Analyse des publications 1998-2004 du département des Cultures pérennes (Cirad-cp)

Bilan

Martine Barale
Cirad Dic / Discom
Décembre 2005

Analyse des publications 1998-2004 du CIRAD-CP

BILAN

Sommaire

1. Contexte de l'étude 3
   1.1. La demande 3
   1.2. Le calendrier de l'étude 3
   1.3. Les intervenants 4
2. Méthodologie 4
   2.1. Les données 4
   2.2. Les outils utilisés 5
3. Résultats 6
   3.1. La relance 6
   3.2. L'analyse 7
4. Difficultés rencontrées et enseignements à tirer 7
   4.1. La relance 7
   4.2. L'analyse 9
5. Les coûts de l'étude 9
   5.1. Les coûts de personnel 9
   5.2. Les coûts de reprographie 10
6. Conclusion 10

1. **Contexte de l'étude**

1.1. La demande

Dans le cadre de la préparation de la revue externe du département CP, prévue pour le mois de novembre 2005, la Délégation à l'information et à la communication (Dic) a été sollicitée début mars par la direction scientifique du département pour réaliser un bilan des publications des agents du CP depuis la dernière évaluation (1997), c'est à dire de 1998 à 2004.
En effet, le rapport qui doit être remis par le département à la commission d'évaluation comporte, selon le modèle élaboré par la Direction scientifique du Cirad, un chapitre consacré à l'évaluation des résultats, dont la production scientifique et technique.

En préparation de l'étude, une relance des publications a été faite auprès de tous les agents encore en activité et ayant appartenu au CP à un moment ou à un autre au cours de la période couverte.

Conformément à la demande, le rapport synthétique fourni comportait des éléments chiffrés sur :

- les types de documents produits par le département,
- l'évolution de la production écrite au cours de la période considérée,

et spécifiquement pour les articles de revues :

- la proportion d'articles publiés dans des revues à facteur d'impact,
- les filières, les disciplines scientifiques et les zones géographiques concernées,
- les co-publications (partenariats).

1.2. Le calendrier de l'étude

Le directeur scientifique du CP a sollicité l'appui de la correspondante du département pour la documentation, M. Barale, fin février 2005. La demande a été officialisée le 25 mars, et a fait l'objet d'un courrier d'accord de la responsable de l'équipe Appui documentaire de la Dic en date du 5 avril 2005.

Une relance générale des publications auprès des agents du CP ayant été décidée, des messages ont été envoyés le 25 mars par la correspondante à tous les agents ayant appartenu à un moment ou à un autre à chacun des cinq programmes du département. Chaque message était accompagné de la liste récapitulative des publications référencées dans la base Agritrop pour la période 1998-2004 et dont un auteur au moins avait appartenu à un moment ou à un autre au programme concerné.

Il était demandé aux agents de vérifier ces listes (qui comportaient un index des auteurs) et de transmettre les documents manquants à la correspondante avant la fin du mois d'avril. En fait, des documents ont été reçus jusqu'à la fin du mois de mai. Ils ont tous été intégrés avant l'extraction des données définitives d'Agritrop.

Une première extraction, réalisée le 10 juin, a permis la constitution d'un fichier de travail pour le repérage des corrections et des ajouts de données manquantes à effectuer dans Agritrop. Ce fichier a également permis la fourniture, le 20 juin, d'un premier tableau de données provisoires sur la base des indicateurs précédemment proposés, afin d'obtenir la validation définitive de ces indicateurs par la direction du CP.

L'extraction du fichier définitif a été réalisée le 7 juillet 2005.
Le rapport d'étude a été fourni le 12 août. La direction du CP l'a jugé tout à fait satisfaisant et n'a pas demandé de complément.

1.3. Les intervenants

L'étude a été réalisée par M. Barale, correspondante du département CP pour la documentation au sein de l'équipe Appui documentaire de la Dic.
Elle a bénéficié de l'appui de plusieurs collègues de la Dic, qu'elle remercie :

- M.F. Fily, pour l'extraction des données d'Atlantis,
- J. Lessard et J. Péricé, pour les extractions d'Agritrop,
- M. Rouly et l'équipe catalogage, pour le traitement prioritaire des documents reçus à la suite de la relance,
- M.C. Deboin et les autres collègues d'Appui documentaire, pour leurs conseils et les retours d'expériences.

