

## ***ANALYSE BIBLIOMETRIQUE DES COLLABORATIONS INTERNATIONALES DE L'INRA.***

---

**MULTON Jean-Louis,**

INRA, Mission Relations Internationales, 147 rue de l'Université, 75338 Paris cedex 07  
[multon@paris.inra.fr](mailto:multon@paris.inra.fr)

**BRANCA-LACOMBE Geneviève,**

INRA, DISI, BP 2078, 06606 Antibes cedex  
[lacombe@antibes.inra.fr](mailto:lacombe@antibes.inra.fr)

**DOUSSET Bernard,**

IRIT, SIG, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex 04  
[dousset@irit.fr](mailto:dousset@irit.fr)

**Résumé** : Ce travail est la réactualisation d'une étude réalisée, il y a deux ans, dans des conditions similaires et qui ne portait que sur deux périodes : 1998 et 1999 avec le début de l'an 2000. Nous avons complété le corpus qui avait été constitué à l'époque à partir des bases bibliographiques sur CD/Rom : Current-Contents, Pascal et Sci. La version CD/Rom de Sci qui ne disposait ni du résumé, ni de mots-clés a été remplacée cette fois-ci par la version en ligne du Web of science qui est beaucoup plus complète mais assez difficile à télécharger. Nous disposons maintenant de quatre années de production scientifique (janvier 1998 à Avril 2002) ce qui nous permet une détection beaucoup plus complète des acteurs et de leurs relations ainsi que des thématiques abordées. Après avoir analysé les collaborations internationales de l'INRA, nous avons pu réaliser un zoom sur deux entités de cet institut : le département TPV (Traitement des Produits Végétaux) et le positionnement de l'INRA aux Caraïbes (centres INRA des Caraïbes et pays de la région qui collaborent avec l'INRA local ou de métropole). Ce complément d'étude montre que le corpus initialement constitué pour les collaborations internationales renferme bien d'autres informations stratégiques. Chaque département, chaque laboratoire, chaque région peut ainsi prendre connaissance de son environnement, de ses forces et de ses faiblesses, de ses relations avec les autres organismes de recherche. Une politique de développement peut en être déduite en prenant en compte de façon rigoureuse des éléments objectifs mesurant la lisibilité de l'activité de recherche et de son évolution. Nous présentons les principaux résultats de cette analyse et les perspectives à moyen et long terme de ce type d'étude.

**Abstract** : This work is the actualization of a study which has been achieved in similar conditions two years ago; it was about only two periods 1998, 1999 and the beginning of 2000. We completed the corpus which was constituted with cd-rom bibliographic databases: Current-Contents, Pascal and Sci. The Sci cd-rom version which had neither abstract nor keywords, has been replaced here by the Web of Science online version, which is more

complete but not easy to load. We have now four years of scientific production (january 1998 to april 2002) which allow to detect actors, subjects and their relations. With the international collaboration analysis of INRA, we achieved a zoom on the two entities' institute: the TPV department (Treatment of Vegetal Products) and the INRA position in the Caribbean. This completed study shows that the corpus, initially constituted for the international collaborations, contains more strategic informations. By this way, each department, laboratory or area can read its environment, its forces and weaknesses, its relations with the others research organisms. Then a development policy can be deduced by the measurement of research activity and its evolution. We present the main results of the analysis and the middle and long-term views.

**Mots-clés** : Veille scientifique et technique - Bibliométrie - Collaborations internationales.

**Keywords** : Science and technology watch - Bibliometry - International collaboration

# **Analyse bibliométrique des collaborations internationales de l'INRA.**

## **INTRODUCTION**

A la suite de l'étude présentée au Colloque VSST'2001 (Veille Stratégique Scientifique et Technologique, 15-19 octobre 2001 – Barcelone) sur la période 1998-1999, une nouvelle étude bibliométrique d'évaluation quantitative des collaborations internationales de la Direction des Relations Internationales de l'INRA a été effectuée sur la période 1998 – 2001, complétée par deux zooms, l'un sur un Département de recherche et l'autre sur une région particulière : la zone Caraïbe.

Comme précédemment :

- Le logiciel retenu pour effectuer cette étude est « Tétralogie » (IRIT Toulouse) identifié comme étant celui qui répond le mieux au cahier des charges fixé. Cependant, la présente étude a bénéficié des nouvelles avancées de « Tétralogie » en matière de cartographie et d'analyse relationnelle.
- Le choix des bases bibliographiques (Current-Contents, Pascal et Science Citation Index,) à partir desquelles a été établi le corpus INRA, a été strictement imposé par la double nécessité d'avoir les noms et adresses complètes de tous auteurs et de recenser toutes les activités de l'INRA.
- Après extraction des données brutes utiles, un travail considérable de normalisation (notamment de traitement des synonymies) a dû être réalisé au niveau des adresses, des noms de pays, de villes et d'organismes, travail pour lequel « Tétralogie » est particulièrement bien adapté.

## **1 - LA NATURE DES PROBLEMES POSES**

### **1.1 - Etudier l'évolution des collaborations internationales de l'INRA**

La démarche que nous avons utilisée il y a deux ans pour la première étude et que nous avons reconduite pour cette actualisation se décompose en cinq points :

- Rechercher des bases consignant l'ensemble des adresses de co-signataires
- Retrouver la totalité des publications dans lesquelles est intervenu l'INRA depuis 1998
- Standardiser l'information: pays, villes, organismes, adresses, auteurs, journaux, ...
- Découper le temps en périodes significatives par rapport aux données disponibles
- Croiser les informations nécessaires à l'étude.

Les évolutions sur quatre périodes homogènes sont alors déduites de ces croisements : nombre de citations dans les bases retenues, nombre de publications nationales, nombre de publications internationales, nombre de co-signatures avec des pays étrangers, stratégies de coopérations, réorientations thématiques, stratégies de publication.

### **1.2 - Analyser l'activité du département TPV**

Ce département n'ayant pas de localisation propre, nous sommes partis de la liste de ses chercheurs pour établir un filtre permettant de récupérer, dans le corpus INRA, l'ensemble de sa production scientifique. Toutes les variations orthographiques des auteurs (plus de 550) ont été détectées et elles ont été synonymées à TPV. Un « retour aux notices » sur ce terme en mode multi-bases nous a permis d'extraire en une seule fois tout le corpus TPV. Celui-ci a ensuite été analysé en détail : détection de tous les acteurs, de leurs relations et des éléments sémantiques associés. Une analyse de l'évolution nous a ensuite permis de dégager, pour ce département, les composantes stratégiques et les signaux faibles.

