



Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com



MÉTHODOLOGIE

L'identité institutionnelle au sein des outils bibliométriques : un enjeu stratégique



Bibliometrics and institutional identity: A strategic issue

J. Sempéré

Université Paris-Saclay, route de L'orme-aux-merisiers, 91190 Saint-Aubin, France

Reçu le 24 août 2017 ; accepté le 26 septembre 2017
Disponible sur Internet le 10 novembre 2017

MOTS CLÉS

Bibliométrie ;
Classements ;
Réputation
académique ;
Signature

Résumé Tout service bibliométrique utilise les deux principaux outils disponibles : Scopus d'Elsevier et Web of Science de Clarivate Analytics. Dans le contexte français de transformation institutionnelle, quelques institutions comme l'université Paris-Saclay ont travaillé avec ces deux entreprises pour mieux définir leur identité. Ce travail de correction des données est utile pour chaque institution afin de mieux comprendre comme les classements internationaux peuvent utiliser leurs données bibliométriques et d'exposer des données plus justes.

© 2017 Elsevier Masson SAS. Tous droits réservés.

KEYWORDS

Bibliometrics;
Rankings;
Academic reputation;
Signature

Summary Any bibliometrics services use the two main tools available: Scopus by Elsevier and Web of Science by Clarivate Analytics. In the French context of institutional transformation, an institution such as Université Paris-Saclay works with these two providers to better define its identity. This work of data curation is useful for each institution in order to well understand how the international rankings can use its bibliometrics data and to expose more correct data.

© 2017 Elsevier Masson SAS. All rights reserved.

Adresse e-mail : julien.sempere@universite-paris-saclay.fr

<https://doi.org/10.1016/j.jemep.2017.09.014>

2352-5525/© 2017 Elsevier Masson SAS. Tous droits réservés.

Pour l'ensemble des acteurs de l'enseignement supérieur et de la recherche, la réputation est devenue un critère majeur en vue de choisir au sein de quel organisme ou de quelle faculté ils effectueront leur cursus, leur post-doc ou avec quel laboratoire ils concluront un partenariat. Si les relations interpersonnelles jouent un rôle majeur et se tissent lors de colloques ou du fait de l'environnement direct du chercheur dès son doctorat (laboratoire de rattachement, directeur de thèse, proximité géographique) et mériteraient une étude prosopographique poussée, cette réputation des institutions est un des critères qui font ou défont les projets de recherche, attirent les financements ou les partenariats.

Cette réputation a été, en partie, normée au sein des classements internationaux si bien que certains d'entre eux, comme le classement QS n'hésitent pas à définir un critère de « réputation académique » [1] ou comme le Times Higher Education qui détermine un *World Reputation Rankings* à partir d'un questionnaire de « réputation académique » effectué par l'éditeur Elsevier auprès de chercheurs du monde entier [2]. Dans ce dernier cas, chaque chercheur interrogé indique les quinze institutions qui, de son point de vue, sont les meilleures en enseignement et en recherche pour sa discipline.

Cette méthodologie pose la question de la visibilité d'une institution et explique en partie la volonté d'acteurs politiques de l'Enseignement supérieur français de développer des campus pluridisciplinaires sous un même nom et de développer une image de marque afin que l'identification des structures et institutions soit plus claire de la part des acteurs hors de France. Cela s'est traduit notamment par le travail autour des noms des communautés d'universités et d'établissements (Comue) qui se sont formées ces dernières années [3].

Un des exemples les plus significatifs a été l'enjeu autour de l'usage du mot Sorbonne dans le nommage des Comues à Paris : Sorbonne-Paris-Cité (regroupant originellement notamment les universités Paris Sorbonne-Nouvelle, Paris-Descartes, Paris-Diderot et Paris Saint-Denis) et Sorbonne-Universités (regroupant avec d'autres institutions les universités Paris-Sorbonne et Pierre-et-Marie-Curie).

Cette problématique de nom de l'institution se décline à chaque niveau de la structure et a posé notamment d'importants débats sur la méthodologie de la signature des articles scientifiques et a été largement développée [4]. Cette signature est un enjeu majeur pour identifier les structures scientifiques au sein desquelles s'effectue la recherche. La question des multi-tutelles est une problématique particulièrement sensible dans le contexte français qui constitue une exception dans le paysage de la bibliométrie mondiale.