2. Méthodologie

2.1. Les données

Les références bibliographiques

Les références des publications 1998-2004 des agents du département CP ont été extraites de la base de données interne Agritrop.
Elles ont été sélectionnées à partir des auteurs (affiliation : cirad-cp) ou de la collectivité auteur (cirad-cp) des documents, et à partir de la date de publication (1998 à 2004).
Conformément à la demande formulée, tous les types de documents ont été pris en compte.

Les données sur les agents

Le système d'information sur les agents du Cirad, Atlantis, a permis de retrouver le nombre de cadres scientifiques appartenant aux différents programmes du département CP au cours de chaque année de la période couverte, nécessaire pour l'établissement des ratios de publications par an et par chercheur.
Il a également fourni les adresses email de tous les cadres scientifiques appartenant à chacun des cinq programmes du département pour chaque année de la période (et travaillant encore au Cirad au moment de l'étude) pour la réalisation des listes de diffusion des messages de relance.

Les données complémentaires ajoutées aux références issues d'Agritrop

Pour les besoins de l'étude, les références bibliographiques des articles ont été enrichies de données, soit manquantes dans Agritrop, soit nécessaires spécifiquement pour l'analyse demandée.

- Facteur d'impact de la revue

Les mentions annuelles de facteur d'impact des revues sont intégrées périodiquement dans Agritrop dans les références d'articles à partir des cdroms du *Journal Citation Report* de l'ISI. Une mise à jour a été faite fin juin, dont l'étude a pu bénéficier. Pour les années 1998 et 1999, l'attribution des facteurs d'impact a été faite rétrospectivement à partir des répertoires papier. Cependant, des contrôles et quelques compléments ont été faits dans Agritrop sur les articles les plus anciens du lot de références.

- Continent de la zone géographique étudiée

Au moment de l'indexation des documents, la documentaliste mentionne, le cas échéant, le ou les pays concernés par l'article traité. Pour permettre une exploitation plus facile de ces informations, les mentions de pays ont été « traduites » en continents (Afrique, Amérique, Asie, Océanie, Moyen Orient, Europe) et ces nouvelles données ont été ajoutées dans un nouveau champ des notices du fichier après extraction.

- Organismes partenaires pour la publication d'articles

Dans les références d'Agritrop, les noms des auteurs sont complétés par leurs affiliations (organismes dont ils font partie). Mais ces mentions sont reprises sur les documents et sont très hétérogènes, en particulier du point de vue du « niveau » pris en compte (université ou institution / département ou programme / laboratoire ou équipe). Après extraction du fichier, elles ont donc été homogénéisées au niveau le plus générique.

2.2. Les outils utilisés

La base documentaire Agritrop sous Loris

La base de données interne Agritrop existe depuis 1986. Outre les fonds des bibliothèques de site du Cirad, elle a pour vocation de référencer toutes les publications des agents du Cirad, ainsi que les travaux des chercheurs accueillis dans ses laboratoires.
Ce référencement s'appuie sur la règle du dépôt institutionnel, qui prévoit l'envoi d'un exemplaire de tous les documents produits par les agents du Cirad à leur correspondante pour la documentation.
Ces documents sont ensuite transmis à l'équipe Infothèque de la Dic, qui assure leur signalement dans Agritrop. La correspondante intervient enfin sur les notices pour enrichir la description du contenu du document (ajout de mots-clés, de codes thématiques et le cas échéant d'une traduction du titre en français) afin de faciliter l'exploitation ultérieure des données.

Le passage en 2000 de la base Agritrop sous Loris s'est accompagné d'un renforcement de la normalisation de la saisie des notices (norme Unimarc) qui permet l'exploitation plus aisée d'un certain nombre d'indicateurs bibliométriques. Plus récemment, un groupe de travail a été

constitué pour étudier les moyens de compléter ces données pour les rapprocher plus encore des types de demandes reçues par la Dic.

Dans le cadre de l'étude des publications 1998-2004 du Cirad-cp, les données exploitées ont été les suivantes :

- appartenance de l'auteur au Cirad,
- auteurs et collectivités
- affiliation (des auteurs),
- date de publication,
- indicateur complémentaire de publication (édité / publié / littérature grise),
- type de document spécifique,
- descripteur matière (filières),
- classification Agris spécifique (codes thématiques),
- descripteurs géographiques.

