



# Google Scholar, Google Print ... un nouvel entrant pour la diffusion de la recherche.

Olivier Ertzscheid

► **To cite this version:**

Olivier Ertzscheid. Google Scholar, Google Print ... un nouvel entrant pour la diffusion de la recherche.. Perspectives documentaires en éducation, ENS Editions, 2005, 62. sic\_00090140

**HAL Id: sic\_00090140**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00090140](https://archivesic.ccsd.cnrs.fr/sic_00090140)**

Submitted on 28 Aug 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Diffusion des résultats de la recherche : une nouvelle donne numérique ?

## **GOOGLE SCHOLAR, GOOGLE PRINT... UN NOUVEL ENTRANT POUR LA DIFFUSION DE LA RECHERCHE ?**

Olivier ERTZSCHEID, McF, Université de Nantes, IUT de La Roche sur Yon.

### **1. INTRODUCTION.**

*« Today we can access a significant part of the last decade of the literature electronically. In 5 to 10 years, this will extend to much of the significant literature of the twentieth century, that is, the 1000 or more most-consulted and higher impact journals. » (GAR 05)*

72 % des universitaires utilisent le moteur Google pour leurs recherches d'articles scientifiques (SWA 05). L'INIST, premier fournisseur français de documents de recherche a fait de Google Scholar<sup>1</sup> son moteur d'accès et annonce d'autres partenariats en cours de négociation<sup>2</sup>. Même la BNF dont on sait les positions de son président, utilise Google comme moteur de recherche interne au sein de sa propre interface<sup>3</sup>. Au-delà de ces symptomatiques remarques préliminaires, c'est toute la chaîne de diffusion de la recherche et des connaissances scientifiques qui se trouve aujourd'hui face à un bouleversement inédit sur 3 plans :

- la massification des accès et des usages d'une part,
- l'élargissement du spectre des pratiques de validation d'autre part,
- le passage enfin d'une situation de diffusion monopolistique (éditeurs commerciaux) à un équilibre tripartit entre ces derniers et les deux nouveaux entrants que sont les moteurs de recherche et les archives ouvertes et institutionnelles.

Ce qui constitua un temps un parangon de stabilité se trouve aujourd'hui à la confluence de forces et d'antagonismes très puissants : les éditeurs scientifiques doivent désormais compter sur des pratiques d'autopublication et d'auto-archivage de plus en plus répandues. Les logiques d'open access, le plus souvent thématiques ou disciplinaires<sup>4</sup> se répandent et emportent l'adhésion de plus en plus de chercheurs. Les institutions elles-mêmes commencent à prendre le relai en mettant en place des politiques d'archivage institutionnel. Cet ensemble de ressources et d'initiatives<sup>5</sup> mettent, pour le secteur de l'édition commerciale, hors de portée marchande une masse de plus en plus considérable de connaissances et de corpus.

L'interrogation actuelle est de savoir dans quelle mesure les outils de recherche, constitués en autant de base de données ou d'entrepôts de la connaissance scientifique, peuvent prendre le relais, voire supplanter les acteurs actuellement dépositaires des seuls indicateurs reconnus pour l'évaluation des publications et de la recherche. Parmi ces acteurs, Google concentre autour de lui l'ensemble des problématiques citées précédemment. Formulée autrement, l'interrogation centrale de cet article pourrait être ainsi résumée : Google a-t-il l'ambition et les moyens de devenir le prochain ISI<sup>6</sup> ?

### **2. UNE INFODOMINANCE PROBLEMATIQUE DE L'ISI.**

Les termes de « bibliométrie » et de « scientométrie » désignent un ensemble de méthodes pour mesurer la production et la diffusion des connaissances scientifiques. Ces deux disciplines ont été fondées par Derek de Solla Price & Vasilij Vasilevitch Nalimov. C'est l'arrivée des outils de

---

<sup>1</sup> <http://scholar.google.com>

<sup>2</sup> [http://www.inist.fr/actualite/article.php3?id\\_article=40](http://www.inist.fr/actualite/article.php3?id_article=40)

<sup>3</sup> <http://www.bnf.fr/pages/zNavigat/frame/recherche.htm>

<sup>4</sup> Voir par exemple le site ArXiv : <http://arxiv.org/>

<sup>5</sup> Le portail ROAR (Registry of Open Access Repositories) en recense la plupart : <http://archives.eprints.org/>