### **1.3 - Analyser l'influence de l'INRA aux Caraïbes**

Pour extraire l'activité des laboratoires INRA des Caraïbes (Guadeloupe, Martinique, Guyane), nous avons travaillé dans un premier temps sur les adresses : recherche des ces trois noms et de tous leurs synonymes dans le corpus INRA. Comme certains organismes non INRA de ces départements travaillent avec l'INRA de métropole, nous avons ensuite éliminé manuellement une quinzaine de publications où l'INRA local n'intervenait pas explicitement. Le corpus Caraïbe ainsi constitué a ensuite été analysé suivant les mêmes principes que celui de TPV. Mais l'influence de l'INRA s'exerce aux Caraïbes essentiellement depuis la métropole, nous avons donc fait un zoom sur les collaborations internationales de l'INRA avec les Pays de la région. Nous avons détecté les laboratoires concernés des deux cotés et les sujets de recherche abordés. Cet inventaire devrait permettre de promouvoir une meilleure coopération locale (entre INRA Caraïbe et pays de la région) puisque la métropole peut servir de pivot pour l'initialisation de nouveaux contacts.

## 2 - DESCRIPTION DES BASES RETENUES

### 2.1 - Base Current-Contents

Principales caractéristiques de cette base :

- Deux formats avant et après janvier 98.
- Plusieurs adresses depuis janvier 98.
- En avance sur l'indexation (articles sous presse)
- Propose un résumé et des mots-clés.

Format de l'adresse :

**RP:** de Lajudie, P; IRD; INRA,AGRO M,CIRAD; Campus Baillarguet,BP 5035; F-34032 Montpellier; France

**IN:** IRD, INRA,AGRO M,CIRAD, Lab Symbioses Trop & Mediterraneennes, F-34032 Montpellier, France; State Univ Ghent, Microbiol Lab, Ghent, Belgium; IRD, Lab Microbiol Sols, Dakar, Senegal

Périodes télé-déchargées et retenues : CD/Rom de 1998 à début 2002.

Certaines publications parues en 96 et 97 sont indexées sur ces CD/Rom.

### 2.2 - Base Pascal

Principales caractéristiques de cette base :

- Deux formats avant et après janvier 96.
- Sciences exactes et appliquées.
- Mots-clés en français, anglais et espagnol.
- Présence d'un résumé.

Format de l'adresse :

**CS-** Departement Etudes et recherches du CETIOM, 174 avenue Victor Hugo, 75116 Paris, France^GIE Cartisol, 20 rue Bachaumont, 75001 Paris, France^Unite associee a l'INRA Organisation et variabilite des genomes vegetaux, Universite Blaise Pascal, Campus des Cezeaux, 63177 Aubiere, France^INRA, Station d'amelioration des plantes et de pathologie vegetale, Domaine de Crouelle, 234 avenue du Brezet, 63039 Clermont-Ferrand, France^INRA-ENSAM, Station de genetique et d'amelioration des plantes, 2 place Viala, 34060 Montpellier, France|

Périodes télé-déchargées et retenues : CD de 1998 à avril 2002.

Certaines publications parues en 96 et 97 sont indexée sur ces CD.

### 2.3 - Base SCI sur le Web of Science

Principales caractéristiques de cette base :

- Plus de 3500 journaux
- Sciences exactes et appliquées
- Pas de descripteur ni de résumé sur le CD/Rom (étude précédente)
- Descripteurs et résumé disponibles sur le Web of science (réactualisation)
- Les citations n'ont pas été récupérées dans le corpus INRA (problèmes techniques à répétition lors du déchargement en ligne).

Format de l'adresse :

**C1** Univ Munster, Dept Dermatol, Von Esmarchstr 56, D-48149

Munster, Germany

Univ Munster, Dept Dermatol, D-48149 Munster, Germany

Univ Hosp Rudolf Virchow, Inst Human Genet, Berlin, Germany

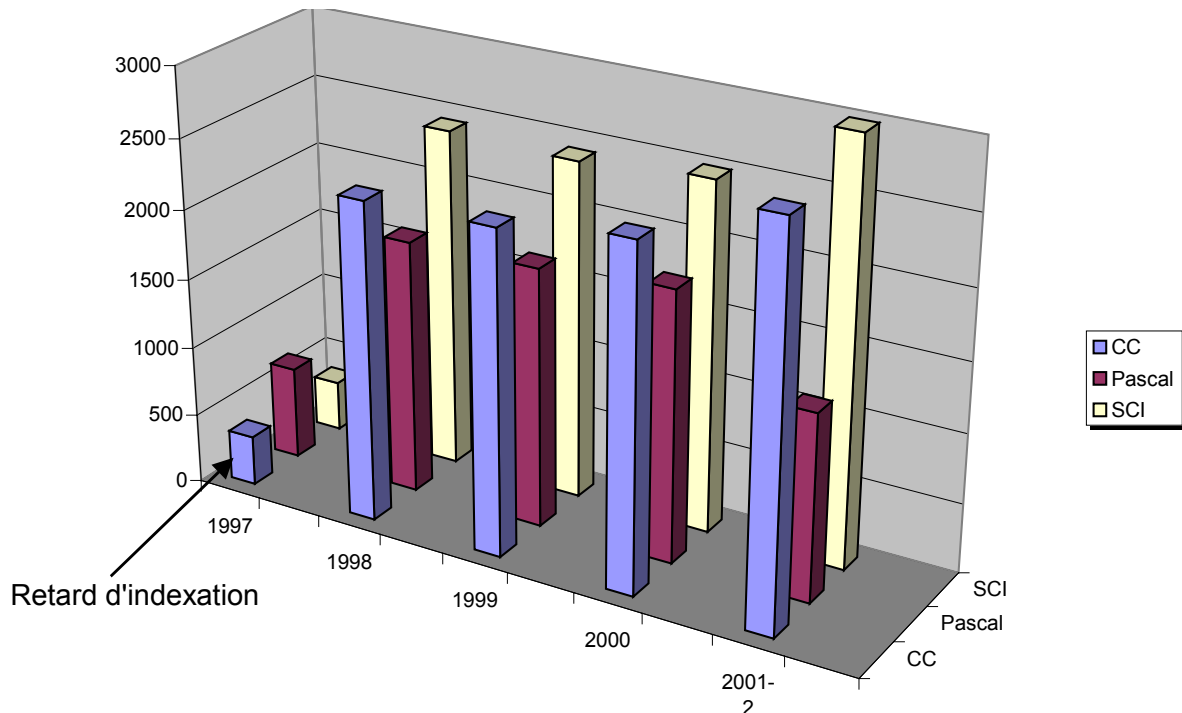
INRA, CEA, Lab Radiobiol Appl, Jouy En Josas, France

NIH, Skin Biol Lab, Bethesda, MD 20892 USA

Périodes télé-déchargées: 1998 à avril 2002.

