

## **Bibliométrie, Contrôle qualité et production d'indicateurs, automatisation sous R**

S. Cariou <sup>a</sup>, L. Frappier <sup>b</sup>

<sup>a</sup> SMART-LERECO, INRA, AGROCAMPUS OUEST, 35000, Rennes, France  
sylvain.cariou@inra.fr

<sup>b</sup> SMART-LERECO, INRA, AGROCAMPUS OUEST, 35000, Rennes, France  
lise.frappier@inra.fr

**Mots clefs :** Bibliométrie, analyse de données, appui à recherche, qualité des données.

Les professionnels de l'Information Scientifique et Technique (IST), aussi nommés documentalistes, de l'Inra ont deux missions assurer la qualité des données bibliographiques des publications déclarées dans l'archive institutionnelle de l'Inra et fournir des indicateurs de production aux différents niveaux de responsabilité.

Nos sources de données sont principalement ProdInra<sup>1</sup>, NORIA, les listes Economie-Gestion du CNRS section 37<sup>2</sup> et de l'HCERES<sup>3</sup>.

Nous projetons de développer un package R permettant d'une part de vérifier la cohérence des informations saisies dans la base de données bibliographique extraite (corpus) et de faciliter les chantiers qualité et d'autre part de produire des indicateurs bibliométriques notamment ceux préconisés dans le cadre des évaluations.

Concernant la cohérence des données, les documentalistes consacrent beaucoup de temps à uniformiser la base de données institutionnelle des publications des chercheurs de l'Inra, ProdInra. Ce travail est indispensable pour l'obtention de listes de publications et d'indicateurs les plus fiables possible [1]. Un chantier qualité sur les données de publications concerne en général un corpus de plusieurs centaines de références bibliographiques (notices), exportées dans un tableur. La mobilisation en temps est pour les professionnels de l'IST, généralement longue d'un ou plusieurs mois, selon le corpus de notices à corriger, les champs plus ou moins difficiles à analyser, la disponibilité des professionnels de l'IST impartie à cette mission. En outre, certaines de ces tâches sont répétitives et communes d'un corpus à un autre. Le package R devra permettre de réduire le temps consacré à la préparation des fichiers, et limiter les sources d'erreurs potentielles.

Concernant la production d'indicateurs, nous avons déjà automatisé la production d'un reporting contenant des indicateurs de production synthétiques. Après analyse des possibilités offertes par R, il nous a semblé opportun pour des questions d'adaptabilité et d'évolutivité des indicateurs et des sources de données à traiter de faire évoluer notre processus d'automatisation vers R.

Les principaux packages R existants sur la bibliométrie, BibliometRics [2], Bibliometrix [3] et BibeR [4] notamment, utilisent des sources de données payantes (Web Of Science et Scopus). Notre source de données est une archive ouverte institutionnelle dont la structure de données et le type de documents sont différents des sources ciblées par les packages existants. Il est à noter qu'en économie, les analyses de la bibliographie ne peuvent se baser uniquement

---

<sup>1</sup> Archive institutionnelle de l'Inra <http://prodinra.inra.fr>

<sup>2</sup> <https://sites.google.com/site/section37cnrs/Home/revues37>

<sup>3</sup> <http://www.hceres.fr/PUBLICATIONS/Documentation-methodologique/Listes-de-revues-SHS>

que sur le Web of Science (~ 50% de la production) ou Scopus (non accessible à l'Inra). Ainsi, nous travaillons sur des classements de revues validés par les pairs, ce qui complexifie le traitement des données et qui ne nous permet pas d'utiliser les packages existants<sup>4</sup>. En plus de l'automatisation d'un contrôle qualité, le package permettra de produire une série d'indicateurs bibliométriques directement mise en forme sous RMarkdown. L'objectif final est de fournir à un public de non-initiés à l'utilisation de R, une interface web leur permettant d'utiliser les deux fonctionnalités décrites.

## Références

- [1] Tang-Chaupitre C., Fouché S., Batifol-Garandel V., Gautret M., Le Hénaff D. (2012) Enjeux et mise en œuvre de chantiers qualité des données dans l'archive ouverte de l'Inra : ProdInra, Cahier des Techniques de l'INRA, Numéro spécial 2012 pp 133-140 <http://prodinra.inra.fr/record/178325>
- [2] Beguería S. (2015) BibliometRics: an R package for bibliometric analysis of scientific production, doi:10.5281/zenodo.834260.
- [3] Aria, M., Cuccurullo, C. (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11(4), pp 959-975
- [4] Liu Y., Li M. (2016) BibeR: a Web-based tool for bibliometric analysis in scientific literature. PeerJ Preprints 4:e1879v1 <https://doi.org/10.7287/peerj.preprints.1879v1>