<sup>6</sup> Institute for Scientific Information : <http://www.isinet.com/>

l'ISI (Institute for Scientific Information -actuel Thompson ISI-), fondé en 1966 par Eugène Garfield, qui leur permettra de se déployer et d'atteindre leur plein potentiel. L'ISI a mis en place trois grands index de citations qui assuraient et régulaient jusqu'ici les pratiques de diffusion et de validation de la science: le premier est le SCI « Science Citation Index » applicable aux sciences dures. Le second, SSCI (Social Science Citation Index) est inauguré en 1973 pour les sciences sociales. Le troisième, AHCI (Arts and Humanities Citation Index) est créé en 1978. En 2004 l'ISI comptait plus de 8500 revues et 25 millions de citations par an (BAU 04). L'écosystème scientifique peut alors être ainsi sommairement schématisé : les chercheurs doivent publier dans des revues de premier rang et disposent pour cela d'indicateurs quantitatifs tels le facteur d'impact.

Les journaux et revues pouvant entrer dans ces index doivent satisfaire à un certain nombre de critères : offrir une régularité de la publication, respecter les conventions internationales de présentation, disposer d'un résumé en anglais et disposer d'un comité d'évaluation par les pairs. Chaque index dispose en sus de critères spécifiques.

Chaque année, 2000 nouveaux journaux sont examinés par les éditeurs de chez Thompson ISI. « *Only about 10% of these journals are selected for coverage.* » (VEI 04) Il est donc tout un pan de la littérature scientifique qui n'entre pas dans cet écosystème régulé sans qu'il soit pour autant raisonnable de considérer comme inintéressante la totalité de cette considérable production scientifique en invoquant les seuls principes comptables de l'ISI<sup>7</sup>.

### **3. DE L'OPEN ACCES AUX MOTEURS DE RECHERCHE : NOUVEAUX ENTRANTS , NOUVELLES PRATIQUES, NOUVELLES METRIQUES.**

Le mouvement du libre accès, les nouvelles dynamiques scientifiques à l'œuvre sur le web, les nouveaux habits des chercheurs en terme notamment d'auto-archivage, les prises de position politiques et aussi parfois polémiques de certains d'entre eux refusant de publier dans des revues n'autorisant pas l'auto-archivage (Lawrence Lessig notamment<sup>8</sup>), ainsi que l'arrivée de nouveaux entrants majeurs aux ambitions clairement affichées (Google notamment), cet ensemble de paramètres vient bousculer les modèles traditionnels de diffusion et de validation de la science en remettant radicalement en cause l'infodominance de l'ISI. Plusieurs paramètres sont à l'origine de ce bouleversement :

- un nombre croissant de journaux de l'ISI adoptent des modes de publication ouverts : « *Our findings suggest that over 55% of the journals and over 65% of the articles indexed in Web of Science in 2003 are produced by publishers who permit some form of self-archiving, and could be made OA by author archiving.* » (VEI 04)
- l'ISI lui-même atteste que (VEI 04) « *an OA distribution model did not appreciably alter the performance of a journal, either positively or negatively.* » Le référencement des journaux « ouverts » au sein de l'ISI demeure cependant très limité (ils ne représentaient en 2004 que 1% des 20 000 journaux du ISI web of Knowledge<sup>9</sup>)
- les modes de publication et de diffusion « ouverts » n'altèrent en rien la reconnaissance de l'activité de recherche au sein des différentes communautés disciplinaires, le facteur d'impact pouvant même s'en trouver augmenté (nombre d'études<sup>10</sup> sur des corpus représentatifs et dans différents domaines de connaissance attestent que les articles et revues OA bénéficient d'un meilleur IF). La seule étude offrant des conclusions sinon opposées, en tout cas fortement minorées, est celle commanditée par ... l'ISI. Dans celle-ci, (TES 04) conclue que « *le facteur d'impact des quelque 148 revues en accès libre étudiées par Thompson ISI est en général comparable à celui des revues à accès payant.* » (ARC 04).

