contents* et comme articles non "ISI" ceux absents des *Current Contents*.

1. Le statut inter-établissement

Les tableaux ci-dessous présentent les calculs liés au test du Khi-deux.

Effectifs réels	Laboratoires		Total
	INSA	inter-établissements	
Articles "ISI"	273	82	355
Articles "non ISI"	119	46	165
Total	392	128	520

Tableau A : les effectifs observés

Effectifs théoriques	Laboratoires		Total
	INSA	inter-établissements	
Articles "ISI"	267,62	87,38	355,00
Articles "non ISI"	124,38	40,62	165,00
Total	392,00	128,00	520,00

Tableau B : les effectifs théoriques

Ecart à l'indépendance	Laboratoires	
	INSA	inter-etablissements
Articles "ISI"	5,38	-5,38
Articles "non ISI"	-5,38	5,38
Total	0	0

Tableau C : les écarts à l'indépendance

khi-2 par case	Laboratoires	
	INSA	inter-etablissements
Articles "ISI"	0,11	0,33
Articles "non ISI"	0,23	0,71
Khi-deux total	1,38	

Tableau D : le Khi-deux par case

Le test est non significatif : $\chi^2_{\text{cal}} = 1.38 < \chi_{\text{th}} = 6.63$ (à 1%). Conclusion, le statut inter-établissement d'un laboratoire n'influe pas sur le nombre d'articles présents dans les *Current Contents*.

2. Le statut EPST

Les tableaux ci-dessous présentent les calculs liés au test du Khi-deux.

Effectifs réels	Laboratoires		Total
	EPST	non EPST	
Articles "ISI"	304	51	355
Articles "non ISI"	122	43	165
Total	426	94	520

Tableau A' : les effectifs observés

Effectifs théoriques	Laboratoires		Total
	EPST	non EPST	
Articles "ISI"	290,83	64,17	355,00
Articles "non ISI"	135,17	29,83	165,00
Total	426,00	94,00	520,00

Tableau B' : les effectifs théoriques

Ecart à l'indépendance	Laboratoires	
	EPST	non EPST
Articles "ISI"	13,17	-13,17
Articles "non ISI"	-13,17	13,17
Total	0	0

Tableau C' : les écarts à l'indépendance

Khi-2 par case	Laboratoires	
	EPST	non EPST
Articles "ISI"	0,60	2,70
Articles "non ISI"	1,28	5,82
Khi-deux total	10,40	

Tableau D' : le Khi-deux par case

Le test est significatif : $\chi^2_{\text{cal}} = 10.80 < \chi_{\text{th}} = 6.63$ (à 1%). Conclusion, le statut EPST d'un laboratoire ESPT influe sur le nombre d'articles présents dans les *Current Contents*.

L'écart à l'indépendance positif (+13.17) pour les articles "ISI" des laboratoires EPST, ainsi que le poids du Khi-deux par case pour les laboratoires non EPST/Articles "non ISI" (50% du Khi-deux total), semblent indiquer que les laboratoires EPST publient plus dans les revues dépouillées par l'ISI que les autres laboratoires.

Annexe 13 Les revues dépouillées ou non par l'ISI

Les 44 revues non dépouillées par l'ISI sont les suivantes :

LIBELLE_PERIODIQUE	ISSN
Entropie	0013-9084
GEOLOGUES	0016-7916
Industria Italiana del Cemento	0019-7637
Ingénieurs de l'Automobile, SIA	0020-1200
International Journal of Refrigeration	0020-6970
Materials Science and Engineering	0025-5416
Matériaux et Techniques	0032-6895
Romanian Journal of Technical Sciences. Applied Mechanics	0035-4074
AMERICAN SOCIETY for TESTING and MATERIALS	0097-4153
Fonderie- Fondateur d'aujourd'hui	0249-3136
Materials Science Forum	0255-5476
Techniques Sciences Méthodes	0299-7258
Revue magrébine de documentation et d'information	0330-9274
Bulletin du Cercle d'Etude des Métaux	0366-4104
Journal of acoustic emission	0730-0050
C.R. Math. Acad. Sci. Paris	0764-4442
American Society of Civil Engineering	0360-0556
Surface Modification Technologies XIV	0844-5389
International Journal of Refractory Metals and Hard Materials	0958-0611
ISET Journal of Earthquake Technology	0972-0405
Tsinghua Science and Technology	1007-0214
Journal of Shanghai Jiao Tong University	1007-1172
Defect and diffusion forum	1012-0386
Solid State Phenomena	1012-0394
International Journal of Rotating Machinery	1023-621X
International Journal of energy, environment, economics	1054-853X
Journal of Flow Visualization and Image Processing	1065-3090
Intelligent Data Analysis Journal	1088-467X
Actualités GFP	1148-1412