Les outils bibliométriques doivent en effet pouvoir indexer ces données de signature et donc les interpréter selon une logique de traitement de masse. Ce traitement peut s'apparenter à du *Big data*. Cela signifie que les données de signature sont analysées, autant faire se peut, de manière compartimentées et semi-automatique. L'outil doit pouvoir reconnaître le nom de l'auteur, son affiliation, le lieu physique du laboratoire ou de l'institut de recherche. Or, une affiliation peut se présenter comme suit :

« Université Paris-Saclay, université Paris-Sud, UVSQ, CESP, Inserm, Villejuif, France ». Cet exemple présente

plusieurs difficultés d'interprétation et de définition au sein des outils de bibliométrie. La virgule joue un rôle essentiel pour permettre d'analyser l'affiliation de l'auteur, cependant, la multiplication des tutelles qui ne présentent pas nécessairement de hiérarchie entre elles, compliquent l'analyse. Cet auteur, qui travaille au sein du Centre de recherche en épidémiologie et santé des populations (CESP) indique donc la structure au sein de laquelle il travaille entre le nom d'une université (université Versailles Saint-Quentin) et un organisme de recherche (Inserm).

L'analyse ne peut donc se fier sur l'ordre logique de description des tutelles, d'autant que chaque Comue peut avoir un ordre de signature différent. Elle doit permettre également d'éviter toute confusion sur des sigles qui, parfois, peuvent prêter à confusion du fait d'homonymie de structures ou de coïncidences. C'est le cas notamment concernant, en France, le nom « ENS » qui peut s'interpréter a minima comme École normale supérieure de la rue d'Ulm, de l'université Paris-Saclay ou de Lyon.

Dans l'exemple cité plus haut, la question de la hiérarchie des valeurs se pose. En effet, l'Inserm est un organisme national, le CESP une structure de recherche multitutelles, les universités Paris-Sud et Versailles Saint-Quentin, des universités et l'université Paris-Saclay, une Comue. L'outil de signalement n'a d'ailleurs pas mentionné l'ultime tutelle du CESP qu'est l'université Paris-Descartes que l'auteur a peut être mentionnée à une autre ligne, non prise en compte ici, car dépendant d'une autre Comue que l'université Paris-Saclay, en l'occurrence la Comue Sorbonne Paris-Cité.

Ainsi, en respectant les règles de signature et en admettant que les outils de bibliométrie admettent des signatures multilignes, l'affiliation pourrait être la suivante :

CESP, université Paris-Sud, université Versailles Saint-Quentin, Inserm, université Paris-Saclay, Villejuif, France.

CESP, université Paris-Descartes, Inserm, Sorbonne-Paris-Cité, Villejuif, France.

Dans ce cas, un outil de bibliométrie doit être capable de comprendre que cette publication « compte pour » la Comue université Paris-Saclay tout comme de chacun de ces membres université Paris-Sud et université Versailles-Saint-Quentin, pour la Comue Sorbonne-Paris-Cité et son membre l'université Paris-Descartes mais aussi l'Inserm comme organisme de recherche national.

Ensuite, chaque organisme de classement interprétera cette double tutelle à l'aune de son classement selon les trois possibilités les plus courantes. Dans un cas, si sont considérées deux tutelles différentes, la valeur de la publication sera divisée d'autant. Dans un second cas, la publication comptera pour chaque tutelle. Enfin, si l'outil de bibliométrie ne parvient qu'à interpréter l'organisme de recherche, l'organisme de classement ne prendra sans doute pas en compte cette publication. En effet, seules les publications liées à une institution (école, université, institut) comptent dans les classements, non celles considérées comme rattachées uniquement à un ou plusieurs organismes de recherche.

Ainsi, au sein de la Comue université Paris-Saclay qui regroupe plus de 310 laboratoires de recherche, la moitié ont pour tutelle uniquement des organismes de recherche nationaux et, en cas de classement, ne compteraient pas dans les publications de l'université Paris-Saclay en l'état. Cette situation est identique dans toutes les Comue. Certes,

cette organisation institutionnelle se construit autour de choix administratifs ou politiques. Cependant, tout cela pose une problématique d'interprétation des données pour les outils bibliométriques.

Si le chercheur doit être prévenu des conséquences de la manière dont il formule sa signature, des acteurs de l'institution doivent entreprendre un travail étroit de collaboration avec les deux leaders mondiaux de la bibliométrie que sont Elsevier avec ses outils Scopus [5] et Scival [6] et Clarivate Analytics et ses outils Web of Science [7] et Incites [8]. Ces deux fournisseurs moissonnent dans des sources qui se recouvrent en partie, essentiellement anglo-saxonnes et en sciences dures les données de publications.

