

Éloge de la complexité : la signature des chercheurs et le millefeuille de l'affiliation institutionnelle dans les processus d'évaluation de la recherche

Praise of complexity: Researchers' signature and the layer cake of the organizational affiliations in the research evaluation process

David Pontille, Directeur de recherche au CNRS, Angèle Séné, Véronique Prêtre, Nathalie Pothier, Alina Deniau, Manuel Durand-Barthez^a, Françoise Girard

^a Institut interdisciplinaire de l'innovation (UMR CNRS 9217), PSL Research University, centre de sociologie de l'innovation, Mines ParisTech, 60, boulevard Saint-Michel, 75272 Paris cedex 06, France

^b Irfu, CEA, université Paris-Saclay, 91191 Gif-sur-Yvette, France

^c Direction de la recherche, CentraleSupélec, université Paris-Saclay, 57070 Metz, France

^d Service information scientifique et technique, OSUC, université d'Orléans, CNRS, 1A, rue de la Férollerie, 45071 Orléans, France

^e Direction de la science, de la technologie et de l'innovation, OCDE, 2, rue André-Pascal, 75775 Paris cedex 16, France

^f Unité régionale de formation à l'information scientifique et technique (URFIST), École nationale des Chartes, PSL Research University, 17, rue des Bernardins, 75005 Paris, France

^g Laboratoire de chimie moléculaire, école polytechnique, CNRS, université Paris-Saclay, 91128 Palaiseau cedex, France

Résumé

La problématique des affiliations dans les publications scientifiques se pose de manière de plus en plus aiguë, à l'heure où le paysage qui structure la recherche se complexifie et alors que les enjeux autour de l'évaluation donnent la mesure de l'hétérogénéité croissante dans la construction d'indicateurs fiables pour les établissements de l'enseignement supérieur et de la recherche.

Summary

This study examines the role of the academic affiliation in research evaluation. The increasing structural complexity of the French research landscape has a massive influence on the construction of reliable indicators for the purposes of research assessment, one of the most damageable downside being the heterogeneity of the bibliometric indicators. The study aims to make the inventory of practices and to suggest solutions, sometimes relying on technology, to overcome the visibility issue of the French research and higher education institutions.

Mots clés

- Bibliométrie;
- Évaluation de la recherche;
- Citation;
- Affiliation;
- Enseignement supérieur

Keywords

- Bibliometrics;
- Research evaluation;
- Citation;
- Affiliation;
- Higher education

Avec la montée des politiques d'évaluation de la recherche, associée au développement des bases de données qui recensent les articles scientifiques (notamment le Web of Science et Scopus) et des classements internationaux des universités (Shanghai, Leiden, U-Multirank...), l'affiliation institutionnelle des productions est devenue une question cruciale. Elle cristallise non seulement des débats sur la visibilité internationale et l'attractivité des organismes de recherche, mais elle s'immisce également dans le quotidien des chercheurs qui, tout spécialement en France, sont désormais pris dans un véritable millefeuille institutionnel [1]. Chaque niveau érigeant sa logique propre (parfois sous forme d'une charte), les chercheurs font face à des injonctions contradictoires : d'une recommandation à l'autre, la priorité est donnée à l'organisme de tutelle (CNRS, Inserm, Inra...), à l'établissement (universités, grandes écoles...), à l'unité de recherche (UMR, UPR), au laboratoire, à l'équipe, ou encore, plus récemment apparues dans le paysage français, aux structures de financement liées aux investissements d'avenir et d'excellence (LabEx, IdEx...) et à la Communauté d'universités et établissements (Comue). Dans un tel empilement, susceptible d'être recomposé à chaque remaniement politique, la visibilité individuelle des chercheurs tend parfois à s'effacer et celle des entités dont ils dépendent est loin d'être optimale. En 2012, un état des lieux (diffusé sous le nom de « Rapport Doccitanist ») avait déjà largement rendu compte de cette situation [2].

Affiliation ... ou l'éloge de la complexité

Si la publication reste le canal privilégié des chercheurs pour communiquer leurs résultats et contribuer à la diffusion internationale de la connaissance scientifique et technique, la signature et l'affiliation en sont les gestes emblématiques. Dans les faits, signature et affiliation ont évolué, mais dans quelle mesure ? Les auteurs n'ont pas toujours conscience de la portée de leur signature et de l'impact réel de ces fameuses affiliations. La technologie en aval est-elle au point pour interpréter ces données, qui sont de véritables clés d'identification des publications ? Et surtout s'agit-il ou non ici, d'un questionnement légitime, dans le contexte de foisonnement de la recherche, ébranlée prioritairement par des préoccupations économiques et des enjeux de rayonnement international ?

À en juger par l'insatisfaction des producteurs d'information scientifique et technique en déficit de reconnaissance, les difficultés chroniques auxquelles sont confrontés les éditeurs et professionnels de l'information pour traiter et repérer les publications dans les bases de données commerciales, on peut s'interroger d'un point de vue purement pragmatique. Car les

procédés actuellement disponibles pour analyser ces briques élémentaires des publications sont infiniment plus complexes qu'on ne l'imagine.

Effectivement, pour être correctement exploitée, l'affiliation doit être clairement libellée et au mieux normalisée selon un modèle. La prolifération de chartes institutionnelles, s'avérant parfois paradoxales pour une même unité de recherche, ne favorise pas cette normalisation.

L'analyse dépend aussi des protocoles d'identification et d'interprétation propres des éditeurs de bases de données et de leurs outils. Les modèles multilignes qui avaient émergé afin de s'adapter aux règles des éditeurs et de contourner les limites imposées des outils par la reconnaissance d'une unique tutelle (« Main Organization »), par défaut universitaire, ont artificiellement démultiplié les lignes d'adresses sur une même publication.

Or, au stade actuel, les mécanismes des bases bibliométriques de référence restent relativement inadaptés au traitement d'adresses hétérogènes et protéiformes, sans intervention additive ou corrective.