### 2.4 - Répartition par bases et par années

Années/Bases	CC	Pascal	SCI	Sigma
1997	357	665	359	1381
1998	2290	1826	2445	6561
1999	2314	1858	2427	6599
2000	2446	1938	2501	6885
2001-2	2808	1337	3039	7184



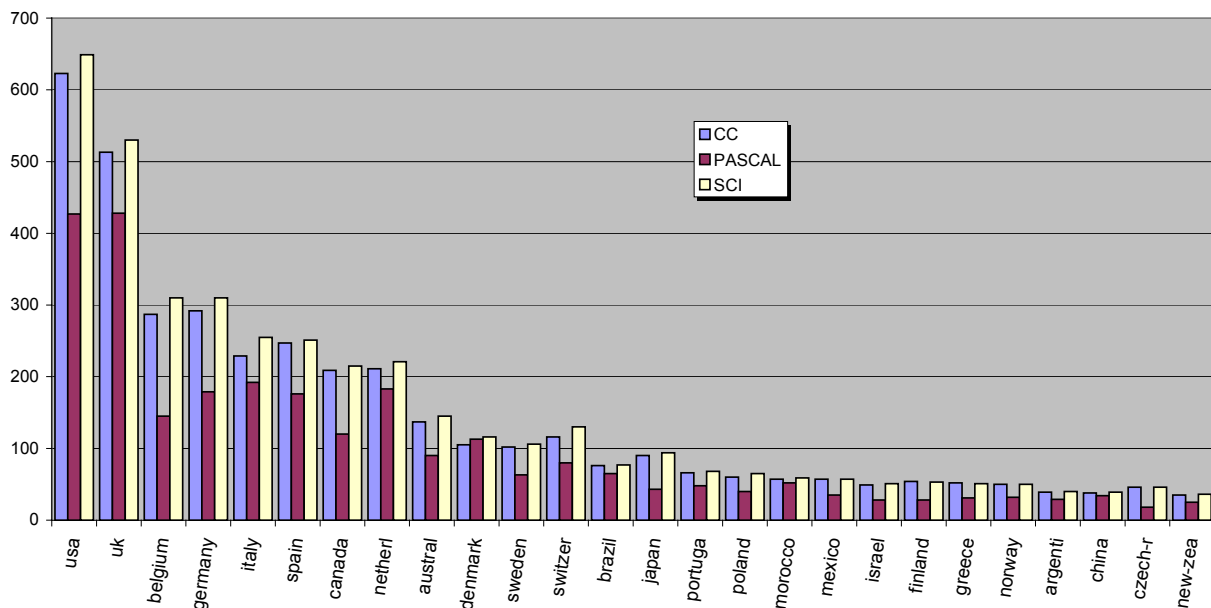
### 3 - L'ANALYSE DES COLLABORATIONS INTERNATIONALES

#### 3.1 - Avantages et inconvénients de l'analyse multi-base

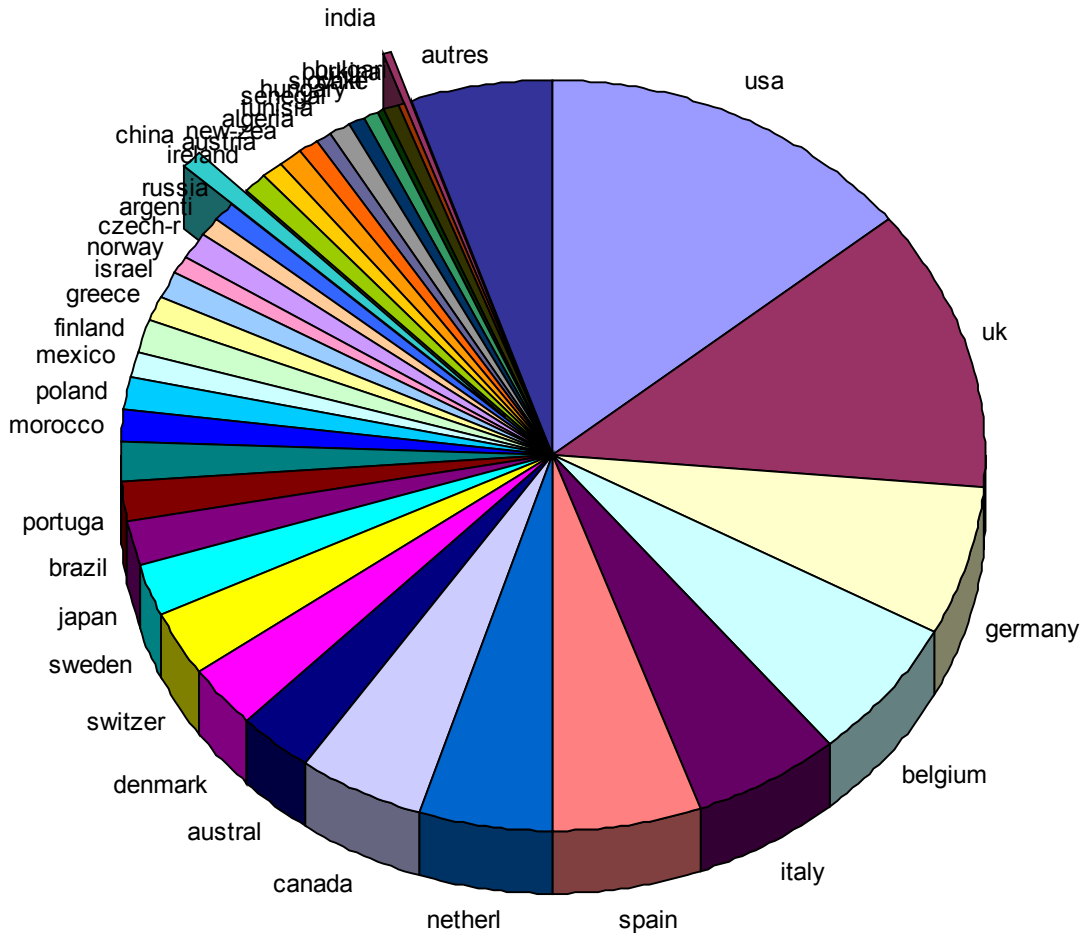
Cette analyse conduite simultanément sur les trois bases présente les caractéristiques suivantes :

- Réponse simultanée des trois bases.
- Etude de l'évolution la plus récente (avant les formats différent et pas de CC).
- Assez bonne homogénéité des quatre sous périodes : 1998, 1999, 2000 et 2001-02.
- Léger décalage de CC et SCI avec PASCAL (retard d'indexation pour PASCAL).
- La dernière sous période (01-02) a donc un déficit sur PASCAL.
- La dernière sous période (01-02) est sous représentée. (Indexation encore incomplète).
- Les valeurs brutes ne sont pas vraiment significatives.
- Il faut travailler si possible en évolution relative