---

<sup>7</sup> Les trois grands indicateurs retenus par l'ISI sont les suivants : le nombre de publications (sur une période de temps donnée), le nombre de citation reçues par les revues et le facteur d'impact Pour plus d'informations sur ces indicateurs, voir [ARC 04]

<sup>8</sup> <http://www.lessig.org/blog/>

<sup>9</sup> l'étude de [VEI 04] dénombre 1190 titres uniques répartis sur 3 repositories ([www.doaj.org](http://www.doaj.org), [www.jstage.jst.go.jp](http://www.jstage.jst.go.jp) et [www.scielo.br](http://www.scielo.br))

<sup>10</sup> (LAW 01) (KUR 04) (ODL 02) (ANT 04)

## 4. L'ORIGINE D'UN INTERET.

L'intérêt de Google pour le secteur de la diffusion scientifique s'inscrit d'abord dans une logique capitaliste, une très forte concentration de quelques acteurs se partageant actuellement les fruits d'un marché en constante expansion. La première motivation de Google est donc bien de nature financière. La seconde tient à la stratégie et au positionnement de la firme qui vise à « rendre universellement accessible l'ensemble des connaissances disponibles. » Si l'on peut considérer comme gagnée – par Google – la bataille de l'information (aussi bien en termes d'accès, de diffusion que de modèle économique – publicitaire – associé), la prochaine bataille à remporter est celle de la connaissance. Il faut enfin prendre en compte un positionnement à reconquérir à une période où l'image de la marque apparaît considérablement ternie par quelques affaires de censure (en Chine) et de protection de la vie privée. A ce titre, gagner la confiance des prescripteurs que sont les chercheurs et autres bibliothécaires revient à acquérir, faute de légitimité, une légitimation qui ne peut être que salutaire. Coûts (et bénéfices attendus), connaissance et confiance constituent donc l'équilibre métastable à l'origine de l'intérêt du moteur pour le secteur de la diffusion scientifique.

### 4.1. LE SCENARIO DE DEPLOIEMENT : UN MOTEUR DEUX TEMPS

#### 4.1.1. PREMIER TEMPS : GOOGLE SCHOLAR.

Le service de Google qui se positionne ouvertement sur le secteur de l'accès à l'information scientifique est Google Scholar<sup>11</sup>, lancé fin 2004. Il donne accès à un corpus de documents composé d'articles scientifiques, de thèses, de pré-publications, de rapports ainsi que de monographies. La base de donnée ainsi constituée provient de différentes sources « crawlées » par Google : éditeurs, sociétés savantes, répertoires et archives de pré et post-prints et autres serveurs universitaires<sup>12</sup>. La force de ce service tient d'abord à l'effet de marque et d'usage dont peut se prévaloir Google. Elle tient ensuite à la mise en œuvre pour chaque document indexé, d'une indication du nombre de citations pointant vers ledit document. A l'image du service développé par Garfield et servant d'indicateur à l'ISI, il ne s'agit pas là de « liens web » pointant vers les documents mais bien de « citations » d'un document à l'intérieur d'un autre. Il s'agit de se placer clairement dans l'héritage des métriques scientifiques proposées par l'ISI même si la technique de calcul reste pour l'instant opaque dans la mesure où le taux de couverture de l'outil Google Scholar, les critères d'inclusion ou d'exclusion des documents indexés ne sont pas fournis aux usagers du service qui n'ont ainsi pas d'autre choix que de comparer empiriquement les résultats fournis par les différents services en concurrence (dépôts d'archives ouvertes et moteurs appliqués aux fonds documentaires des éditeurs commerciaux). Ces derniers éléments sont autant de problématiques lacunes : l'occultation délibérée de la couverture et de la profondeur de l'outil porte, par effet de ricochet, atteinte à la qualité statistique des mesures de citation fournies.

Les services concurrents à Google Scholar sont actuellement de deux types : il s'agit d'une part des réservoirs (repositories) d'archives ouvertes, et des entrepôt mis à disposition par les éditeurs scientifiques eux-mêmes, tels le service Scirus d'Elsevier<sup>13</sup>. Pour les premiers, leur force vient de l'interopérabilité des protocoles de mise à disposition (OAI-PMH) et de la gratuité du service offert. Pour les seconds, c'est la nature qualitativement homogène et la garantie de « validation » qui est mise en avant. Les documents référencés sont ensuite accessibles sous plusieurs formes : gratuitement pour certains d'entre eux et sous forme payante pour les autres, avec une hybridation qui peut parfois entretenir chez l'utilisateur une illusion de gratuité lorsque celui-ci accède à des articles payants en interrogeant le service via une bibliothèque universitaire qui aura forfaitairement payé au prix fort cet accès au texte intégral.