European Journal of Automation	1166-3049
Revue Européenne des éléments finis	1250-6559
Revue Internationale de CFAO et d'Informatique graphique	1266-0175
Mécanique et industrie	1267-3889
Essais Industriels	1290-0206
International Journal of Forming Processes	1292-7775
Revue Internationale de Génie Electrique	1295-490X
Philosophical Magazine. A, Physics of condensed matter. Structure, defects and mechanical properties	1364-2804
European Journal of Mechanical and Environmental Engineering	1371-6980
Archives of Physiology and Biochemistry	1381-3455
New Astronomy	1384-1076
Journal of Micromechatronics	1389-2258
Bulletin of Geology Engineering and Environment	1435-9529
Urban Water	1462-0758
Genome Biology	1474-760X

Les 29 revues dépouillées par l'Isi sont les suivantes :

LIBELLE_PERIODIQUE	ISSN
Annals of the New York Academy of Sciences	0077-8923
Asymptotic Analysis	0921-7134
Canadian Geotechnical Journal/Revue Canadienne de Géotechnique	0008-3674
Data&Knowledge Engineering Journal	0169-023X
GEOTECHNIQUE	0016-8505
High Temperatures-High Pressures	0018-1544
High-temperature materials and processes	0334-6455
International Journal of Experimental Heat Transfer	0891-6152
International Journal of fracture	0376-9429
J. Acoust. Soc. Am.	0001-4966
Journal La Houille Blanche	0018-6368
Journal of Applied Polymer Science	0021-8995
Journal of Cement and Concrete Composites	0958-9465
Journal of Fluid Engineering	0098-2202
Journal of materials research	0884-2914
Journal of Sound and Vibration	0022-460X
Journal of Sports Sciences	0264-0414
Journal of the American Ceramic Society	0002-7820
Journal of the European Ceramic Society	0955-2219
L'actualité chimique	0151-9093
Materials in engineering	0261-3069
Non destructive testing and Evaluation international	0963-8695
Polymer	0032-3861
Polymer composites	0272-8397
Polymer journal	0032-3896
Revue de Métallurgie	0035-1563
Silicates industriels	0037-5225
Tribology International	0301-679X
Wear	0043-1648

Annexe 14 les articles absents de façon injustifiée des *Currents Contents*

Les 39 articles absents des *Currents Contents* pour une raison inconnue sont :

CD_LABO	CD_PROD	TITRE	LIBELLE_PERIODIQU UE	ISSN
CETHIL	3	Nucleate pool boiling under DC electric field	International Journal of Experimental Heat Transfer	0891-6152
CETHIL	4	Comparisons between two models of cooling surfaces using blowing	Annals of the New York Academy of Sciences	0077-8923
CETHIL	11	Determination of spectral radiative properties of open- cell foam : hemispherical-directional and directional-directional spectral transmittance measurement	High Temperatures-High Pressures	0018-1544
CETHIL	490	Numerical study of the blowing impact on the flow past a circular cylinder	Journal of Fluid Engineering	0098-2202
GEMPPM	42	Mechanical behaviour above T _g of a plasticised PVC reinforced with cellulose whiskers ; a SANS structural study	Polymer	0032-3861
GEMPPM	43	Modulus recovery kinetics and other insights into the Payne effect for filled elastomers	Polymer composites	0272-8397
GEMPPM	44	Viscoelastic properties of a plasticised PVC reinforced with cellulose whiskers	Journal of Applied Polymer Science	0021-8995
GEMPPM	370	Sources of dispersion in eddy current conductivity measurements performed on aluminium alloy aeronautic products	Non destructive testing and Evaluation international	0963-8695
GEMPPM	374	Influence of the mechanical loading history on the stress assisted two way memory effect in a TiNiCa alloy	Journal of materials research	0884-2914
GEMPPM	382	Micromechanical modelling of rate and temperature dependent fracture of glassy polymers	International Journal of fracture	0376-9429