Comme les chercheurs, les professionnels de l'information scientifique et technique sont interpellés par cette problématique dans les laboratoires, les directions des établissements et organismes, quand ils sont sollicités pour fournir des indicateurs bibliométriques (appel à projet, investissement d'avenir, contrat européen, rapport HCERES ...), car ils se heurtent à un processus compliqué dès qu'il s'agit de constituer des corpus de publications ciblés à l'aide des affiliations et de calculer des indicateurs fiables. Des solutions techniques sont souvent développées localement par les organismes, mais elles demeurent inefficaces dans une stratégie de visibilité internationale à long terme.

Face à ce constat, les réseaux Renatis et Urfist de Paris se sont associés en 2013, pour mener une réflexion et des actions sur ce sujet primordial. Ce travail a conduit à la publication de deux articles : « Le millefeuille des affiliations françaises dans les publications scientifiques » [2] et « Le millefeuille des affiliations : où en sommes-nous ? » [3], ainsi qu'à l'organisation d'un colloque sur ce thème en juin 2015 [4], en présence d'acteurs majeurs de l'information scientifique et technique. Le groupe de réflexion a abordé la question des affiliations sous un angle original, à la fois technique et politique en mettant en avant un certain nombre de préconisations.

La signature scientifique : un geste multiple

La signature des articles scientifiques est la pratique par laquelle les savants ont historiquement attaché leur nom propre aux comptes rendus des expérimentations qu'ils produisaient [5]. Associée au développement des sociétés savantes et des premières revues, la signature était d'emblée un geste multiple. Accompagnant l'annonce de résultats originaux, préalablement validés par les pairs, elle est tout d'abord liée à l'« enregistrement » de la priorité des découvertes. Ce lien est attesté sur le long terme, de la tenue des registres consignnant la présence des témoins aux expériences publiques du XVII^e siècle [6] jusqu'aux cahiers de laboratoire contemporains [7]. La signature est ensuite articulée à une logique de l'« adresse ». Pointant vers le nom de ceux qui endossent le crédit et la responsabilité des articles, elle entretient un rapport direct avec le réel. Contrairement à un texte littéraire (roman, nouvelle, poème...) dont le nom de l'auteur, réel ou fictif, suffit généralement à son identification, un article scientifique inclut nécessairement une adresse effective afin que son origine géographique soit clairement identifiée et localisée dans le monde [8]. Autrement dit,

un article scientifique n'est pas seulement, comme de nombreux autres documents, adressé à certains auditoires, il désigne également une personne physiquement traçable et conduit vers un lieu de production précis. Enfin, la signature est intimement une affaire d'« affiliations ». Signe du lien élaboré entre chercheurs, elle donne les filiations intellectuelles au sein d'un domaine de recherche, par le truchement de la cosignature et de la citation. Simultanément, à travers la logique de l'adresse, elle décline également l'affiliation institutionnelle des travailleurs de la preuve.

Cette multiplicité des registres d'action de la signature en fait à la fois toute la richesse, mais aussi toute la complexité. À partir du moment où les publications ont été de plus en plus régulièrement cosignées [9], la confusion s'est installée tant sur le versant de l'enregistrement que sur celui de l'affiliation. La cosignature rend plus difficile l'identification des contributions respectives des signataires, mais elle multiplie simultanément les origines : à l'équation simple « un article, un auteur (origine épistémologique), une institution (origine géographique) » se substitue une équation à plusieurs inconnues « un article, x auteurs, y institutions ». En outre, avec l'essor des collaborations internationales, les valeurs de x et y ne sont pas nécessairement équivalentes, un même auteur pouvant mentionner deux institutions, ou bien deux départements dans la même institution, ou encore deux adresses dans la même institution et une troisième dans une autre institution. Cette situation a considérablement réorienté les enjeux vers les modalités d'indexation des publications¹.

Une indexation controversée

L'indexation des articles s'accompagne cependant de questions qui ne sont pas totalement inédites, ni spécifiques au cas français. À partir des années 1960, plusieurs chercheurs en sciences biomédicales ont émis des réserves, voire formulé des critiques, à l'égard des pratiques en vigueur. Certains dénonçaient le processus de semi-anonymisation engendré par l'allongement des listes de noms, estimant que la visibilité de leur contribution n'était plus suffisamment garantie [11]. Parallèlement, d'autres s'inquiétaient des problèmes d'indexation et de citation, en particulier dans des catalogues comme l'*Index Medicus* qui, publié de 1879 à 2004 par la National Library of Medicine (NLM), compilait les sommaires d'environ 2600 revues de sciences de la vie et de médecine, d'abord américaines, puis internationales².

Alors que le champ « auteur » de l'*Index Medicus* contenait les noms de tous les signataires d'un article de 1960 à 1963, il était limité aux trois premiers signataires de 1964 à 1977 afin d'endiguer le phénomène des cosignatures. Mais face aux pratiques des chercheurs, assorties de revendications, il a à nouveau été étendu à tous les noms de 1978 à 1983. L'indexation des citations dans la base de données Medline a commencé en 1966, sans restriction du nombre de noms jusqu'en 1983. En revanche, à partir de 1984, face à l'inflation exponentielle de la cosignature et à l'apparition des articles comprenant plus de 100 noms [12], la NLM prenait une mesure commune pour l'*Index Medicus* et Medline consistant à ne retenir que les dix premiers noms, suivis systématiquement de « et al. » lorsque les articles en comprenaient onze et plus [13]. De la sorte, tous les noms supplémentaires étaient ignorés dans le processus d'indexation. Au milieu des années 1990, la NLM cédait à nouveau à la pression de chercheurs (notamment en sciences de la vie), en étendant sa politique d'indexation dans Medline de 10 à 25 noms [14]. À partir de 1995, les 24 premiers signataires et le dernier nom de chaque article étaient indexés, tandis que les noms supplémentaires entre ces deux bornes étaient remplacés par « et al. ».