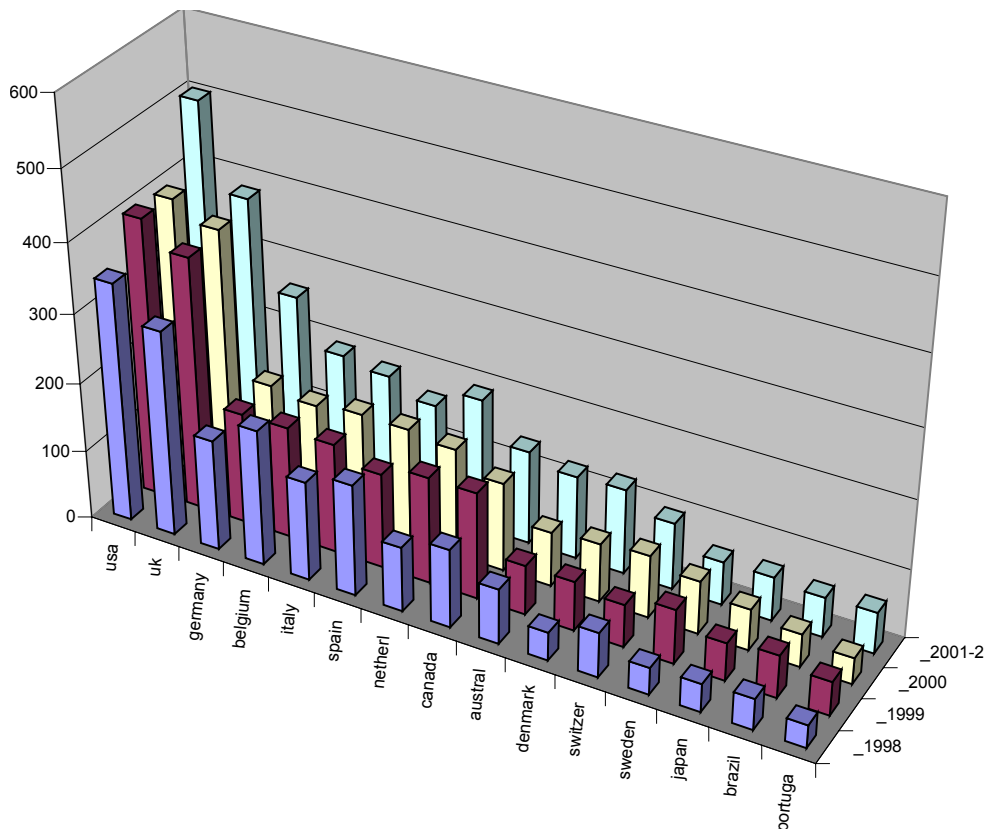
#### 3.2 - Citations dans les trois bases par pays



### 3.3 - Répartition des pays pour l'ensemble du corpus



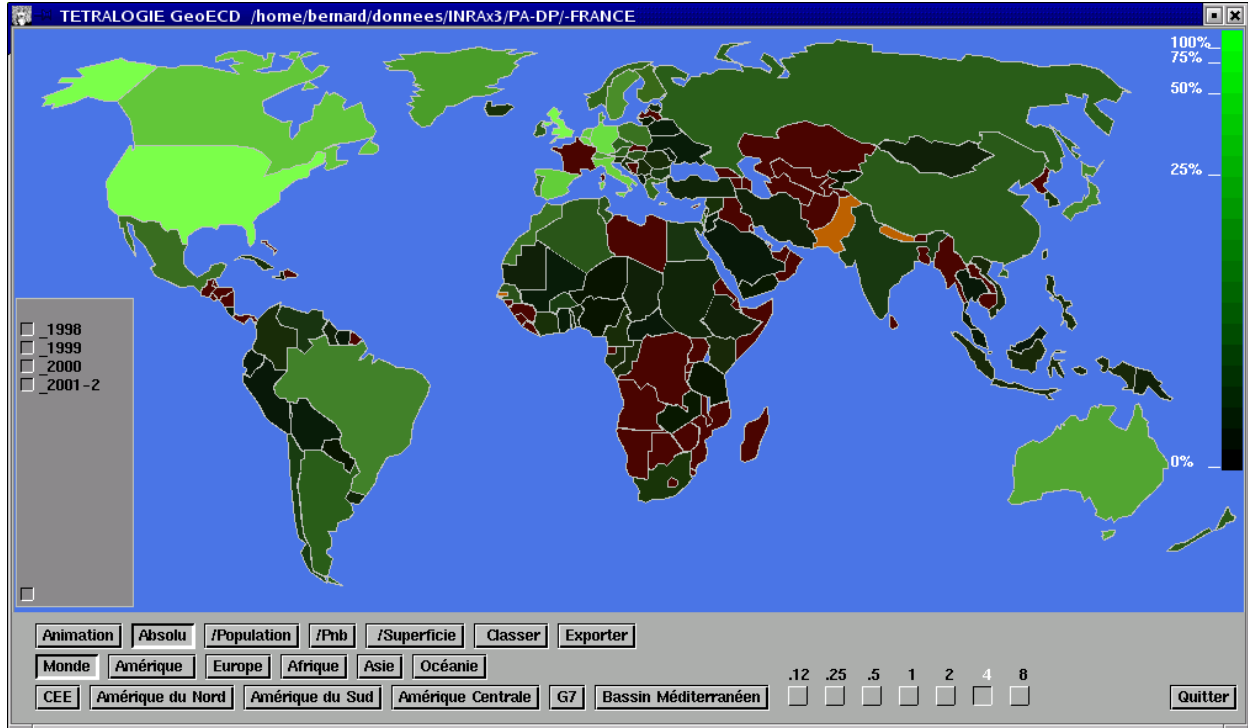
### 3.4 - Présence absolue dans les 3 bases sur les 4 périodes



