

## **La base d'articles Web of Science**

**Suzy RAMANANA-RAHARY**

**Responsable de corpus, OST**

J'évoquerai, pour ma part, la base d'indicateurs de bibliométrie scientifique de l'Observatoire des Sciences et des Techniques. Un certain nombre d'entre vous connaissent déjà le sujet des données et des indicateurs. Je me propose cependant d'opérer un rappel quant aux données sources. J'enchaînerai ensuite sur la méthode de construction d'un indicateur.

Les données sources correspondent à un ensemble de publications se trouvant dans les revues. Nous possédons une licence d'exploitation de base bibliométrique d'indicateurs, qui nous est fournie par Thomson. L'ensemble est constitué de journaux regroupés selon trois thématiques : sciences de la matière et de la vie, sciences sociales, sciences humaines. Les critères d'intégration des journaux dans la base sont des critères de régularité, de rigueur et de visibilité. L'intégration est extrêmement sélective. Les informations sont ensuite structurées, avec l'information centrale que représente l'article et l'intégralité des informations qui permettent de disposer des indicateurs. Le système d'organisation autorise également un lien avec des informations de journaux, lien sur lequel se fondera la construction des agrégations thématiques. Enfin, il existe une spécificité à la base d'indicateurs de bibliométrie : dans les publications, se trouve un système de référencement permettant de construire les indicateurs de visibilité ou indicateurs de citation.

Pour indication, sur une année, nous dénombrons environ 1 million de publications en sciences de la matière et de la vie. A cette occasion, il est important de noter une autre spécificité de la base : une connexion entre les différents produits (un journal peut appartenir à la fois aux sciences de la matière et de la vie, aux sciences sociales, aux sciences humaines).

La construction des indicateurs commence par une phase d'extraction de l'information, avec une partie centrale constituée de l'expertise. En fin d'exploitation des données sources et en fin de processus, grâce à l'utilisation d'une nomenclature spécifique à l'Observatoire des Sciences et des Techniques se basant sur l'utilisation d'un référentiel commun, nous aboutirons à trois types d'indicateurs : les indicateurs de production, les indicateurs de partenariat et les indicateurs de visibilité. Il convient de noter que l'attribution des productions aux différents acteurs géographiques s'opère de manière automatique, les champs de Thomson apparaissant particulièrement bien renseignés. L'identification au niveau des institutions demande en revanche de mobiliser la totalité de l'adresse, avec l'aide de vos repérages. L'intégration du type de compte à l'Observatoire des Sciences et des Techniques s'effectue, quant à lui, sans souci au niveau de l'acteur géographique. L'unité de compte reste l'article, l'intérêt du compte fractionnaire utilisé étant d'être consolidable à tous les niveaux (si vous regroupez les acteurs d'une discipline, vous ne rencontrerez aucune difficulté à établir un compte toutes disciplines confondues). Pour la partie institutionnelle, nous récupérons les informations données au niveau du repérage, sans spécification particulière sur l'institution (il ne s'agit pas de l'intérêt premier de la production). Ensuite, nous réalisons une combinaison entre les deux niveaux d'identification définis (l'identification thématique et l'identification de la contribution de l'acteur) après laquelle il ne nous reste plus qu'à agréger les données pour aboutir à l'indicateur.

Pour conclure, il convient de noter que la méthodologie de programmation par les données pour laquelle l'Observatoire des Sciences et des Techniques a opté allie à la fois la gestion des règles par nomenclature et la gestion modulaire des traitements afin d'adapter les indicateurs selon les besoins de chaque étude et d'en assurer simultanément la fiabilité. Il s'agit du principe de base de chaque corpus de l'Observatoire des Sciences et des Techniques. Merci.

**Ghislaine FILLIATREAU**

Il convient principalement de noter qu'il est possible de moduler les fractionnements et les logiques de comptes selon le résultat souhaité. Le travail sur chaque adresse et sur chaque donnée unitaire offre la possibilité d'opérer des choix de calcul d'indicateurs qui permettront de parler soit de contribution, soit de participation, et d'attribuer des règles de crédit d'un article avec une pondération selon des choix qui peuvent être des choix politiques. Nous travaillons les données très finement, très précisément, avant utilisation de systèmes de consolidation, qui n'empêchent pas d'opérer des choix préalablement, notamment dans les unités mixtes. Nous maîtrisons totalement les choix d'attribution des ratios que nous donnerons pour la participation d'un établissement. La démarche est liée à l'architecture de notre base et au fait que nous travaillons et structurons jusque dans le moindre détail l'information qui nous est livrée par le fournisseur. Bien entendu, nous recevons la base *Web of Science* en copie ; nous ne travaillons pas en ligne.