Parallèlement, les revues biomédicales cherchaient à économiser de la place, notamment lorsqu'elles publiaient des recherches cliniques impliquant divers centres à l'échelle internationale. Plutôt que de lister l'ensemble des noms des nombreux cliniciens engagés dans l'étude, seuls quelques uns figuraient comme signataires, quand les autres étaient uniquement mentionnés dans une annexe finale de la publication. Ces noms n'apparaissant ni au sommaire du numéro, ni sous le titre de l'article, ils bénéficiaient d'une faible visibilité. Certains chercheurs s'y sont explicitement opposés : « Maintenant, lorsqu'il peut y avoir une vingtaine d'auteurs sous le titre d'un papier, l'identité de la plupart disparaît dans le désert de l'index annuel des auteurs (...) Il peut y avoir n'importe quel nombre d'auteurs – 200 n'est pas inhabituel – dans une annexe ou une liste de remerciements. Ce système peut être un coup mortel porté aux nombreux « als » et « autres », comme certaines revues les désignent dorénavant. » [15].

Bien qu'ils signaient la publication, les noms de certains participants étaient regroupés sous un sigle abrégatif comme « le groupe de recherche X », « et al. » ou « autres ». Dans ces conditions, les chercheurs concernés estimaient très difficile de faire valoir leur contribution à la recherche et donc de faire figurer ces publications dans leur curriculum vitae.

L'essor des noms collectifs

Ce phénomène de semi-anonymisation s'est amplifié avec l'emploi de noms collectifs, non plus à côté, mais à la place même des noms des chercheurs : par exemple, 22 % des 185 articles publiés dans la catégorie « contributions originales » du *Journal of the American Medical Association* en 2001, contre 6 % des 172 articles publiés dix plus tôt [16]. Avec le développement des projets multicentriques et pluridisciplinaires, l'usage des noms collectifs est devenu courant pour désigner les différentes études réalisées aussi bien en recherche clinique qu'en génomique, tels *The World Health Organization*, *the Women Health Initiative*, *the Arabidopsis Genome Initiative*, ou encore *the International Human Genome Sequencing Consortium*.

Certains chercheurs estimaient que cette option est particulièrement adéquate pour signaler le partage du crédit à égalité entre les différents contributeurs à une recherche multicentrique [17]. Or, pour d'autres, les noms collectifs posaient des problèmes d'indexation des informations dans les bases de données bibliographiques avec des conséquences sensibles sur la visibilité des contributions [18]. Dans Medline, jusqu'en 2001 les noms collectifs n'étaient pas indexés dans le champ « auteur », mais à côté du « titre » de l'article, seuls figuraient comme « auteur » les noms de personnes qui étaient éventuellement listés. Désormais, une rubrique « nom collectif » est disponible. Dans le *Science Citation Index*, les articles sont indexés de différentes manières, soit avec le nom collectif (par exemple : Ischemic Optic Neuropathy Decompression Trial), soit avec son abréviation (IONDT), soit avec le premier nom du « writing committee », soit avec le premier nom de la liste des investigateurs. En conséquence, un même article est sujet à différentes modalités de citation et le calcul de son impact est nécessairement très variable, voire approximatif.

L'emploi des noms collectifs, proscrit dans la plupart des milieux scientifiques, afin d'éviter toute dilution du crédit et de la responsabilité individuelle est, en revanche, une pratique courante en physique des particules. Le nom de chaque collaboration (e.g. *ATLAS Collaboration*, *CMS Collaboration*) est systématiquement inscrit sur l'ensemble des productions écrites : notes internes, rapports intermédiaires, articles de revues, communications, thèses de doctorat... Il pointe simultanément vers l'instrument monumental

qui sert à réaliser des expérimentations et l'assemblée des chercheurs qui l'ont conçu, construit, calibré, et qui le maintiennent en état de marche jour après jour. Mais ce n'est pas tout. L'importance du nom collectif est telle dans les collaborations contemporaines de physique des particules que celui-ci figure dorénavant seul sous le titre des articles et comme adresse électronique du « corresponding author ». Il instaure une personne morale, constitutive de l'origine géographique et épistémologique qui garantit aux expérimentateurs, employés par diverses institutions et engagés dans un projet de longue haleine, de parler d'une seule et même voix. Chaque occurrence du nom collectif s'articule toutefois à une liste qui compte des milliers de noms de chercheurs (y compris les membres décédés) et de leurs institutions de rattachement. Ainsi, la publication proposant l'estimation la plus précise du boson de Higgs par les collaborations ATLAS et CMS du CERN comporte 5154 signataires. On imagine sans peine le casse-tête que peut représenter une indexation sans erreur de ce seul article...

De l'affiliation... à l'évaluation

À l'échelle supérieure, la problématique des affiliations concerne deux aspects : l'un, spécifiquement documentaire, l'autre étant directement lié à l'évaluation de la recherche, de ceux qui la font, de l'organisme qui les encadre, en termes de positionnement et de caractérisation.

Plusieurs acteurs sont donc impliqués dans cette problématique, rappelons ici leur rôle respectif :

- les chercheurs sont les auteurs des publications et parfois, de manière subjective ou unilatérale, prescripteurs des affiliations ;
- les organismes/établissements sont les prescripteurs des chartes de signature, et acteurs de l'évaluation pour leur structure ;
- les tutelles ministérielles pilotent la stratégie pour la recherche au niveau national ;
- les professionnels de l'information scientifique et technique au sein des laboratoires ou des établissements, sont chargés de la collecte des corpus de publications et de l'établissement des indicateurs ;
- les producteurs de bases de données, sont fournisseurs des outils utilisés pour la bibliométrie et contributeurs du signalement des publications. Ils peuvent cumuler plusieurs rôles dont celui d'éditeurs (ex. Elsevier) ;
- les éditeurs des revues imposent parfois un format qui leur est propre en matière de libellé des affiliations.