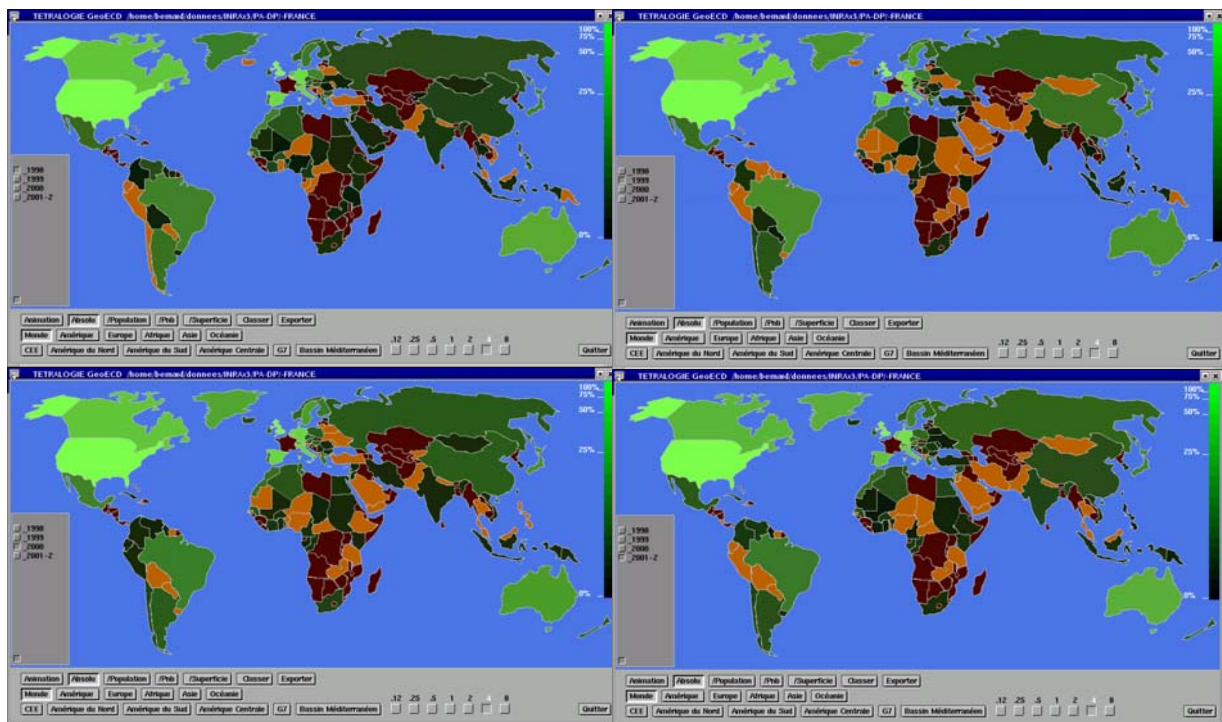
### 3.5 - Cartes d'implantation des collaborations INRA

#### 3.5.1 - Carte couvrant l'ensemble des périodes de 1998 à début 2002

La France a été enlevée de cette représentation, où figurent en grenat les pays absents, en orange les pays présents avant 1998 et perdus depuis, en nuances de vert (du noir au vert vif) le nombre d'articles co-signés, suivant une échelle privilégiant les signaux faibles.



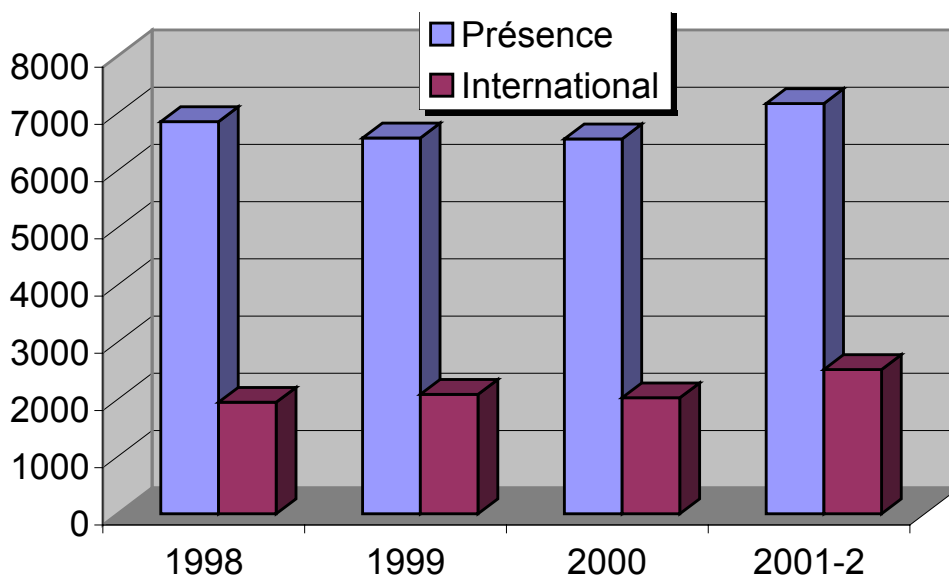
#### 3.5.2 - Cartes comparatives pour les quatre périodes retenues



### 3.6 - Evolution des collaborations internationales sur la périodes 1998 à 2001-2

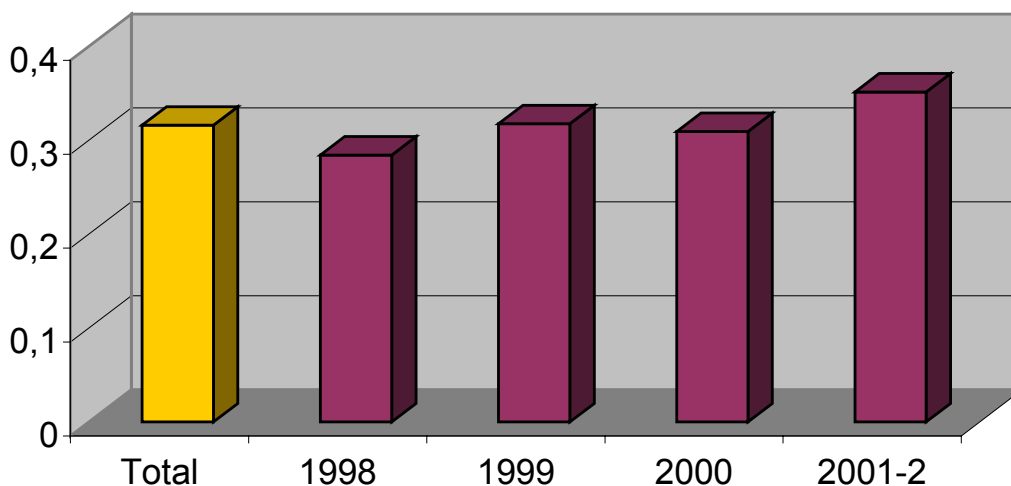
#### 3.6.1 - Répartition des références dans le temps

En valeur absolue le nombre des publications internationales augmente. Mais cette mesure n'est pas fiable car les chiffres de chaque période ne sont pas obtenus dans les mêmes conditions. Les années 1998 et 1999 sont pratiquement indexées à 100%, pour 2000 il subsiste un léger déficit d'indexation surtout pour Pascal, enfin la période 2001-02 couvre 16 mois avec un fort déficit d'indexation pour 2001 et une indexation très faible des 4 premiers mois de 2002.