Le logiciel bibliographique EndNote

Les références bibliographiques extraites d'Agritrop ont été importées dans EndNote pour constituer une base de données locale sur laquelle ont été effectués les corrections et enrichissements nécessaires pour les besoins spécifiques de l'étude :

- suppression des « faux doubles » (même rapport en deux langues, même document sur deux supports),
- ajout dans un champ supplémentaire des continents à partir des mots-clés géographiques,
- ajout dans un champ supplémentaire des affiliations des auteurs non Cirad (partenaires) homogénéisées.

L'exploration des données et les comptages ont été ensuite réalisés directement sous EndNote.

Excel et Word

Les résultats ont été reportés dans un tableau Excel, et les données les plus significatives ont été traduites sous forme de graphiques.

Le rapport a été rédigé sous Word, et les graphiques importés d'Excel et placés en annexe.

3. Résultats

3.1. La relance

Une relance générale a été effectuée par messagerie sur la base de listes récapitulatives des publications 1988-2004 des agents du Cirad-cp par anciens programmes.
36 % des destinataires des messages de relance ont répondu dans un délai de 2 mois.

148 documents manquants dans Agritrop (hors rapports internes) ont été reçus à la suite de cette relance. 23 rapports internes ont été réceptionnés pendant la même période, dont 2 ou 3 rapports anciens transmis vraisemblablement à la suite de la relance.
Ces chiffres correspondent à un complément d'environ 6 % de documents fournis à la suite de la relance par rapport aux documents préalablement présents dans Agritrop.

3.2. L'analyse

2517 références ont été extraites d'Agritrop.
Après suppression de quelques doublons et d'un nombre important de « faux doubles » (même publication sur deux supports ou en deux langues différentes), un fichier de 2415 références a été utilisé pour l'analyse.

Les indicateurs suivants ont été produits :
- nombre total de publications et évolutions annuelles,
- proportions de publications vraies, littérature grise, thèses et mémoires, et évolutions,
- pour les publications vraies : proportions des différents types de documents concernés, et évolutions,
- pour les articles : proportions dans des revues avec/sans facteur d'impact ISI, nombre traitant de chaque filière, nombre dans chaque grande discipline scientifique, nombre par zone géographique, et évolutions annuelles,
- pour les articles également : nombre d'articles d'auteur(s) CP seul(s) et nombre d'articles co-écrits et évolutions annuelles ; nombres d'articles par grands types d'organismes partenaires (Cirad hors CP, Inra, Ird, Cnrs, universités et institutions en France, au nord et au sud, organisations internationales, sociétés privées).

Conformément à ce qui avait été demandé, une synthèse des principaux résultats a été rédigée à l'intention de la direction du département. En annexe étaient également fournis le tableau complet de données ainsi que quatre graphiques :
- publications CP 1998-2004 : évolution par grands types de documents (publications vraies, thèses et mémoires, littérature grise)
- publications vraies : répartition par grands types de documents (actes, communications et posters ; articles ; ouvrages et chapitres ; autres types)
- publications vraies : évolutions annuelles (en nombre de documents)
- publications vraies : proportions annuelles des différents types de documents (en pourcentage du total).

4. Difficultés rencontrées et enseignements à tirer

4.1. La relance

Le dépôt institutionnel des publications

La règle du dépôt obligatoire de toutes les publications auprès de la Documentation a été établie de longue date au Cirad. Elle a été rappelée en octobre 2003 dans une lettre officielle

du directeur de la Dist aux directeurs de départements dans laquelle la base Agritrop est positionnée comme « un des instruments du dispositif d'évaluation des équipes scientifiques ». On constate cependant que cette base souffre encore de lacunes importantes.

En conséquence, les divers travaux d'analyse des publications commandés par les équipes de recherche, les départements ou la direction du Cirad sont souvent précédés d'opérations de relance très lourdes à mener.
Dans le cas de l'étude des publications 1998-2004 du CP, environ 60 % du temps de travail de la documentaliste correspondante a été consacré à cette relance : extraction et préparation des listes de références récapitulatives par programme, établissement des listes de destinataires des messages de relance, échanges avec les chercheurs (plus de 450 messages échangés), tris, contrôles, enregistrement et transmission des publications reçues.
A cela il faut ajouter le temps consacré au traitement en urgence de ces documents par l'équipe de catalogage, qui n'est généralement pas programmé et perturbe les activités de l'équipe.

