International collaborations in Research Institute: Bibliometric study

Odile BEDU

Inra, French National Institute for Agricultural Research, France.

Abstract

The analysis of co-publications between INRA and their foreign partners provides indicators which interest the INRA (French National Institute for Agricultural Research) and particularly the Department of International Relations. An annual bibliometric study has been led since 2005, which allows us to follow several quantitative and qualitative indicators. Data are provided by the Web Of Science which provides all the necessary information for this work in the Adress Field.
Once the location of the Inra's partner countries is complete, the corpus is constituted. The most important work is the harmonization of the writings of the addresses, essential before the statistical treatment of the data. This is realized with a software allowing the analysis of surveys, from a statistical or textual point of view. The global results confirm intuitive data: increase of the number of co-publications and broadening of the number of our partner countries. A country analysis is more relevant; it allows the listing of the partner institutions with precision. This work is led by the 20 principal countries with which Inra co-publishes; the fields of research (Subjects categories) and temporal evolution of the collaborations are also analysed. Partners networks are also identified. A regrouping of these countries, by geopolitical identity, allows to be better situate INRA within world agronomic research. The data acquired allows us to answer to requests from the Department of International Relations of our Institute, as well as from several Government Ministries.

Résumé

Les collaborations internationales dans un institut de recherche : étude bibliométrique — L'analyse des copublications entre l'Inra et ses partenaires étrangers fournit des indicateurs qui intéressent l'INRA (Institut National de la Recherche Agronomique) et particulièrement la Mission des relations internationales. C'est pourquoi une étude bibliométrique annuelle est menée depuis 2005 ; elle permet de suivre plusieurs indicateurs, quantitatifs et qualitatifs. Les données proviennent du Web of Sciences qui donne, dans le champ adresse, toute l'information nécessaire pour ce travail. Après le repérage des pays partenaires, le corpus est constitué ; un gros travail d'harmonisation des écritures des adresses des auteurs est nécessaire avant de traiter les données. Celui-ci est réalisé avec un logiciel permettant l'analyse d'enquêtes, d'un point de vue de l'analyse statistique ou textuelle. Les résultats globaux confirment des données intuitives : progression du nombre de co-publications, élargissement du nombre de nos pays partenaires ; une analyse par pays est plus pertinente : elle permet de repérer avec précision les institutions partenaires. Ce travail est mené pour les 20 pays principaux avec lesquels l'Inra co-publie, ainsi que les champs de recherche et leur évolution temporelle ; les réseaux de partenaires sont eux aussi identifiés. Un regroupement de ces pays par ensemble géopolitique permet de mieux situer l'Inra dans la recherche agronomique mondiale. Ces données permettent de répondre à des demandes de la direction de l'institut ainsi que de plusieurs ministères.

Introduction

The national Institute for Agricultural Research (INRA, France), for several years attempts to quantify the degree of international cooperation. Bibliometric analysis appears among the held indicators since the volume of articles published by Inra allows a statistical treatment. The aim is to develop knowledge on articles signed jointly by INRA and its partners of foreign countries.
Several studies have already been led by various authors (Hardy-Dessources, 2006), at the request of Department of International Relations (DIR). The methodology was revised and adapted to study a new corpus, starting from the publications of year 2006 to 2008. This corpus allows to quantify the influence

and the impact of research, and to follow several quantitative and qualitative indicators which inform with precision about Inra's partners. The main ones are the geographical origin of the contributors and their field address.

Methodology

State of the Art, bibliography

Many research institutes are interested by quantitative analysis of research performances. So, the scientific literature is numerous on the subject. In France, we know the scientific works of the Cnrs (National Center for Scientific Research, Coutrot, 2008), of OST (Observatory of Sciences and Technilogies, Filliatreau, 2008); abroad, we know the annual report of the CSIC (Spanish National Research Council) or vast study of European Union onScience in Africa (WAAST ; GAILLARD,2002). Some others may be guided directly to agricultural production: for example rice (Morooka, 2008).