#### 3.6.2 - Part des publications internationales

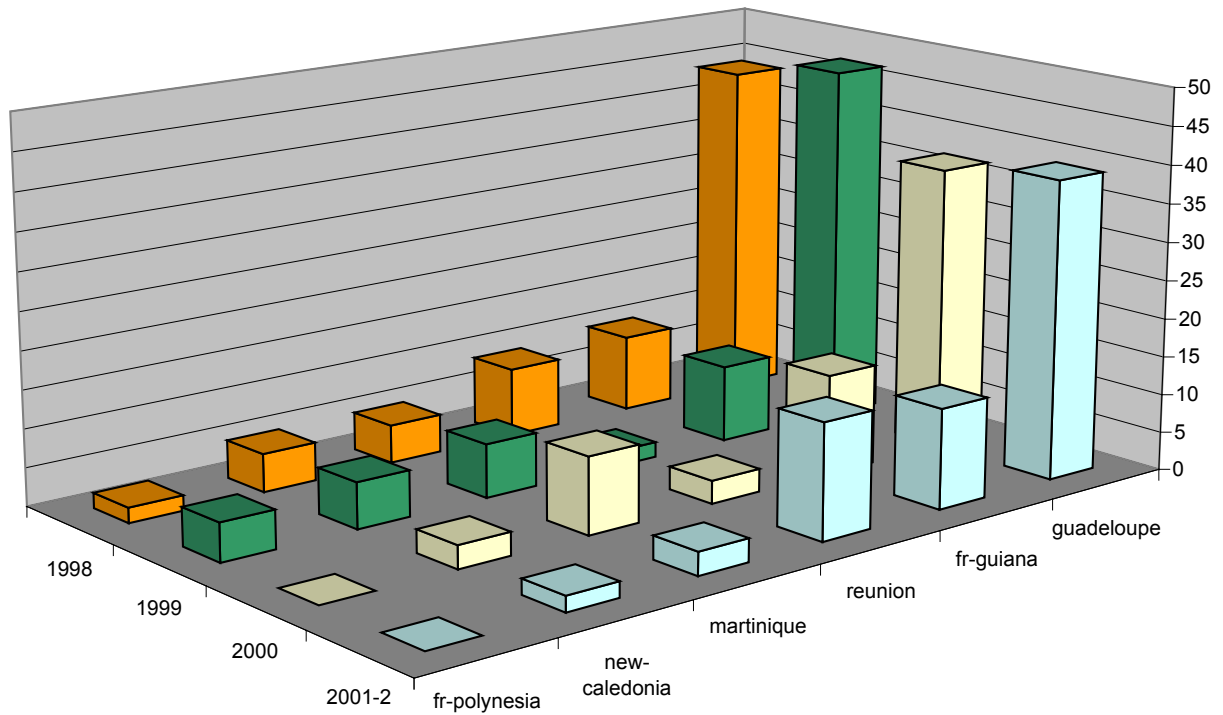
Si nous travaillons en proportion de publications internationales, nous remarquons une progression de 5 points par rapport à 1998. Les publications visibles de l'INRA sont donc de meilleure qualité, d'autant plus que la couverture des bases augmente et que des revues de second rang sont maintenant indexées.



#### 3.6.3 - Présence et évolution dans les Dom/Tom

Nous pouvons, par le biais des adresses, retrouver l'origine régionale des publications référencées. Nous avons ainsi pu isoler la production des Dom/Tom (INRA ou collaboration directe avec l'INRA). La Guadeloupe est sans conteste en première position, mais nous constatons un léger déclin des Dom/Tom sauf pour la Réunion.





### 3.6.4 - Evolution entre 1998-99 et 2000-01 dans les trois bases

#### Codes des couleurs utilisées dans les cartes géographiques :

En **grenat** : pays absents sur l'ensemble des périodes.

En **orange** : pays perdus après 2000 mais présents avant.

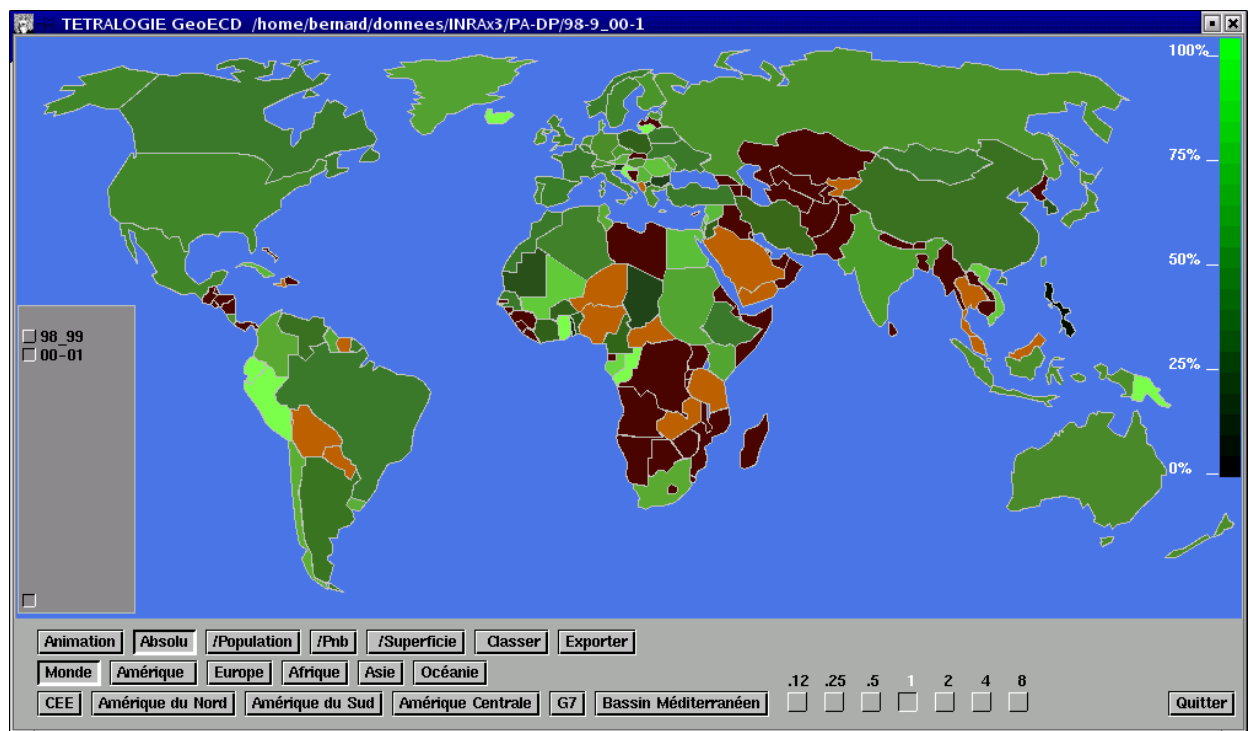
En **vert** part de la période 2000-01 sur la totalité.

**Vert sombre** diminution des collaborations

**Vert moyen** (50%) maintient

**Vert clair** augmentation des collaborations.

Dans l'exemple ci-dessous : l'Islande augmente, la Pologne diminue.