GEMPPM	620	Oxygen vacancies relaxation and domain wall hysteresis motion in cobalt-doped barium titanate ceramics	Journal of the American Ceramic Society	0002-7820
GEMPPM	624	Slow crack growth in zirconia ceramics : from the single crystal to the composites	Materials in engineering	0261-3069
GEMPPM	627	Effect of initial particle distribution on the grain morphology of CA6	Journal of the European Ceramic Society	0955-2219
GEMPPM	635	Thermomechanical behaviour of long fibres ceramic-ceramic composites	Silicates industriels	0037-5225
GEMPPM	653	Thermal conductivity and mechanical properties related to microstructure of a high alumina refractory castable	Silicates industriels	0037-5225
GEMPPM	870	Elaboration de composites modèles Al-ZrO ₂ par métallurgie des poudres	Revue de Métallurgie	0035-1563
GEMPPM	928	Les céramiques : futurs développements	L'actualité chimique	0151-9093
LaMCoS	684	Micropolar effects in the coupling of a thin film past a porous medium	Asymptotic Analysis	0921-7134
LaMCoS	720	Numerical simulations with 3DWAPE considering shallow water range dependent environments	J. Acoust. Soc. Am.	0001-4966
LaMCoS	1118	Fixed versus Free-Floating stretcher mechanism in rowing ergometers: a race rate dynamic analysis.	Journal of Sports Sciences. A paraître	0264-0414
LaMCoS	1334	A numerical simulation and experimental study of the interaction between a pure high-velocity waterjet and targets : contribution to investigate the decoating process	Wear	0043-1648
LaMCoS	1403	Numerical and experimental investigations on rolling contact fatigue with dented surfaces	Tribology International	0301-679X
LaMCoS	1430	New models in micropolar fluids and application to lubrication	Asymptotic Analysis	0921-7134
LDMS	135	Balancing of machinery with a flexible variable-speed rotor	Journal of Sound and Vibration	0022-460X
LDMS	179	Review of the book: Rotating Machinery Vibration, by M.L. ADAMS and M. DEKKER	Journal of Sound and Vibration	0022-460X
LIRIS	687	Conceptual Framework for Document Semantic Modelling: an application to document and knowledge management in the legal domain	Data&Knowledge Engineering Journal	0169-023X
LMFA-INSA	15	Exploration par Anémométrie Laser Doppler de l'écoulement dans une pompe hélico-axiale.	Journal La Houille Blanche	0018-6368
LMFA-INSA	17	Analyse de l'écoulement instationnaire dans l'entrefer roue-diffuseur d'une pompe centrifuge.	Journal La Houille Blanche	0018-6368

LPCI	102	Electrochemical properties of poly(2-Mercaptobenzimidazole) films prepared in methanol alkaline solvent on copper zinc alloy	Polymer journal	0032-3896
LPCI	106	On-line continuous monitoring of a purification process by plasma using direct analysis of metals in the flue gases by ICP-OES	High-temperature materials and processes	0334-6455
LPCI	107	Gas temperature in contracted atmospheric pressure discharges sustained in cylindrical tubes by microwaves at 2450 MHz	High-temperature materials and processes	0334-6455
MAPLY- INSA	10	Micropolar effects in the coupling of a thin film past a porous medium.	Asymptotic analysis	0921-7134
MAPLY- INSA	16	Homogenization and two-scale analysis of the transient compressible Reynolds equation	Asymptotic analysis	0921-7134
URGC	985	Geotechnical survey and mechanical parameters in urban soils : part two. Towards a hierarchy of numerical modelling related uncertainties	Canadian Geotechnical Journal/Revue Canadienne de Géotechnique	0008-3674
URGC	986	Geotechnical survey and mechanical parameters in urban soils : part one. Modelling soil variability and inferring representative values case of Lyon subway line D extension	Canadian Geotechnical Journal/Revue Canadienne de Géotechnique	0008-3674
URGC	1034	Etude de la migration d'un lixiviat dans un sol non saturé par la méthode TDR	GEOTECHNIQUE	0016-8505
URGC	1083	Metakaolin and calcined clays	Journal of Cement and Concrete Composites	0958-9465
URGC	1179	Durabilité des ciments phospho-magnésiens	Silicates Industriels	0037-5225
URGC	1204	Modélisation des flux polluants en hydrologie urbaine : évolution depuis les années 1960 et perspectives pour les années 2000	La Houille Blanche	0018-6368