Most of these bibliometric works study citations ranking to measure their influence in dissemination of information. As the DIR is interested first in the international collaborations, we looked for similar studies. The wealth of the information supplied by the addresses of the authors has been studied, in spite of its difficulties; see for example (Bador ; Lafouge, 2006) or RAMANANA-RAHARY S, 2007).

Setting up of the corpus

The first step was the choice of the bibliographical data bank. The necessity of finding a relevant information on the addresses of the authors quickly directed us on the WOS (Web of Science), only base to give all the addresses. The known consequences are that it will supply a cluster of bibliographical references resulting essentially from scientific journals, that the number of communications at congresses will be low and technical articles will be missing. On the other hand, WOS is very well-known by the scientific community and the information provided can be easily verified; the up-dating is very fast and Social Sciences are included.

A basic search query was constituted by taking into account the variability for the writing of Inra and the publication years:

[AD=(Natl* same Inst* same Rech* same Agr) or (AD=INRA*) or (OG=INRA)] and PY= (2006, 2007, 2008).

The analyze function of the Wos is used to refine the result by countries; the result can be download in a tabulated delimited format. After downloading, the consolidation of the data is important to correct the errors about countries; it is certain that such a large search query produces a lot of noise, specially among the French speaking countries, but not only (INRAN, Rome, Italy, for example). Besides, the name of the regions or states must be harmonised to keep a unique name: Scotland, Welsh, North Ireland, England for the UK. At this time, the corpus decreased from to 4446 references to 4383, i.e. 1,42%.

This method has the inconvenience of:
- not to locate the researchers who publish under the name of their Joint Research Unit, without indicating Inra. After checking on the center of Montpellier we note 5 % of errors due to this fact.
- adding to the corpus researchers' publications published in an UMR where the name of Inra appears, but the researcher may not belong to Inra. We so can locate countries which are not of first order for Inra, but it certainly does not significantly affect the list of our main for the first 20 countries with whom they work.

Corpus analysis

The software held for the data processing allows analyses of several types: textual analysis or statistic analysis, for quantitative and qualitative results. The corpus of bibliographic record is considered a statistical survey and bibliographic fields are questions; these questions can be textual or numerical, open ended or closed ended (unique, multiple choice or ranking scale options).

Dictionaries are used to reduce lexical variety, to end up to Germany, instead of the name of the German Länders for example. Relationships between several kinds of data (questions) or inside a question can be analyzed to provide cross-tabulation or mapping.

Indicators choice

Contrary to numerous bibliometric studies, the reputation of the journals, number of citations by articles, is not essential. One advantage is not to wait the 2nd year to obtain the impact factors; it allows to follow the evolution of collaborations annually, and to work with the references of the previous year. That is also why journals (source journals) are not a core information for this study. These data are studied by other working groups in the Inra; they provide this information to the General Direction.

The first level of analyse is a macro level: **countries** involved in co-publications and the **temporal evolution.**

At a second level, **Subject Categories** (SC) provide data on all activities in one area. They characterize a Source Journal; we can find until 6 SC by Journal. Even if we know they apply to the Source Journal and not the article, it is a precious indicator. We can also say they are the only one that provides information on the content of the publication. In fact, Keywords and Keywords Plus are not controlled vocabulary and the statistical dispersal is such as they do not bring any valuable information.

The **organisations addresses** is the most important of the indicators for this study; these data allow to a qualitative understanding of whom are the main partners in foreign countries. Unfortunately, it is the less immediate.

In the WOS, addresses are structured on organisation, sub-organisation, zip code city, province, and country. This is the theoretical plan, but the facts are different. The variability of the writing of the addresses must be corrected to allow a classification of number of papers per institutions inside a country.