C'est sur cette double articulation entre gratuité de l'accès d'une part et niveau de profondeur et de couverture d'autre part qu'est en train de se jouer la partie. Continuant à s'installer au cœur des

---

<sup>11</sup> <http://scholar.google.com>

<sup>12</sup> pour une fiche signalétique complète sur le fonctionnement de Google Scholar, voir <http://urfist.univ-lyon1.fr/GoogleScholar.pdf>

<sup>13</sup> <http://www.scirus.com>. Là encore voir la fiche proposée par l'Urfist de Lyon : <http://urfist.univ-lyon1.fr/SCIRUS.pdf>

institutions et des organismes de recherche et des bibliothèques universitaires, il y a fort à parier que dès que Google pourra, à son entier avantage, faire état des sources utilisées pour le service Scholar et mettre en avant des fonds d'archives ouvertes pouvant eux-mêmes se prévaloir d'une validation au moins qualitativement équivalente – à des degrés divers – aux fonds des revues « classiques », il communiquera sur ce point. Mais il sera alors le seul à pouvoir mettre en avant un modèle de gratuité totale pour l'utilisateur. Même si ladite gratuité est entièrement financée par la publicité, ce qui pose des questions sur lesquelles nous reviendrons plus tard.

S'il pêche encore par son manque de profondeur et d'homogénéité dans le recensement, Google est d'ores et déjà positionné au dessus des autres services équivalents, que ceux-ci proviennent de l'édition commerciale ou de l'archivage ouvert et public. L'expérience du « Google Web » a par ailleurs prouvé que le moteur était en capacité de doubler presque instantanément le nombre de pages accessibles<sup>14</sup>, étant le seul à pouvoir compter sur une infrastructure technologique (nombre de serveurs et de Datacenters) d'une telle densité.

#### **4.1.2. SECOND TEMPS : BIBLIOTHEQUE ET GOOGLE BOOKS**

La documentation scientifique ne se compose pas uniquement d'articles de revues mais compte aussi sur l'ensemble des documents et monographies disponibles. Ainsi, le projet bibliothèque (SAL 05) ambitionnant de numériser les textes disponibles dans le domaine public par le biais d'accords avec quelques prestigieuses bibliothèques peut également entrer dans la stratégie de Google pour s'accaparer la diffusion de la recherche. Ce projet est actuellement fusionné avec le troisième service Google Books<sup>15</sup>.

Initialement nommé « Google Print » ce service concerne l'ensemble de la politique de numérisation de Google (SAL 05). D'après les chiffres cités par Lawrence Lessig, Google Books mettrait à disposition 18 millions d'ouvrages dont :

- 16% sont dans le domaine public
- 9% sont sous presse et couverts par le copyright
- 75% sont encore sous copyright « théorique » mais ne sont plus imprimés.

Notons que pour les services fusionnés « Books » et « bibliothèque », le taux de citation n'est pas pris en compte comme c'est le cas dans Google Scholar et c'est l'interface habituelle de Google qui prévaut.

#### **4.2. QUELLES PERSPECTIVES ?**

Les ambitions de Google s'accroissent mal d'humilité : « organiser les connaissances à l'échelle de la planète et les rendre universellement accessibles. » Avant l'arrivée de Google Scholar, « *Thompson ISI has provided the only source of large scale « inverted » bibliographic data of the sort required for impact analysis.* » (BEL 05) Or Scholar rend disponible le même type de données. Cette arrivée marque donc un clair et inédit bouleversement du rapport de force existant, certains faisant le choix de la crispation. Ainsi Elsevier lui ayant interdit l'indexation de ses articles, Google « *includes hits for more than a million Elsevier articles indexed as abstracts.* » (BEL 05)

C'est l'opposition frontale entre deux modèles opposés qui marque une nouvelle ligne de fracture dans le modèle économique de l'édition scientifique sur le web : le tout gratuit selon Google et le tout payant selon Elsevier avec Scopus, la souscription annuelle pour les institutions étant fixée à partir de 25 000 dollars et pouvant aller jusqu'à plusieurs centaines de milliers de dollars.

Les forces en présence sont donc Google Scholar, Elsevier avec son service Scopus et son moteur de recherche Scirus, Thompson ISI avec le service Web Of Science, l'ensemble des archives ouvertes et institutionnelles et les fonds propres numérisés par les bibliothèques ou les organismes de recherche.