Les objectifs des acteurs sont souvent différents, et leur conjugaison peut aboutir au contraire des effets escomptés ; on distingue notamment :

- les préoccupations identitaires et de reconnaissance au sein des communautés scientifiques pour les chercheurs ;
- le renforcement de leur présence dans les classements internationaux pour les organismes et les établissements ;

- le besoin d'indicateurs de positionnement de la recherche nationale dans le concert mondial pour les instances ministérielles et les tutelles ;
- le besoin de lever les difficultés de l'identification des corpus documentaires dans les bases de données pour les professionnels de l'IST ;
- la production de services aux clients dans le cadre d'une vision commerciale et rentable, pour les producteurs de bases de données ;
- l'alignement sur le modèle anglo-saxon du traitement des affiliations pour les éditeurs de revues et/ou producteurs de bases de données, face au système complexe d'affiliations français qu'ils ne comprennent pas toujours.

Les professionnels de l'information gèrent des opérations de recensement des publications scientifiques de leur institution en vue de l'évaluation. Ces opérations sont particulièrement délicates et rendues très difficiles du fait de l'immense variété d'intitulés d'affiliations mentionnés en tête des publications, abréviations, sigles, entités à géométrie variable et nécessitent l'élaboration de stratégies chronophages en vue de simuler et restituer toutes les modalités sémantiques imaginables en l'espèce.

Par ailleurs, l'usage ou le développement d'outils sophistiqués de retraitement de l'information, parfois coûteux et d'une ergonomie peu intuitive, est généralement indispensable. Or, temps, moyens financiers, matériels et humains font souvent défaut, a fortiori lorsque les délais impartis pour exécuter ces tâches sont excessivement réduits.

Il est clair que dans une logique de classement, le recensement des publications, i.e. des preuves patentes du travail effectué par les chercheurs, est capital. Si le recensement est effectué dans des conditions d'incomplétude notoire trop souvent teintée d'un flou inadmissible, la catégorisation et le positionnement des chercheurs sont directement et négativement impactés. En ce sens, le problème est loin d'être exclusivement technique et documentaire. Il est déontologiquement important car les conséquences d'un travail de collecte, d'inventaire et de dénombrement inaboutis se répercutent inmanquablement sur l'appréciation qu'exerce une autorité administrative décisionnaire en regard du travail fourni par les unités de recherche et les entités qui en ont la tutelle.

Les éditeurs et producteurs commerciaux ont leur part de responsabilité dans cette situation

Avec l'avantage d'un certain monopole, les producteurs de bases de données bibliométriques fonctionnent sur des schémas qui leur sont propres. Tout au long du cycle éditorial, l'affiliation indiquée dans un article peut subir maintes transformations susceptibles d'engendrer des erreurs d'interprétation ou d'affectation vis-à-vis des signataires de la publication. Les analyses bibliométriques erronées du fait de distributions algorithmiques, erratiques ou aléatoires, ou de manipulations éditoriales, portent préjudice à l'évaluation.

Certes, les éditeurs déploient des solutions conséquentes sur le plan technique, mais elles restent figées dans une logique qui ne répond pas forcément à l'attente concrète des usagers qui sont leurs clients. Par exemple, le problème de repérage de plusieurs organismes ou établissements dans une même ligne d'adresse ne peut être significativement amélioré que si les éditeurs partent du principe qu'un article doit être affecté à toutes les tutelles signataires.

Ce paradigme reviendrait à rendre caduque la notion d'unique organisation principale (« Main Organization ») prise en compte dans l'adresse, qui implique quelques contournements d'usage.

Sur le marché, on constate pourtant que certains logiciels d'analyse de données n'ont aucune peine à extraire les informations requises des fichiers de sortie de ces bases bibliométriques commerciales, pour fournir des statistiques fiables (Intellixir...). Par ailleurs, les organismes de recherche sont contraints de développer ensemble ou individuellement leurs propres outils d'identification des publications à leur niveau (comme *Opalia* et *Sampra*, abordés plus loin dans une analyse des « initiatives internes »).

De la théorie à la pratique

Prenons des exemples récents de trois Comue(s) qui ont publié leurs recommandations.

La place de la Comue dans la signature est souhaitée :

- soit en dernière position : recommandations de l'université Paris-Saclay (05/2015) : « nom de l'unité de recherche, liste des tutelles membres, Université Paris-Saclay, adresse postale de l'unité » [19] ;
- soit en deuxième position : Recommandations de PSL (10/2015) : « Nom de l'établissement, PSL Research University, [institut ou département], EPST, [numéro unité mixte], laboratoire, [équipe], [adresse], code postal, ville, France » [20] ;
- soit en première position : recommandations de Sorbonne universités (11/2015) : « Sorbonne Universités, Nom établissement ou organisme tutelle 1, Nom établissement ou organisme tutelle 2... Nom de l'Unité » [21].

L'argument invoqué pour favoriser telle ou telle position dans l'affiliation serait systématiquement un meilleur repérage par les agences de bibliométrie et de classement. Qui a raison ? Qui a tort ? Quelle sera finalement la meilleure stratégie ? Que fera le chercheur d'un laboratoire qui a pour tutelles deux Comue(s) ?

Du côté des producteurs de bases de données

Afin de mieux illustrer la problématique des affiliations sur les deux principales bases bibliométriques utilisées dans les institutions, visualisons des exemples à partir du Laboratoire d'astrophysique de Marseille et ses « dérivés ». Nous verrons la complexité de son identification malgré les velléités de regroupement.

Pour Web of Science

Effectuons une recherche sur le module « Organization enhanced » censé réunir plusieurs affiliations assimilables sous une appellation commune. Saisissons « Marseille ». En queue de la liste des 18 réponses, nous retenons la forme anglaise générique : « University of Aix-Marseille ». En la développant, nous obtenons une très longue liste de plusieurs dizaines de composantes. Parmi elles, nous relevons deux séries totalisant 16 dénominations abrégées désignant le Laboratoire d'Astrophysique de Marseille qui figure dans les champs « Address ». Une première série commence par les trois lettres LAB ; une autre commence

par l'acronyme LAM. Entre les deux, l'ordre alphabétique fait apparaître la forme unique « Laboratoire Astrophys Marseille ».