It is, by far, the longest work to make. After the advice of experts, decision was taken to harmonize at least the first two levels of the organization name and the name of the city. Nevertheless, the last level is very interesting too. The software can help to spot and correct the errors by the textual analysis function, when the error of writing only results on an inversion of a word group in the address. Most of them must be located manually, specially when addresses are written in several languages *plus* English. Addresses are checked on web of these institutions if possible. Return to publication is sometimes necessary. They are confirmed by DRI members in charge of relationships with these countries. The work is done for the twenty top countries and for India. The result is given in global count; a fractional is difficult to establish without being able to verify exactly the institutional membership of the authors.

Main results

Several indicators suggest that the involvement of Inra in international research activities is increasing.

The most obvious is the **number of publications**; the increase is more than 29/%: it is more than the national average which is 20 % (figure OECD, 2008).

YEAR	Nb. cit.	Fréq.
2006	1279	29,20%
2007	1480	33,80%
2008	1624	37,10%
TOTAL OBS.	4383	100,00%

A second interesting indicator is concerning countries co-publishing with Inra

We notice several mains aspects.

The number of countries is quite permanent, more than one hundred. Some countries with a low count disappear or reappear according to the years. The interest of these collaborations is marginal; they are certainly joined into a wider network.

Three groups emerge:

- − The ranking of the first top countries does not change much: Usa, UK (#England), Germany, Spain, Belgium, and Italy are at the top level. They are industrial countries where scientific research strengths tend to be very strong.
- − After these six countries, the decline of the number of co-publications is very fast. After the ninth rank, this number falls under 5% of publications. The group from 5% to 1% produces a list of 21 countries; in this class, some of them have a regular progress. China, for example has a ratio at almost 3 %, a figure three times higher than in 2002; Tunisia was at 1,5% and is now at 3%. Most of these countries are belonging to OECD. For Inra, one of the questions is to know if these countries will have a strong interest for medium-term.
- − The last one represent 16% of the 4383 publications; in these, 2,4% of countries is under 0,1%.

Organisations co-publishing with Inra

The main interest of collaborations by country is to allow a fine knowledge of the organizations. After the harmonization in the writing of the addresses, the list of our partners is very precise. Information on organisations and sub-organisation is clear; we can have the information on the complete address if necessary. It allows to classify in a hierarchy our partners. If we compare to previous analysis, we can notice that the increase of the number of publications with a country is accompanied by a higher number of partners.

The list of the first partner addresses is interesting: it shows the case of very targeted collaborations [Table 1].

10 first partners addesses	Nb. cit.
Univ Ghent, Flanders Interuniv Inst Biotechnol (VIB), Dept Plant Syst Biol, B-9052 Ghent, Belgium	25
TEAGASC, Dairy Prod Res Ctr, Moorepark Fermoy, Cork, Ireland	18
Macaulay Land Use Res Inst, Aberdeen AB15 8QH, England	17
Max Planck Inst Mol Plant Physiol, 14476 Golm, Germany	17
Inst Food Res, Norwich NR4 7UA, Norfolk, England	14
Max Planck Inst Biogeochem, Jena, Germany	12
Univ Coll Dublin, Sch Agr Food Sci & Vet Med, Dublin 4, Ireland	12
Univ Helsinki, Dept Forest Ecol, 00014 Helsinki, Finland	12
Nestle Res Ctr, CH-1000 Lausanne, Switzerland	11
Rothamsted Res, Harpenden AL5 2JQ, Herts, England	11
Univ Cambridge, Dept Plant Sci, Cambridge CB2 3EA, England	11
Wageningen UR, Anim Sci Grp, Lelystad, Netherlands	11

