



www.cnrs.fr



Direction de l'Information
Scientifique et Technique

3 rue Michel-Ange
75016 PARIS

DISTInfo13 /mars 2015

Elsevier annonce le lancement de SciVal Trends qui exploite et analyse les données d'usage des chercheurs utilisant Scopus et Science Direct.

L'information :

Elsevier a annoncé¹ le 25 février le lancement d'un nouveau module au sein de son outil analytique *SciVal* baptisé *SciVal Trends*. Ce module permet pour la première fois d'utiliser au sein de la plate-forme d'analyse d'Elsevier les « traces » laissées par les usagers des services *ScienceDirect* (bibliothèque de périodiques en ligne) et de *Scopus* (moteur de recherche intégrant bases bibliographiques et bases de citations). Jusqu'ici les analyses de *SciVal* portaient exclusivement sur les publications scientifiques. Or comme le précise le communiqué de presse d'Elsevier un tiers des chercheurs ne sont pas publiants : les centres d'intérêts de ces chercheurs étaient jusqu'ici une « zone aveugle » inaccessible aux analyses de *SciVal*. En combinant les données d'usage (très volumineuses : on est là dans la logique du « Big Data ») de ses services en ligne les plus utilisés et la puissance analytique des technologies propriétaires de *RELX Group* (nouveau nom du groupe *Reed Elsevier*), *SciVal Trends* va permettre à ses usagers de visualiser sur la base de mots-clés définissant un profil thématique les « traces » laissées par ces utilisateurs anonymes et non publiant - mais aussi les traces des chercheurs-publiant utilisant les services en ligne d'Elsevier dans leurs recherches documentaires. Cette nouvelle fonctionnalité s'appuie sur la technologie *High Performance Computing Cluster* (HPCC), une technologie propriétaire du Groupe RELX (connu jusqu'ici sous la raison sociale *Reed Elsevier*). Cet outil est l'une des « dorsales technologiques » du groupe sur laquelle s'appuie aussi bien la division *LexisNexis* pour analyser des fonds de jurisprudence ou des fonds d'information brevets que la division Elsevier au travers de *SciVal*. Ce dernier service permet désormais d'analyser 75 billions (75 000 milliards) de données métriques d'usages actualisées chaque semaine. L'économie numérique est une économie des données et de la trace : les technologies analytiques traitant des données d'usage des services en ligne étaient jusqu'ici réservées à des acteurs comme *Google* ou *Facebook*. Cette valorisation par des technologies de Big Data, dans le cadre de *SciVal Trends*, des « traces » (anonymisées) laissées par les usagers des services en ligne est une première dans le champ de l'IST.

L'analyse de la DIST :

SciVal Trends fait entrer l'IST dans l'ère du Big Data et du Data Mining. Il y a un an, en janvier 2014 Elsevier annonçait la refonte profonde de son outil analytique.

¹ <http://www.elsevier.com/about/press-releases/science-and-technology/elsevier-launches-the-new-generation-scival,-providing-dynamic-real-time-analytics-and-insights-into-the-global-research-landscape#sthash.M6VFPF5s.dpuf>

Au-delà de l'argumentaire technique, le communiqué de presse *Elsevier* publié l'an dernier indiquait : «*Le monde de la recherche est confronté à la globalisation et à une compétition accrue (...). Les structures de recherche académiques, gouvernementales ou industrielles sont soumises à une pression qui va croissant pour optimiser leurs ressources et maximiser les avantages compétitifs que sont leurs domaines d'excellence dans un contexte mondial de plus en plus encombré. Aujourd'hui l'économie de la recherche est si vaste et si complexe, comptant 7 millions de chercheurs, qu'il est de plus en plus nécessaires de s'appuyer sur des données analytiques fiables (NDLR : pour piloter les activités de recherche dans ce contexte concurrentiel) ».*

SciVal est l'outil de ce qu'*Elsevier* désigne du terme de « *Research Intelligence* » (intelligence au sens anglo-saxon de renseignement et de compréhension analytique)², un concept qui intéresse les responsables des stratégies scientifiques des grands organismes de recherche mais aussi à un niveau plus opérationnel les responsables des grands « clusters » de recherche structurés par disciplines voire les chercheurs désireux de se situer sur la « carte » analytique de leur champ de compétence. Mais les entités qui ont la charge de la gouvernance de la recherche ou les agences de financement sont aussi des utilisateurs potentiels de ce type d'outil. Les tarifs très élevés de ce type d'outils les réservent de toute façon aux grandes structures. *Elsevier* a investi lourdement pour développer ces outils de « *Research Intelligence* » sans probablement qu'aujourd'hui le retour sur investissement soit assuré. Les marges très élevées de l'édition de revues scientifiques et de la mise à disposition de services de recherche (*ScienceDirect*, *Scopus*) sur de larges corpus d'articles de recherche ont permis hier de financer cet effort de R&D sur des outils analytiques traitant de grands volumes de données et aujourd'hui financer leur exploitation commerciale dans une logique patiente d'« évangelisation ». Pour *Elsevier*, l'horizon stratégique pour ce type d'offre est de 4 à 5 ans. Le numéro 1 mondial de l'édition scientifique à une longueur d'avance sur tous ses autres concurrents, y compris *Thomson Reuters* qui est l'autre promoteur des outils de « *Research Intelligence* ». Or pour de nouveaux entrants sur le marché de ce type d'outils la courbe d'apprentissage est longue, de l'ordre de dix ans avant de disposer d'une offre commerciale mature. Dix ans, c'est à peu près le temps depuis lequel *Elsevier* laboure ce nouveau sillon. Depuis dix ans, *Elsevier* a construit patiemment ce qui est le fondement de ses annonces d'aujourd'hui et ce qui sera son avantage compétitif unique de demain.

Cette actualité est par ailleurs une illustration du fait que les grands groupes « for-profit » de l'édition scientifique, qui sont aussi de grands groupes diversifiés, ont désormais un avantage compétitif « build-in » et durable sur les groupes de plus petite taille sur le marché de l'IST. Car le facteur « taille » est ici décisif : taille des corpus dont disposent ces grands éditeurs, volumes d'usage sur les services en ligne qu'ils proposent, volumes des données captées sur ces services, importance des ressources financières, capacité à rentabiliser des investissements technologiques transversaux (ici les technologies du Text et du Data Mining) sur un portefeuille d'activité diversifié allant au-delà de l'édition scientifique. Le nom du numéro 1 du secteur de l'IST dans dix ans est déjà

² http://www.elsevier.com/_data/assets/pdf_file/0007/188296/elsevier-research-intelligence-brochure.pdf