Un effort important reste donc à faire au niveau des chercheurs du Cirad pour permettre un référencement régulier et complet de toutes les productions écrites de l'établissement.

L'identification des agents

La relance a été réalisée sur la base des anciens programmes du CP (la réorganisation en unités de recherche date de janvier 2005 et n'est donc pas prise en compte dans l'étude).

Par chance, le département existe depuis longtemps, et sa structure en 5 programmes « filières » (Cacao, Café, Cocotier, Hévéa, Palmier) n'a pas changé au cours de la période considérée. Seuls quelques agents ont changé de programme entre 1998 et 2004. Leurs publications apparaissaient alors dans les listes récapitulatives des deux programmes auxquels ils avaient appartenu, et ils ont dans ce cas reçu les messages destinés aux agents de ces deux programmes.

Bien entendu, l'établissement de ces listes de publications et des listes de diffusion correspondantes aurait été beaucoup plus complexe si la période d'étude avait inclus des changements importants de la structure du Cirad ou du département.

Pour répondre à des demandes d'analyse des publications sur plusieurs années, il est absolument indispensable de pouvoir disposer d'un système d'information qui garde en mémoire, sans limitation de durée et sous une forme facilement exploitable, les changements d'affectation des agents au cours de leur carrière au Cirad.

Dans Agritrop, les affiliations des auteurs ont aussi une importance cruciale. Dans le cas des auteurs Cirad, elles sont importées d'Atlantis en tenant compte de l'année de la publication traitée. Pour les autres auteurs, elles sont reprises sur le document.

D'une manière générale, les évaluations des publications au sein des structures de recherche nécessitent obligatoirement qu'une attention particulière soit portée par les auteurs aux affiliations qu'ils font apparaître sur leurs publications. Cela est plus vrai que jamais avec le développement des équipes mixtes de recherche, qui sont très souvent mises en avant aux dépens des organismes de rattachement des chercheurs, ce qui nuit grandement à la visibilité des travaux du Cirad dans les bases de données internationales.

4.2. L'analyse

Le problème des « faux doubles » dans Agritrop

Selon la norme de catalogage Unimarc, chaque document physique doit faire l'objet d'une notice dans Agritrop. Ainsi, deux rapports de la même mission rédigés en deux langues donnent lieu à deux notices. C'est aussi le cas du compte rendu d'un congrès publié sous forme d'un document papier et d'un cdrom.
Cette procédure, qui peut sembler logique quand il s'agit de gérer un fonds physique, entraîne un gonflement artificiel du nombre de publications d'un auteur ou d'une structure, préjudiciable pour l'édition de listes de publications et dans le cadre des études bibliométriques réalisées à partir des données de la base.

Pour éviter à terme un fastidieux travail de contrôle et de suppression de ces « faux doubles » dans les fichiers extraits d'Agritrop à des fins d'analyse de publications, une réflexion devrait être menée sur les moyens d'identifier les notices concernées et sur les documents à privilégier dans un tel cas pour l'élaboration de listes de publications.

La catégorisation des articles de revues

En plus du facteur d'impact des revues dans lesquelles sont publiés les articles, qui est un indicateur chiffré mais délicat à exploiter, les demandes d'analyse des publications font souvent mention de l'existence d'un comité de lecture, qui garantit un certain niveau de qualité des articles publiés dans une revue.
Cette caractéristique spécifique à chaque revue ne fait l'objet d'aucun référentiel complet directement exploitable. Elle n'est pas actuellement entrée dans les références d'articles dans Agritrop, et doit donc, le cas échéant, être ajoutée dans les fichiers extraits pour les analyses de publications.
Là aussi, une réflexion devrait être menée pour trouver le meilleur moyen de référencer régulièrement les revues à comité de lecture.