### 3.7 - Conclusions globales de cette nouvelle étude

Le nouveau système de cartes géographiques interactives propre à « Tétralogie » permet ainsi de formuler les principales conclusions suivantes :

Les publications internationales de l'INRA

- Se concentrent sur moins de pays
- Leur nombre augmente
- Leur proportion augmente (de 27 à 34%) par rapport aux publications franco-françaises.

On peut en conclure que la qualité et le niveau des publications augmentent (publications acceptées dans les grandes revues référencées dans les bases).

Par comparaison avec l'étude précédente (analyse 1998-99), la période 2000-01 montre dans la dynamique des collaborations internationales de l'INRA :

- L'arrêt du déclin africain
- La confirmation du déclin en Asie mineur et centrale, sauf Méditerranée
- La retombée de la Chine
- L'augmentation de l'Inde et du Vietnam
- L'émergence de certains sujets de recherche à l'international.

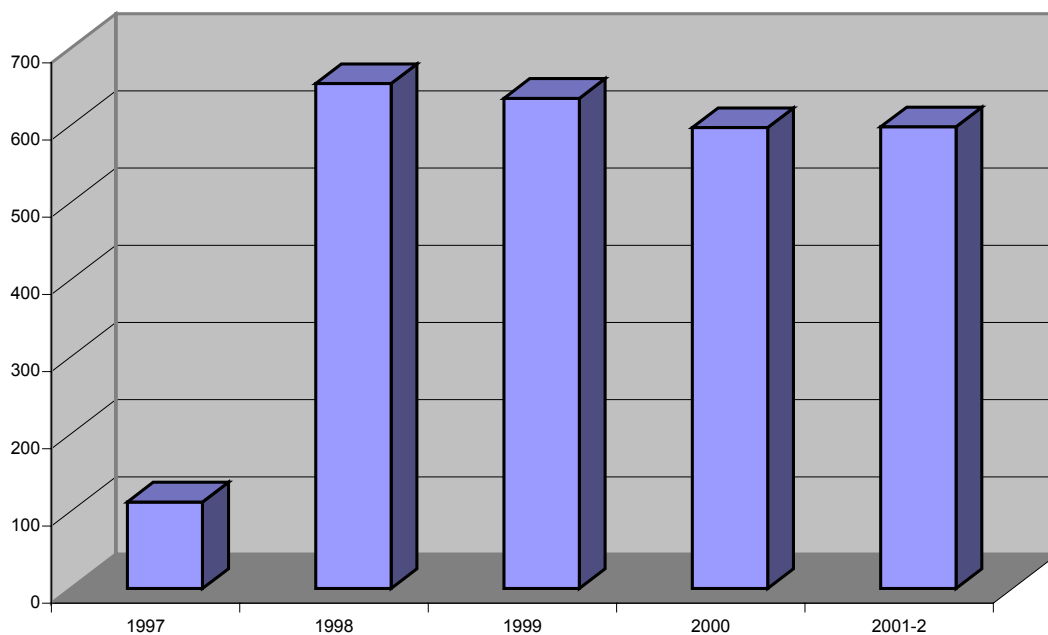
## 4 - ANALYSE DU DEPARTEMENT TPV

### 4.1 - Identification du département

La liste de tous les auteurs de TPV et des variations orthographiques recensées a été synonymisée au mot TPV. Cette méthode permet soit de filtrer la production de TPV dans le corpus INRA, soit d'extraire un sous corpus TPV en recherchant automatiquement les notices bibliographiques correspondantes dans les trois bases étudiées.

ABBAL-P	TPV	AGEORGES, A	TPV
ABBAL, P	TPV	AGUIE-BEGHIN-T-V	TPV
ABECASSIS-JOEL	TPV	AGUIE-BEGHIN-V	TPV
ABECASSIS-J	TPV	AGUIE-BEGHIN, V	TPV
ABECASSIS, J	TPV	AGUIEBEGHIN-V	TPV
AGEORGES-AGNES	TPV	AGUIE-BEGLUN, V	TPV
AGEORGES-A	TPV		

Ci-dessous la répartition dans le temps des articles retenus.



### 4.2 - Principaux résultats

Dans cette étude, nous avons détecté la totalité des acteurs connexes à TPV : auteurs, journaux, pays, villes, organismes, universités, autres centres INRA, les thèmes de recherche et les signaux faibles. Nous avons croisé

toutes ces entités entre elles afin de savoir qui fait quoi avec qui et depuis combien de temps. Nous avons aussi étudié l'évolution des équipes de recherche et leurs collaborations nationales et internationales. Par contre, nous n'avons pas encore utilisé nos méthodes d'indexation du texte libre (Titres et Résumés) afin de rendre homogène les trois bases sur le plan sémantique. En effet, une intervention des experts du domaine est nécessaire pour valider les dictionnaires de mots composés qui sont extraits du corpus par des méthodes de détection statistique. Une fois cette étape réalisée, il sera possible de connaître, avec beaucoup plus de précision qu'en utilisant les mots-clés, l'étendue du front de recherche actuel et les émergences de nouveaux sujets.