LAB ASTROHYS MARSEILLE

LAB ASTRON MARSEILLE

LAB ASTROPHYS

LAB ASTROPHYS MARSEILLE

LAB ASTROPHYS MARSEILLE LOOM

LAB ASTROPHYS MARSEILLE TRAVERSE SIPHON

LAB ASTROPHYS MARSEILLE TRAVERSE SIPHON LES TROIS

LAB ASTROPHYS MARSEILLE UMR 6110

LAB ASTROPHYS MARSEILLES

LAB ASTROPHYS MARSIELLE

LAB ASTROPHYS MARSILLE

LABORATOIRE ASTROPHYS MARSEILLE

LAM

LAM LAB ASTROPHYS MARSEILLE

LAM OAM

LAM TRAVERSE SIPHON

Total : 813 réponses (le 24 mai 2016). Au vu des références les plus récentes (février et mai 2016) la mention « Organization enhanced » qui regroupe ces variantes est University of Aix-Marseille. Examinons maintenant d'un peu plus près la mention LAM OAM, sous-ensemble d'OAMP (M pour Marseille et P pour Provence). L'Institut Pythéas (Observatoire des Sciences[NP2] de l'univers de l'université d'Aix-Marseille, Unité Mixte[NP3] de Services UMS 3470) a remplacé en 2012 l'Observatoire Astronomique[NP4] de Marseille-Provence qui regroupait 3 unités : le Laboratoire d'Astrophysique[NP5] de Marseille (LAM), l'Observatoire[NP6] de Haute-Provence (OHP) et le département Gassendi.

•Par le biais de l'« Organization enhanced », sous University of Aix-Marseille, avec l'équation : (INST PYTHEAS AND University of Aix-Marseille) OR (OBSERV HAUTE PROVENCE CNRS PYTHEAS UNIV AIX MARSE AND University of Aix-Marseille) OR (OSU PYTHEAS AND University of Aix-Marseille) nous obtenons 5 réponses.

•En saisissant : « pytheas » tout court dans le champ « Address » nous obtenons 43 réponses

•En limitant les 813 réponses obtenues ci-dessus avec l'ensemble des variantes de signatures à la période 2012–2016, soit depuis la création de l'institut Pythéas, nous obtenons 135 réponses contre 43 avec Pythéas tout court dans le champ Address et 5 en « Organization enhanced ».

Nous constatons aisément le grand différentiel entre la combinaison des variantes de signatures correspondant aux composantes de l'Institut Pythéas et la saisie du nom « officiel » de cette entité englobante tant en « Organization enhanced » qu'en champ adresse « brut ».

Examinons une d'entre elles :

•Inst Pytheas, UMR CNRS 7249, MIO, F-13288 Marseille 09, France.

Elle comporte le numéro d'UMR 7249, tandis que nous connaissons également l'Institut sous le numéro d'UMS 3470 (voir ci-après RNSR).

Pour Scopus

Sur le même exemple, nous voyons l'ambiguïté manifeste entre deux noms d'affiliations assortis d'un identifiant numérique propre à la base Scopus (24 mai 2016).

1) LAM Laboratoire d'Astrophysique de Marseille, Traverse du Siphon-Les trois Lucs, Marseille, France. Affiliation ID : 60007671 ; 1702 réponses.

Other name formats : Laboratoire d'Astrophysique de Marseille Lab. d'Astrophysique de Marseille, LAM.

Dans ce signalement, nous voyons la mention de la « Traverse Siphon » repérée sur la liste WoS.

En limitant les 1702 réponses obtenues ci-dessus avec l'identifiant 60007671 du LAM à la période 2012–2016, soit depuis la création de l'institut Pythéas, nous obtenons 242 réponses contre 46 avec Pythéas tout court dans le champ « Affiliation Name ».

2) Inst. d'Astrophysique de Marseille, 2 Place Le Verrier, Marseille, France. Affiliation ID : 100564665. Sous ce signalement identifié par un numéro d'affiliation Scopus, nous ne récoltons que 4 références. Peut-être ne devons-nous cette signature « Inst. ... » qu'à une approximation de la part des auteurs ?

Ces exemples sous-tendent les recouvrements propres à ces deux bases assez largement illustrés dans la littérature ; voir notamment l'étude de Philippe Mongeon et Adèle Paul-Hus de l'université de Montréal, intitulée *The journal coverage of Web of Science and Scopus : a comparative analysis* [22].

De nombreuses démonstrations opérées par des professionnels de l'IST sont consultables en ligne ; voir par exemple l'intervention de Doriane Ibarra (École des Mines de Paris) à la Journée d'étude « Affiliations » Renatis-Urfist du 19 juin 2015 [23].

Signalons également que La Direction de l'information scientifique et technique (DIST) du CNRS a commis en novembre 2015 une note publique précisant :

« Il convient de noter d'emblée que les deux bases peinent à repérer correctement et de façon exhaustive les bibliographies des laboratoires. L'offre globale internationale n'a pas encore réussi à fournir une solution opérationnelle pour traiter l'hétérogénéité de présentation des affiliations : c'est un obstacle important à l'expression de la mixité des co-tutelles » [24].

Du côté des éditeurs des publications

Premier exemple

Voyons la retranscription d'une affiliation par le journal *Organic Geochemistry* (éd. par Elsevier).

L'affiliation est indiquée de la façon suivante :

« Institut National de Recherches Archéologiques Préventives, UMR 5608 TRACES, Université de Toulouse 2–Le Mirail, France ».

La retranscription dans la base Scopus est identique :

« Institut National de Recherches Archéologiques Préventives, UMR 5608 TRACES, Université de Toulouse 2, Le Mirail, France ».

Celle que nous relevons dans la base du Wos est différente :

« Univ Toulouse 2, Inst Natl Rech Archeol Prevent, UMR TRACES 5608, Le Mirail, France ».