Les sociétés de classements contractualisent avec ces deux sociétés pour constituer une partie des sources qui leur servent afin d'établir l'évaluation portant sur l'impact de la recherche au sein des institutions classées. Ainsi, en 2017, QS [9] et Times Higher Education [10] utilisent des données de Scopus, ARWU utilise lui des données de Web of Science [11] pour son classement global. Si les deux géants se livrent une lutte âpre pour constituer la source unique de tel ou tel classement générique ou thématique, un autre enjeu est celui de leurs clients.

En effet, il convient pour chaque institution de comprendre comment les données sont transmises et interprétées par les organismes de classement. Certains récupèrent la donnée de manière brute pour reconstituer ensuite une cartographie des publications de l'institution, d'autres utilisent les données fournies par Elsevier ou Clarivate Analytics ainsi que l'indexation qui en a été faite. C'est sur ce point précisément qu'il est primordial pour une institution d'agir sur la qualité des données fournies et sur leur interprétation.

Chacun des deux fournisseurs d'outils bibliométriques dispose d'un service de travail de la qualité des données. Ce service a la charge notamment de définir les institutions telles qu'elles apparaissent sur leurs interfaces. Scopus a pris, pour l'heure, le parti de ne rien normaliser et de filtrer l'ensemble des données fournies en signature pour les interpréter. Web of Science normalise des éléments de la signature brute afin de normaliser les différents cas de figure. Cette normalisation porte notamment sur les types d'institutions. Ainsi, « university » ou « université » sera normalisé « Univ. » permettant un premier rapprochement par ce système de troncatures.

Ensuite, la définition de la structure dépend d'un travail minutieux entrepris par l'institution elle-même. Un premier obstacle évident à ce travail et à cette collaboration est la capacité à se doter de l'outil et à échanger avec le fournisseur afin de pouvoir déclarer les contours de son institution. En effet, ces outils sont coûteux et demandent donc un réel investissement qu'on est en droit d'attendre d'une institution qui souhaite posséder une aura internationale.

L'obstacle principal réside cependant dans la capacité à proposer à définir son institution auprès d'Elsevier ou de Clarivate Analytics. Le principe de cette définition est de pouvoir détecter toute publication qui appartient à l'institution considérée et de la donner à voir dans toute sa complexité à un utilisateur extérieur : qu'il soit institution de classement ou chercheur en recherche de partenariat dans un domaine propre. Si nous considérons la signature citée précédemment : « Université Paris-Saclay, université Paris-Sud, UVSQ, CESP, Inserm, Villejuif, France », elle

présente l'avantage d'être relativement complète pour une partie des tutelles du laboratoire. La Comue Sorbonne-Paris-Cité n'a aucune chance ou presque de voir cette publication rattachée à elle sauf si elle a défini le CESP comme une institution qui lui est en partie rattachée.

Dans ce cas, la définition du lieu géographique est essentielle. Par exemple, le Web of Science ne prend pas en compte la définition brute d'une unité de recherche telle que le CESP. Il conviendra alors d'indiquer dans sa définition Sorbonne-Paris-Cité le rattachement du CESP Villejuif, France. Cette définition s'effectue dans un tableur dont les champs récupèrent les données suivantes : variante du nom de l'institution, ville d'implantation, pays. Ainsi selon ces trois informations, une instance est rattachée à une institution ou non. Les organismes de recherche français atteignent ainsi des définitions de plusieurs milliers de ligne car chaque signature qu'un chercheur rattaché à une de leurs unités de recherche a pu écrire doit trouver dans ce tableur une ligne qui la définit selon la logique des trois données précitées.

Une autre difficulté est le croisement des définitions. Toujours dans l'exemple considéré, soit l'université Paris-Saclay échange avec les autres tutelles du CESP de son périmètre et propose une définition signalant l'ensemble des tutelles du laboratoire : université Versailles Saint-Quentin, université Paris-Sud, Inserm et université Paris-Saclay, soit elle ne le fait pas et chacune de ces structures doit avoir défini la relation à ce laboratoire sous peine qu'il ne soit pas rattaché à l'ensemble de ses tutelles dans les outils de bibliométrie.