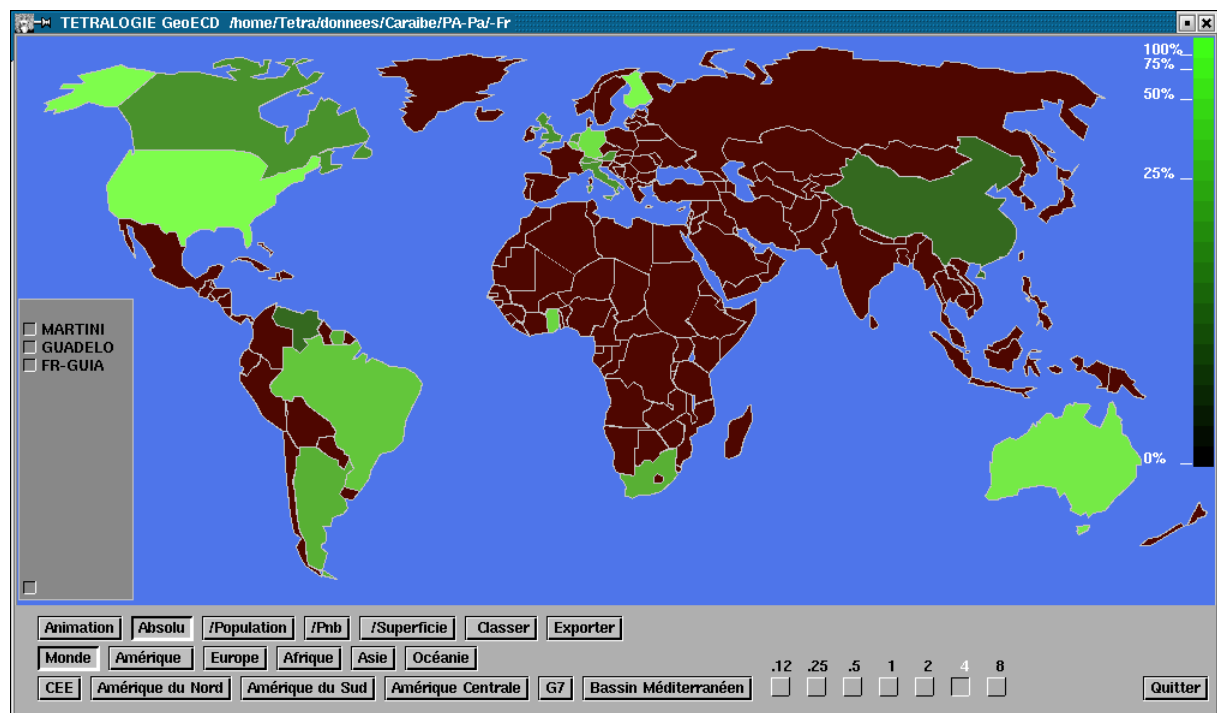
## 5 - ANALYSE DE L'INFLUENCE DE L'INRA AUX CARAÏBES

### 5.1 - Les collaborations avec les Caraïbes

	CC	PASCAL	SCI
<b>GUADELOUPE</b>	<b>58</b>	<b>53</b>	<b>63</b>
MEXICO	57	35	57
<b>FR-GUIANA</b>	<b>16</b>	<b>14</b>	<b>15</b>
CUBA	9	7	9
VENEZUELA	10	5	13
<b>MARTINIQUE</b>	<b>8</b>	<b>8</b>	<b>9</b>
COLOMBIA	9	2	8
GUYANA	4	0	2
COSTA-RICA	1	4	1
SURINAME	1	1	1
JAMAICA	1	0	1
HAITI	0	1	0

### 5.2 - Les laboratoires INRA des Caraïbes

Si on se restreint aux laboratoires INRA des Caraïbes explicitement identifiés par leur adresse, nous pouvons dresser la carte des collaborations internationales. Nous remarquons, par exemple, l'absence du Mexique, de la Chine. En fait, ces pays collaborent avec l'INRA de métropole et un laboratoire en agronomie de la région Caraïbe qui n'est pas explicitement l'INRA (Cirad, Orstom, Ifremer, ...).



## CONCLUSION

Cette seconde étude complétée par deux zooms spécifique a permis de montrer l'étendue des possibilités de la bibliométrie pour extraire d'un corpus conséquent (plus de 30 000 références) des informations stratégiques pour la conduite et le développement d'un grand institut de recherche. L'identification précise des acteurs et la détection de toutes leurs relations permet de mieux comprendre le fonctionnement d'une telle entité. Le suivi de ses relations internationales dévoile les forces et les faiblesses en terme de collaboration et de coopération ainsi que des stratégies peu visibles autrement. Chaque département peut être analysé séparément notamment sur le plan sémantique (domaine plus ciblé que l'agronomie dans son ensemble), ses caractéristiques structurelles et son fonctionnement sont alors mieux compris, ses orientations mieux évaluées, sa lisibilité mesurée, son influence délimitée. Pour une région comme celle des Caraïbes, le but d'une telle étude est de détecter tous les partenaires locaux potentiels, les contacts qu'ils ont déjà avec l'INRA et les domaines dans lesquels ils collaborent. Une extension naturelle de cette étude est la mise en ligne de l'analyse du corpus INRA sur un portail (Internet ou Intranet). Les dictionnaires et les matrices sont alors compilés dans une base de données afin de permettre à un utilisateur de faire sa propre analyse (micro analyse) en scrutant son environnement (acteurs et thèmes de recherche). Des statistiques en ligne doivent l'aider à appréhender les éléments stratégiques qui le concernent, à se situer dans le contexte national et international et à détecter à temps toute innovation pouvant l'intéresser. La macro analyse, que nous venons de produire, lui servant de référentiel tout au long sa navigation et de ses investigations.

## BIBLIOGRAPHIE

[KARO99] S. Karouach, T. Dkaki, B. Dousset

*Visualisation interactive de classifications d'informations*. 8<sup>èmes</sup> journées d'études sur les systèmes d'information élaborée: Bibliométrie - Informatique stratégique - Veille technologique, (Ile Rousse Corse France), CD-ROM, p. 45-61, 27 septembre-1<sup>o</sup> octobre 1999.

[ROUX99] C. Roux , B. Dousset

*Une méthode de détection des signaux faibles: application à l'émergence des Dendrimères*. Veille stratégique, scientifique et technologique : VSST'98, pp 349-357, (Toulouse, France), octobre 1998.

[DOUS00] B. Dousset, T. Dkaki, J. Mothe

*Information mining in order to graphically summarize semi-structured document*. 17<sup>th</sup> international CODATA Conference, (Baveno Italie), 15-19 octobre 2000.

[HUBE00] G. Hubert, J. Mothe, A. Benammar, T. Dkaki, B. Dousset, S. Karouach

*Textual document Mining using graphical interface*. International Human Computer Interaction, HCI International 2001 , New Orleans (USA). Lawrence Erlbaum Associates - Publishers , Mahwah - New Jersey, pp 918-922 (volume 1), 05-10 août 2001.

[KARO01] S. Karouach, B. Dousset

*Visualisation interactive pour la découverte de connaissances*. Veille stratégique, scientifique et technologique : VSST'01, Vol 1, pp 291-300, (Barcelone, Espagne), octobre 2001.

[MOTH01] J. Mothe, C. Chrisment, T. Dkaki, B. Dousset, D. Egret

*Information mining: use of the document dimensions to analyse interactively a document set*. 23<sup>rd</sup> BCS European Colloquium on IR Research: ECIR, Darmstadt. BCS IRSG, pp 66-77, 4-6 avril 2001.

[MULT01] J.-L. Multon, G. Lacombe, B. Dousset

*Analyse bibliométrique des collaborations internationales de l'INRA. Veille stratégique, scientifique et technologique : VSST'01, Vol 1, pp 261-270, (Barcelone, Espagne), octobre 2001.*

[SOSS01] D. Sosson, M. Vassard, B. Dousset

*Portail pour la navigation en ligne dans les analyses stratégiques. Veille stratégique, scientifique et technologique : VSST'01, Vol 1, pp 347-358, (Barcelone, Espagne), octobre 2001.*

[DOUS02] B. Dousset, S. Karouach

*Collaboration interactive entre classifications et cartes thématiques ou géographiques. 9<sup>èmes</sup> rencontres de la société francophone de classification, (Toulouse France), 16-18 septembre 2002.*