Organization Enhanced Name(s)

Centre National de la Recherche Scientifique (CNRS)

PRES Université de Toulouse

Université Toulouse Jean-Jaurès

Université de Toulouse

PSL Research University Paris

Ecole des Hautes Etudes en Sciences Sociales

Second exemple

Dans la revue *Microbial Ecology* (éd. par Springer) une affiliation est indiquée comme suit :
Laboratoire Microorganismes : Génome et Environnement, UMR CNRS 6023, Clermont Université, Université Blaise Pascal, BP 80026, 63171, Aubière Cedex, France

Scopus la reprend à l'identique : Laboratoire Microorganismes : Génome et Environnement, UMR CNRS 6023, Clermont Université, Université Blaise Pascal, BP 80026, Aubière Cedex, France

Le Wos la formule ainsi : Univ Clermont Ferrand, Clermont Univ, Lab Microorganismes Genome & Environm, UMR 6023, CNRS, F-63171 Aubière, France

Organization Enhanced Name(s)

Centre National de la Recherche Scientifique (CNRS).

Conclusion

Une simple analyse de ce type, à partir des bases bibliométriques et bibliographiques de renommée internationale, démontre qu'une identification pleine d'incertitudes conduit à un facteur de déperdition bibliométrique important et de facto impacte forcément l'évaluation des entités (individuelles ou collectives).

Cela met en évidence le caractère incontournable de l'affiliation. Homogénéisée et correctement interprétée, elle est une garantie majeure de visibilité pour la recherche.

Une prise de conscience générale émerge sur ce contentieux et, dans l'urgence, des dispositifs sont mis en place pour rechercher des solutions.

Des solutions politiques et académiques

Le Ministère de l'enseignement supérieur et de la recherche (MESR), pilote de la stratégie pour la recherche au niveau national

En 2006, le Gip OST [25] a pris l'initiative de lancer un chantier, nommé Normadresses, sur le libellé des adresses françaises. Le travail a été mené en partenariat avec le ministère et l'ensemble des institutions de recherche concernées afin de produire une série de recommandations communes, à partir d'informations livrées par l'éditeur de la base, Thomson-Reuters, pour aboutir à une amélioration du libellé des adresses françaises présentes dans le Web of Science.

L'une des conclusions du rapport [26] (codée RT4) est singulièrement significative et déconcertante : « le libellé des adresses des publications des UMR pose de nombreuses questions. En effet, beaucoup de choix seraient théoriquement possibles, et les évaluer n'est pas du ressort du présent groupe de travail. Cependant, il est clair qu'il serait très souhaitable de définir des règles communes claires. »

Par la suite, nous constatons la quasi-généralisation « officieuse » du mode multilignes puis, fin juin 2015, un retour tout aussi officieux vers le monoligne.

Les dernières recommandations (février 2016) de l'Inspection générale de l'administration de l'éducation nationale et de la recherche (IGAENR), pour simplifier les signatures et opérations de recensement des publications nationales [27], méritent d'être soulignées. Elles proposent que selon les signatures de référence qui seraient accessibles en ligne sur le Répertoire national des structures de recherche (RNSR) [28] et la forme retenue, chaque chercheur « utilisera les structures monoligne ou multiligne en fonction des prescriptions des éditeurs où son article sera publié ». Nous ne pouvons pas dire qu'il y ait là motif à se rassurer.

Plusieurs initiatives tentent de résoudre, au plan national, la problématique des affiliations :

- Conditor [29] est une expérience menée par un consortium institutionnel (ABES, CNRS [DASTR, DIST/INIST/CCSD, INSHS], Inra, Inria, IRD, MESR, université Paris Dauphine, université de Bordeaux) en vue de recenser l'ensemble de la production scientifique de l'ESR (articles, ouvrages, congrès, thèses, rapports ...). Chaque entité participante a fourni ses données publiées durant l'année 2011, constituant ainsi un corpus sur lequel ont travaillé plusieurs filières professionnelles : documentalistes, bibliométriciens, gestionnaires d'archives ou de bases de données, informaticiens, curateurs... Les informaticiens ont développé un modèle fondé sur des appariements de caractères afin d'agréger des affiliations et des noms d'auteurs. Ceux-ci sont identifiés par leur IdRef (référentiel national, géré par l'ABES, comportant pour chaque auteur répertorié des informations le concernant et une liste de publications). Les entités d'affiliations sont identifiées par un code issu du RNSR ;
- le RNSR : créé et administré par le département des systèmes d'information de la DGESIP-DGRI du MESR, ce répertoire a pour but d'attribuer aux différentes tutelles (État, agences de moyens, opérateurs de la recherche...) un identifiant national unique, véritable clé du processus d'évaluation.

Si nous reprenons l'exemple précédent du laboratoire LAM : il porte l'identifiant 201220274P dans le RNSR, et UMR 7326.

Il est mentionné que le LAM est une composante de l'unité de recherche 201220900 V : OSU Pythéas qui elle-même regroupe d'autres composantes :

- unité de recherche 201220274P : Laboratoire d'astrophysique de Marseille ;
- unité de recherche 201220322S : Centre européen de recherche et d'enseignement de géosciences de l'environnement ;
- unité de recherche 201220339 K : Institut méditerranéen de biodiversité et d'écologie marine et continentale ;
- unité de recherche 201220349 W : Institut méditerranéen d'océanographie

À condition qu'il soit bien configuré (respect des arborescences et traçabilité des filiations), nous voyons bien la puissance et l'intérêt d'un tel outil développé sur le plan national car il pourrait assurer une meilleure reconnaissance des structures dans les bases de données commerciales.

L'Observatoire des sciences et des techniques (OST)

L'Observatoire des sciences et des techniques (OST), GIP rattaché au HCERES, gère une base de données issue d'une révision augmentée et le cas échéant corrigée du Web of Science. Particulièrement minutieux, ce travail fait cependant l'objet de reprises par les professionnels de la documentation localement afin de combler des lacunes ou des inexactitudes inévitables et parfois nombreuses. D'où l'intérêt d'entreprises telles que Conditor adossée au RNSR. Malheureusement, le statut expérimental de Conditor mériterait une reconduction qui n'est pas présentement acquise.