Table 1. 10 first partners addesses

It is also interesting to see which are the other countries (or even partners) involved in these publications. This allows to visualize networks of science [Table 2]. We can also point pairs of partners.
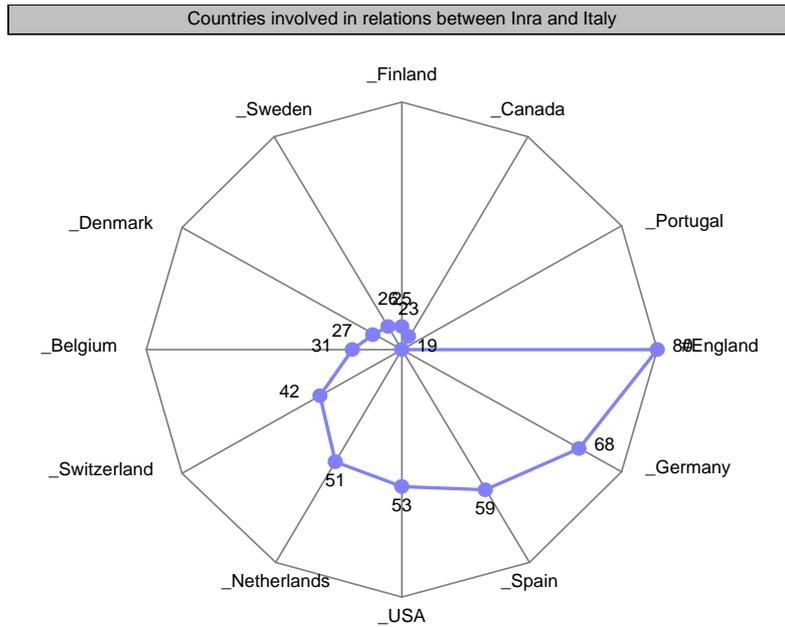
Countries involved in relations between Inra and Italy

(radial chart with countries: _Finland, _Canada, _Portugal, #England, _Germany, _Spain, _USA, _Netherlands, _Switzerland, _Belgium, _Denmark, _Sweden; values: 26, 25, 23, 27, 31, 19, 80, 68, 59, 53, 51, 42)

Table 2. Countries involved in relations between Inra and Italy

Subject Categories

The global result is of a limited interest; it is not surprising to see that the first SC are **Plant Sciences** and **Biochemistry & Molecular Biology.** But country by country[Table 3], this information is very strong, to know at the same time the strongest subjects of our partners or to know the best partners in some themes[Table 4]. It is also interesting to discover that the main thematic with a partner is not the expected one.
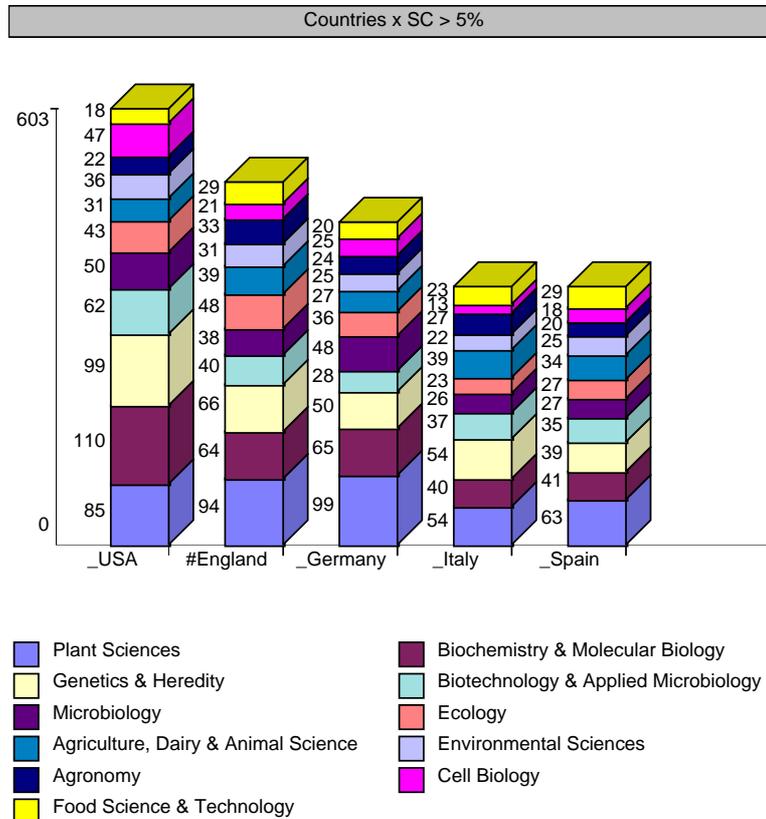