Cette situation peut ainsi devenir plus complexe dans le cas des définitions auprès de Scopus qui propose une arborescence des tutelles. Il est donc possible de dire que le CESP est une structure fille de l'université Paris-Sud, elle-même structure fille de l'université Paris-Saclay dont elle est membre pour l'ensemble de ses laboratoires. Dans ce cas, il y a un héritage qui permet d'améliorer la recherche d'exhaustivité des déclarations de tutelles et donc la qualité des données sur les outils bibliométriques.

Cependant, il y a de nombreux écueils à cela, notamment la difficulté des acteurs sur le terrain à maîtriser les incidences des définitions effectuées aux différents niveaux : Comue, organisme de recherche, institution universitaire et grandes écoles. En effet, il peut y avoir des contradictions quand les périmètres ne sont pas clairement définis et il est difficile de comprendre comment les outils bibliométriques interprètent des déclarations contradictoires.

Partant de ce constat, il est utopique d'imaginer la définition parfaite qui prendrait en compte l'ensemble des publications d'une structure du fait de l'incertitude de la complétude des signatures comme des définitions de celles-ci. Il convient d'une part d'effectuer une pédagogie sur les incidences de la qualité de la signature auprès des acteurs de la recherche. D'autre part, les professionnels de l'information scientifique et technique ont tout intérêt à poursuivre une double dynamique :

- travailler en réseau à toutes les échelles afin de proposer des définitions communes de leurs différentes structures ou des méthodologies qui évitent redondance et contradictions ;
- travailler conjointement avec les fournisseurs des outils pour permettre de faire comprendre le contexte de

changement institutionnel en cours et d'améliorer la qualité et la rapidité de la prise en compte du contexte français.

Ce n'est qu'au prix de ce travail sur la qualité des données que les institutions françaises se donneront mieux à comprendre à l'échelle internationale sur des outils bibliométriques utilisés de plus en plus par les grandes institutions comme des outils de pilotage ou de partenariat et dégageront une compréhension suffisante de la destinée de leurs données sur ces plateformes afin d'engager un travail d'approfondissement et de traitement des données à l'échelle des organismes de classement.

Déclaration de liens d'intérêts

L'auteur déclare ne pas avoir de liens d'intérêts.

Références

- [1] QS Intelligence Unit. (page consultée le 26 septembre 2017) [En ligne <http://www.iu.qs.com/university-rankings/indicator-academic/>].
- [2] Times Higher Education (page consultée le 26 septembre 2017). [En ligne <https://www.timeshighereducation.com/world-university-rankings/world-reputation-rankings-2017-methodology/>].
- [3] Ministère de l'Enseignement supérieur et de la Recherche, (page consultée le 26 septembre 2017). Les regroupements universitaires et scientifiques : une coordination territoriale pour un projet partagé [En ligne <http://www.enseignementsup-recherche.gouv.fr/cid69475/les-regroupements-universitaires-et-scientifiques-une-coordination-territoriale-pour-un-projet-partage.html>].
- [4] Boudry C, Durand-Barthez M. Publications en libre accès en biologie-médecine : historique et état des lieux en 2016. *Ethics Med Public Health* 2017;3:169–81 [En ligne <http://linkinghub.elsevier.com/retrieve/pii/S2352552517300233/>].
- [5] Scopus. (page consultée le 26 septembre 2017). [En ligne <https://www.elsevier.com/solutions/scopus>].
- [6] Scival. (page consultée le 26 septembre 2017). [En ligne <https://www.elsevier.com/solutions/scival>].
- [7] Web of Science (page consultée le 26 septembre 2017) [En ligne] <https://clarivate.com/products/web-of-science/>.
- [8] Incites (page consultée le 26 septembre 2017) [En ligne <https://clarivate.com/products/incites/>].
- [9] Scopus (page consultée le 26 septembre 2017). QS renews agreement to use Scopus data for its World University Ranking. [En ligne : <https://blog.scopus.com/posts/qs-renews-agreement-to-use-scopus-data-for-its-world-university-ranking>].
- [10] Scopus (page consultée le 26 septembre 2017). Times Higher Education choose Scopus data for its WORLD University Ranking. [En ligne : <https://blog.scopus.com/posts/times-higher-education-choose-scopus-data-for-its-world-university-ranking>].
- [11] Academic Ranking of World, University, (page consultée le 26 septembre 2017). Methodology for Shanghai ranking's Global Ranking of Academic Subjects [En ligne <http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/Methodology-for-ShanghaiRanking-Global-Ranking-of-Academic-Subjects-2017.html>].