Le cadre de cohérence des systèmes d'information du MESR

Le cadre de cohérence des systèmes d'information du MESR est une initiative de ce ministère, amorcée en juin 2015 et menée à terme en avril 2016. Les gestionnaires des S.I. des établissements sont invités à harmoniser un certain nombre de standards pour faciliter les échanges. L'identification des entités de l'ESR est jugée prioritaire afin de faciliter, d'une part, le recensement de la production scientifique et, d'autre part, en aval, l'évaluation proprement dite. En effet, l'article L111-1 du Code de la recherche, définit le fait qu'au-delà de l'accroissement des connaissances, la politique nationale de la recherche et du développement technologique vise à : « partager la culture scientifique, technique et industrielle (...) et valoriser les résultats de la recherche au service de la société ». Il existe donc une prise de conscience à un niveau administratif élevé de la nécessité d'harmoniser les moyens de reconnaissance des items de production scientifique. Nous sommes en droit d'espérer que cette prise de conscience dépasse la rhétorique de déclaration de bonnes intentions.

Initiatives internes au sein des laboratoires ou des établissements, permettant la collecte des corpus de publications et la production d'indicateurs

Logiciel Sampra [30]

SAMPRA est un système de gestion et d'analyse des publications développé par l'équipe qui a conçu le logiciel SIGAPS.

Pour mémoire, SIGAPS [31], initialement développé au CHRU de Lille en 2002, avait pour objectif d'aider au recensement et à l'analyse des publications scientifiques référencées sur PubMed, pour un établissement ayant des activités de recherche médicale (plus de 150 établissements français : 32 CHU/CHR, 20 CLCC - Centres de lutte contre le cancer - ...).

Aujourd'hui SAMPRA, en lien avec l'application WAAN [32] de Thomson-Reuters, contribue notamment à l'homogénéisation des affiliations. Ses fonctionnalités :

- recenser, rapidement et de manière fiable, l'ensemble des publications d'un établissement ;
- produire des analyses (tableaux et graphiques) par chercheur, équipe, laboratoire, ...
- accroître la visibilité d'un établissement, en transmettant à Thomson-Reuters, la liste des variantes d'adresses pour une identification unique dans le Web of Science (« Organization enhanced »).

OPALIA

Base bibliométrique développée par l'École nationale des Ponts & Chaussées [33] dans le cadre de la Comue Paris-Est, qui a fourni le serveur, et par une SSI. AgroParisTech, organisme extérieur à la Comue fait aussi partie des partenaires.

Opalia est un outil commun à tous les établissements pour capitaliser, recenser et localiser l'ensemble des publications sur une seule plateforme, à partir du WoS, de Scopus et de HAL. Ce système permet de tagguer chacune des publications avec un nom de laboratoire. Opalia génère, en outre, l'import des données de Sherpa/Romeo pour les politiques des éditeurs.

Localement, tout le travail réside dans l'injection des données en amont et le traitement par chaque établissement ; ensuite les données sont aisément exploitables.

Solutions et attentes des professionnels de l'information scientifique et technique

Les attentes restent vives. La responsabilisation des auteurs et l'expertise des professionnels de l'information sont deux paramètres fondamentaux de la gestion des affiliations. Cependant, ils doivent être impérativement complétés par des solutions techniques plus adéquates.

Dans cette perspective, deux conclusions se détachent, l'une politique, l'autre technique.

À l'échelle politique

Porter la problématique à une échelle globale ou nationale permettrait d'éviter l'individualisation parfois excessive que reflète politiquement l'obésité contre-productive des affiliations dans les publications scientifiques actuelles.

De nombreux outils ou référentiels fédératifs existent ou sont en devenir, tels que Conditor, le RNSR, ainsi que de nombreux identifiants auteurs IDref, IDhal, ISNI ... Ils pourraient être développés et interconnectés pour offrir à la communauté scientifique un système cohérent, pérenne et univoque dédié à l'identification des corpus de publications.

Les professionnels de l'information scientifique et technique ont la compétence requise pour contribuer utilement à la conception des outils et enrichir les bases de données existantes.

À l'échelle technique

Une seconde façon de poser le problème nous mène plus près des solutions techniques et de la prise en compte des technologies du Web de données (Linked Data), du Web sémantique et de l'interopérabilité des données.

En amont, l'amélioration de la robustesse des indicateurs produits par les éditeurs permettrait d'aboutir à des recensements bibliométriques plus exhaustifs et homogènes. Les grands éditeurs de bases de données ont les moyens de développer des outils particulièrement sophistiqués, susceptibles de résoudre le problème des affiliations à très large échelle.

En premier lieu, la contrainte d'agencement des composantes d'une affiliation ne devrait plus constituer un frein technique à l'identification des publications et à l'unification des données, relatives à une institution (contrainte à l'origine de la règle d'écriture multiligne).

Pour l'heure, l'optimisation de ce repérage est assujettie au travail considérable, hyperminutieux et chronophage, effectué par des professionnels de la documentation dans un souci de valorisation de la production scientifique des établissements, organismes et laboratoires. De plus, c'est une opération coûteuse qui in fine profite aux éditeurs, pour un résultat relativement éphémère.

Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts. Par ailleurs, les opinions et les arguments exprimés ici sont ceux des auteurs et ne reflètent pas nécessairement les vues officielles de l'OCDE ou de ses pays membres.