Countries x SC > 5%

(stacked bar chart for _USA, #England, _Germany, _Italy, _Spain)

Legend:
- Plant Sciences
- Biochemistry & Molecular Biology
- Genetics & Heredity
- Biotechnology & Applied Microbiology
- Microbiology
- Ecology
- Agriculture, Dairy & Animal Science
- Environmental Sciences
- Agronomy
- Cell Biology
- Food Science & Technology

Table 3. Subject Categories Countries > 5% for the top 5 countries

SC	Nb. cit.	Fréq.	SC	Nb. cit.	Fréq.
Genetics & Heredity	54	13,2%	Plant Sciences	594	13,6%
Plant Sciences	54	13,2%	Biochemistry & Molecular Biology	495	11,3%
Biochemistry & Molecular Biology	40	9,8%	Genetics & Heredity	343	7,8%
Agriculture, Dairy & Animal Science	39	9,5%	Microbiology	331	7,6%
Biotechnology & Applied Microbiology	37	9,0%	Agriculture, Dairy & Animal Science	319	7,3%
TOTAL OBS.	**409**		**TOTAL OBS.**	**4383**	

Table 4. 5 main subject categories for Italy and for the complete corpus.

The factorial analysis provides data on perspectives, publication year, country, subject categories [Table 5].
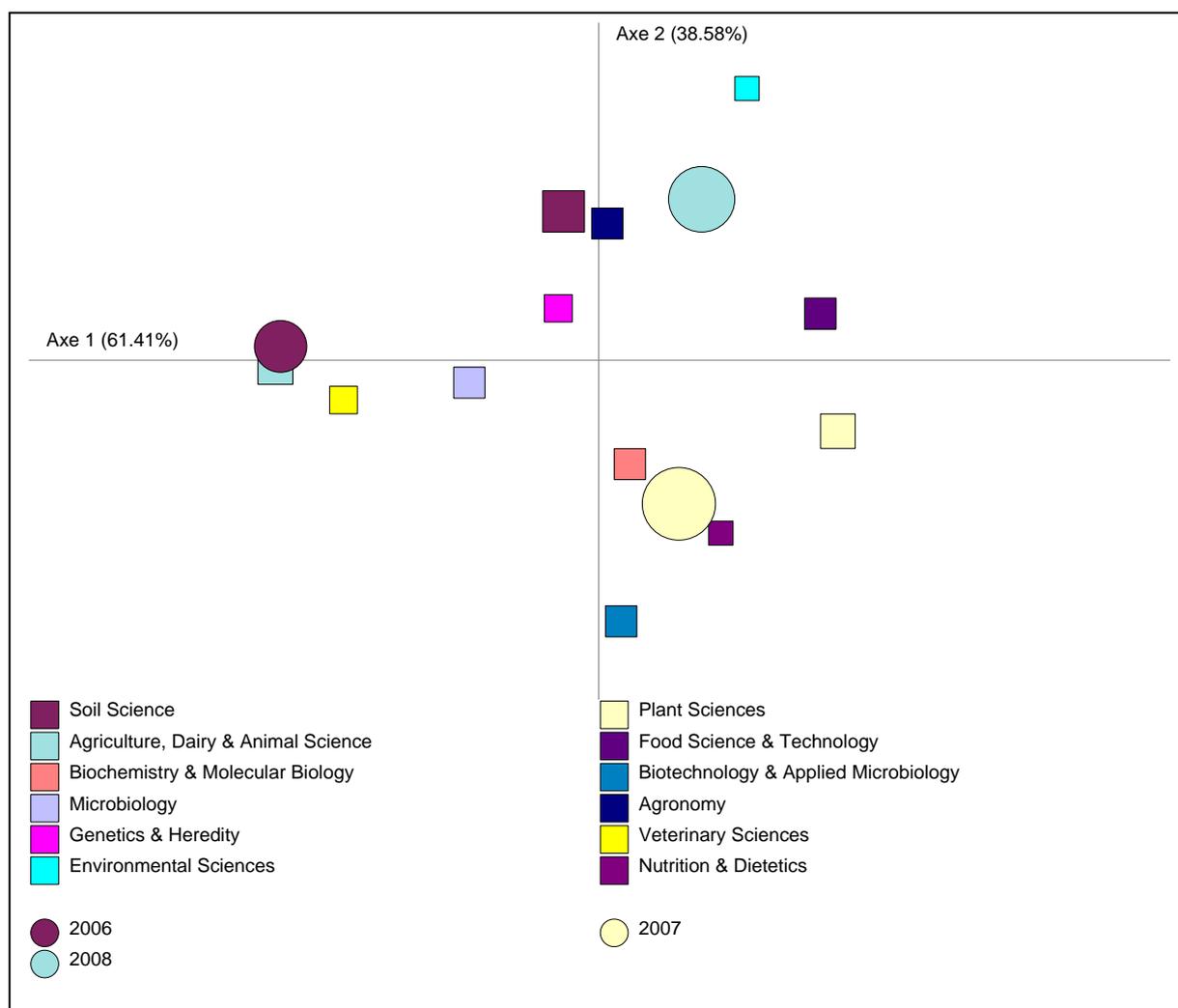