5. Les coûts de l'étude

5.1. Les coûts de personnel

La documentaliste correspondante qui a réalisée l'étude de publications y a investi, entre le 28 février et le 12 août 2005, l'équivalent de 31 journées de travail, qui se répartissent comme suit :

Relance **17 journées**

Préparation des listes de publications récapitulatives par programme
Préparation des listes de diffusion email par programme
Préparation des messages de relance et envoi
Echanges avec les destinataires par mail, téléphone et sur place
Contrôle des documents signalés comme manquants et/ou reçus
Recherche de compléments d'information sur les documents (congrès, éditeur, ISSN)

Photocopies, impression des documents reçus sous forme électronique
Transmission pour catalogage, suivi, bilan.

Analyse **14 journées**

Interface avec le Cirad-cp, coordination Dic
Extraction préliminaire, contrôles sur index, corrections dans Agritrop
Production d'une série d'indicateurs provisoires pour validation
Extraction définitive et transfert du fichier de travail sous EndNote
Repérage et suppression des « faux doubles »
Ajout de données complémentaires dans les notices, homogénéisation
Comptages et mise en forme des données chiffrées dans Excel
Préparation des graphiques dans Excel
Rédaction du rapport synthétique.

Les investissements des autres intervenants ont été plus ponctuels, à l'exception de l'équipe catalogage qui a eu à se mobiliser dans l'urgence et à réorganiser temporairement son plan de travail pour traiter en priorité les documents CP reçus à la suite de la relance.

5.2. Les coûts de reprographie

La relance a donné lieu à l'envoi en urgence par les chercheurs d'un nombre important de documents, dont plusieurs thèses et des mémoires, souvent sous forme électronique. Compte tenu des délais assez courts et de l'éloignement d'une part importante des chercheurs concernés, aucune restriction n'a été imposée quant au support des documents transmis dans le cadre de cette relance. Il en a résulté un coût non négligeable (bien que non évalué précisément) pour l'impression sur papier de documents électroniques, dont des documents en couleur.

La généralisation de la transmission électronique des documents est de nature à faciliter le dépôt institutionnel des publications, en particulier celles des expatriés. Mais elle implique que des moyens suffisants soient alloués à l'équipe Appui documentaire de la Dic qui est le point focal de ce dépôt.

6. Conclusion

La Dic est de plus en plus souvent sollicitée pour réaliser des études de publications ou des analyses bibliométriques à partir des données d'Agritrop. Or, cette base de données n'a pas été conçue au départ pour répondre à ce type de demande.
Il faut donc poursuivre l'adaptation de cet outil à cette nouvelle fonction, qui est stratégique et qui prend de plus en plus d'importance.

Cela passe par les étapes suivantes :

- **renforcement du dépôt institutionnel**, dont l'importance est mal connue et qui, pour beaucoup de chercheurs du Cirad, n'est pas encore un automatisme,
- amélioration de l'**enregistrement des auteurs et de leurs affiliations**,

- développement d'un **système de repérage des notices multiples** d'une même production
- prise en compte des **revues avec ou sans comité de lecture**.

Les études de publication s'inscrivent dans la démarche d'évaluation des collectifs de recherche. Dans ce contexte, il est important de travailler collectivement à la définition de **critères d'évaluation des publications à la fois réalistes, véritablement significatifs et cohérents avec ceux utilisés dans les organismes proches**. Les indicateurs correspondants pourront alors être tous inclus dans les références d'Agritrop pour pouvoir être exploités à la demande.

A l'extérieur du Cirad, et en particulier dans les banques de données internationales qui font référence pour les instances d'évaluation de la recherche, la visibilité des travaux du Cirad est conditionnée par la mention de l'organisme sur les publications elles-mêmes. Un effort doit être fait dans ce domaine, en particulier dans le contexte actuel des équipes mixtes de recherche. **Les chercheurs doivent être largement informés de l'enjeu de la normalisation des adresses, et encouragés à valoriser leur appartenance au Cirad**, sous peine de voir s'effondrer la part des publications du Cirad comme des autres organismes de recherche français dans les grands référentiels bibliographiques.

Dans le même esprit d'amélioration de la visibilité des travaux du Cirad, une réflexion collective doit être menée sur la participation de la communauté Cirad dans les nouveaux vecteurs de publication (revues électroniques en accès libre) et dans le mouvement des archives ouvertes.