Références

1. M. Dassa, A. Deniau, M. Durand-Barthez, N. Pothier, A. Séné
Le millefeuille des affiliations françaises dans les publications scientifiques
Doc-Sci Inf, 51 (4) (2014), pp. 12–16
2. N. Roquefère, M. Villeneuve, V. Vincent
Affiliations des producteurs de la recherche publique française : état des lieux sur les pratiques de signature des publications scientifiques des laboratoires affiliés au CNRS en Midi-Pyrénées et Languedoc-Roussillon. Réseau Doccitanist ; CNRS (2012)
http://doccitanist.lirmm.fr/IMG/pdf/2011_doccitanist_affiliations_rapport.pdf
3. F. Girard, A. Séné, N. Pothier, A. Deniau, V. Prêtre, M. Durand-Barthez
Le millefeuille des affiliations : où en sommes-nous ?
I2D, 53 (1) (2016), p. 84
4. Journée d'étude Renatis-Urfist « Affiliations dans les publications scientifiques françaises »
École nationale des Chartes, Paris (2015) <http://urfist.enc.sorbonne.fr/veille-et-recherche/rencontres-scientifiques/rencontres-urfist/je-affiliations-dans-les-publications-scientifiques-francaises>
5. D. Pontille
La Signature scientifique : une sociologie pragmatique de l'attribution
CNRS Éditions, série « CNRS sociologie », Paris (2004) 200 p
6. S. Shapin, S. Schaffer
Léviathan et la pompe à air. Hobbes et Boyle entre science et politique
La Découverte, Paris (1993) 468 p
7. <https://www.curie.asso.fr/-Cahier-de-laboratoire-national-.html>.
8. M. Biagioli
Documents of documents. Scientists' names and scientific claims
A. Riles (Ed.), Documents: Artifacts of Modern Knowledge, Michigan University Press (2006), pp. 127–157
9. D. Pontille
Signer ensemble. Contribution et évaluation en sciences
Economica, coll. « Études sociologiques », Paris (2016) 208 p
10. J.S. Katz, B.R. Martin
What is research collaboration?

Res Policy, 26 (1) (1997), pp. 1–18

11. L. Duley
Anonymity of authorship
Lancet, 345 (8961) (1995), p. 1372
12. T. Charen
Authors! Authors!
NLM Tech Bull (225) (1988), p. 14
13. N. Wright
Author policy for NLM indexes and databases
NLM Tech Bull (283) (1995), p. 17
14. L. Schulman, N. Wright
New Author Policy for NLM indexes and databases
NLM Tech Bull (286) (1995), p. 17
15. M.F. Oliver
AI, or the anonymity of authorship
Lancet, 345 (8951) (1995), p. 668
16. A. Flanagan, P.B. Fontanarosa, C.D. DeAngelis
Authorship for research group
JAMA, 288 (24) (2002), pp. 3166–3168
17. C.L. Meinert
In defense of the corporate author for multicenter trials
Control Clin Trials, 14 (3) (1993), pp. 255–260
18. K. Dickersin, R. Scherer, E.S.T. Suci, M. Gil-Montero
Problems with indexing and citation of articles with group authorship
JAMA, 287 (21) (2002), pp. 2772–2774
19. http://urfist.enc.sorbonne.fr/sites/default/files/mdb/Leboeuf_Signature%20commun%20UPSaclay.pdf.
20. <http://www.bib.mines-paristech.fr/Donnees/data12/1263-2015-Note-Signatures-publications-mal-j-1.pdf>.
21. http://www.sorbonne-universites.fr/fileadmin/user_upload/SorbonneUniversites-signature_publications.pdf.
22. Mongeon Ph, A. Paul-Hus
The journal coverage of Web of Science and Scopus: a comparative analysis
Scientometrics, 106 (1) (2016), pp. 213–228
23. D. Ibarra
Bibliothèque MINES ParisTech. Solutions actuelles, aide à l'identification et au référencement, solutions technologiques existantes : prototype, outils et

référentiels. Table ronde. Journée d'étude Renatis-Urfist « Affiliations dans les publications scientifiques françaises »
École nationale des Chartes, Paris (2015)

24. D. Egret, K. Takeda, C. Weil Miko, R. Fabre
Rapport de mission. 'Fiabiliser l'analyse des publications scientifiques'
CNRS (2015) <http://www.cnrs.fr/dist/z-outils/documents/Rapport%20de%20mission%20Elsevier%20Thomson%20Reuters%20novembre%202015.pdf>
25. <http://www.obs-ost.fr>.
26. O. Hologne, L. Martineau, S. Ramanana
Rapport d'étape sur l'analyse des instructions émises par les établissements à l'attention de leurs chercheurs concernant les publications scientifiques
(2007) <http://prodinra.inra.fr/record/194172>
27. Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche (France)
Simplification du fonctionnement des établissements d'enseignement supérieur et de recherche et de leurs laboratoires. Rapport n° 2016-014
(2016) http://cache.media.enseignementsup-recherche.gouv.fr/file/Simplification/94/0/2016-014_Simplification_fonctionnement_etablissements_569940.pdf
28. <https://appliweb.dgri.education.fr/rnsr/>.
29. <http://www.bibliothequescientifiquenumerique.fr/conditor/>.
30. P. Devos, S. Bihan
SIGAPS/SAMPRA. Journée d'information : « Bibliométrie, scientométrie et métriques alternatives : quels outils pour quelles stratégies ? »
Bibliothèque universitaire des Langues et Civilisations (BULAC), Paris (2015)
http://adbu.fr/competplug/uploads/2015/04/SIGAPS-SAMPRA_vuSB.pdf
31. C. Derancourt, P. Devos, N. Moore, J. Rouvillain
Modèle de valorisation des publications médicales scientifiques : le logiciel SIGAPS
Ann Dermatol Venereol, 141 (8–9) (2014), pp. 531–535
32. <http://ipscience-help.thomsonreuters.com/incitesLiveESI/ESIGroup/citationThresholds/thresholdsESI/institutionsESI.html>.
33. F. Bordignon
L'identification des publications de l'École des Ponts ParisTech. Présentation en ligne. Journée d'étude Renatis-Urfist « Affiliations dans les publications scientifiques françaises »
École nationale des Chartes, Paris (2015)
<http://urfist.enc.sorbonne.fr/sites/default/files/mdb/FBordignon.pdf>

Auteur correspondant.

1 : Pour une discussion pionnière sur la relation entre prolifération des adresses, présence dans les bases de données, et mesures de la collaboration, voir [10].

2 : La NLM gérait parallèlement la base de données en ligne MEDLINE (Medical Literature Analysis and Retrieval System Online), consultable à distance dès 1972, qui a pris définitivement le relais en 2004.