Table 5. Factorial analysis of Subject Categories for Brazil, 2006-2008

Conclusion

These indicators of collaborations are used by the DRI to check the validity of international policy: indeed, the international collaborations are driven at once on proposition of the researchers and by decision of the Management Board. The number of exchanges of researchers (in coming or out coming) and the number of grant holders welcomed in Inra consolidate these figures. All this help the direction to form new contacts or to sign new Agreements. The systematic use of Essential Science Indicators of ISI is going to increase to develop new partnerships.

References

Article

BADOR P; LAFOUGE T, 2006, La difficulté d'accéder aux adresses des chercheurs français dans les bases de données bibliographiques : l'exemple d'une université lyonnaise. *Documentaliste. Sciences de l'information* ; vol. 43, 1, pp. 28-35.

COUTROT, L. 2008. *Sur l'usage récent des indicateurs bibliométriques comme outil d'évaluation de la recherche scientifique.* Bull. Of socilogical Methodology, 100, Mis en ligne le 01 octobre 2008. URL : http://bms.revues.org/index3353.html. Consulté le 15 février 2010.

Books

WAAST, R. ; GAILLARD, J. 2002, *La science en Afrique à l'aube du 21ème siecle.* Paris (FRA), Bruxelles : MAE, CE, 2000, 172 p. multigr.

FILLIATREAU G. dir. *I*2008. *Indicateurs de sciences et de technologies : rapport OST.* Economica, 562 p.

OECD 2009. *Science, Technology and Industry Scoreboar.* OECD. 145 p.

Congress

RAMANANA-RAHARY S., 2007. Addresses of French research laboratories in the WoS. In: Affiliation institutionnelle des laboratoires de recherche et production scientifique, Séminaire ADEST, Paris, 28 septembre. 6 p.

MOROOKA K., 2008. Bibliographic analysis of Japanese Rice Research. In: Agricultural Information and IT. Proceedings of IAALD 2008, pp. 597-603.

Thesis

Hardy-Dessources O, 2006. *Analyse Bibliométrique des collaborations internationales de l'INRA à travers ses publications.* Master 2ème année Sciences de l'Information et des Bibliothèques (SIB) Option Gestion De l'Information Scientifique et TEchnique (GISTE), Université Claude Bernard, Lyon1, France. 36 p. + ann.