L'appétence de Google est ici comme ailleurs considérable, grignotant progressivement l'ensemble du « marché ». Il a ainsi passé récemment un accord avec le projet MUSE<sup>16</sup> qui rassemble

---

<sup>14</sup> [http://urfistinfo.blogs.com/urfist\\_info/2004/11/le\\_web\\_double\\_d.html](http://urfistinfo.blogs.com/urfist_info/2004/11/le_web_double_d.html)

<sup>15</sup> <http://books.google.com>

<sup>16</sup> <http://muse.jhu.edu/>

270 revues éditées par près de 40 éditeurs. Google développe sa stratégie commerciale à destination de ces trois piliers de la culture scientifique que sont les bibliothèques<sup>17</sup> (en proposant de numériser gratuitement leurs fonds), les éditeurs scientifiques et les éditeurs plus « généralistes » avec Google Books. Comme l'indique Aileen McHugh, directeur des publications électroniques du projet MUSE : « *We believe that this new Google capability will help subscribers increase their usage of MUSE content, enhancing its value and usefulness as a research tool at libraries. (...) Working with Google is also beneficial to publishers who participate in MUSE as content from their journals is more visible to researchers and students,* »<sup>18</sup>

Fidèle à ses habitudes, Google communique peu, préférant laisser courir les rumeurs.

« *Google says almost all 'major publishers' have allowed the full text of their papers to be searched, although it declined to provide a list of those involved. The engine also searches abstracts from online archives such as PubMed and the NASA Astrophysics Data System, and the complete text of physics preprints on the arXiv server. In total, almost half a billion documents are thought to be covered.* » (BUT 04).

(BEL 05) en basant sur l'analyse de 203 publications, « *collectively cited by more than 4000 other publications* » en déduit que « *GoogleScholar seems competitive in terms of coverage for materials published in the last twenty years ; before then ISI seems to dominate.* » (BEL 05) note également que « *Publications in books (...) and conference proceedings are much more likely to be available via GoogleScholar; conversely, journal articles are better indexed via ISI.* »

En se positionnant de la sorte, Google peut avec son service Scholar, devenir très rapidement LE médiateur référent (unique ?) de l'ensemble de la diffusion scientifique, compensant par la masse et l'indice de citation son manque de prise en compte de métadonnées structurées selon le protocole OAI-PMH.

D'autres indices laissent supposer comme autant de possibles, l'acquisition ou la fusion avec Scholar de services tels CiteSeer<sup>19</sup>, index autonome de citation développé par Steve Lawrence qui permet d'afficher en sus du titre de l'article dans lequel le mot-clé est contenu, le contexte dans lequel le terme apparaît. Or Steve Lawrence a récemment quitté le NEC à Princeton pour rejoindre la firme de Mountain View (GAR 05). Il est probable que le projet Google Scholar mettra donc en place les outils développés par Lawrence, enterrant de fait l'actuel Citeseer.

## 5. REPUTATION N'EST PAS CITATION

En brisant ainsi progressivement un monopole (celui de l'ISI), Google en conforte un autre : le sien. Si l'on déplace le débat sur un autre terrain qui est celui de l'évaluation des articles accessibles dans des services de diffusion et d'accès aux résultats de la recherche scientifique un certain nombre de questions refont surface comme en écho à l'idée qui fût à l'origine du succès planétaire de la firme. Il faut en effet se souvenir que le moteur Google fût initialement conçu à Stanford par Serguei Brin et Larry Page, qui avaient alors pour projet d'indexer le fonds de leur bibliothèque universitaire. Et que le principe algorithmique de son célèbre PageRank est la transposition au web des modèles bibliométriques développés par Garfield pour des corpus de publications scientifiques. Or ces derniers modèles permettent de disposer d'indicateurs qualitatifs fiables qui reposent à part égale sur la mathématique statistique utilisée et sur l'homogénéité d'un corpus dont chaque document a d'abord été soumis à une évaluation par le biais d'un comité scientifique. Ainsi dans le cadre des mesures et des indicateurs délivrés par l'ISI c'est l'autorité qui prévaut alors que l'algorithmie du moteur Google ne peut s'établir que sur la notoriété d'un document. Comme le fait remarquer (BIL 06) : « *Trust, authority, and reputation are central to scholarly publishing, but the model of the Internet is almost antithetical to the trust model of academia.* » Google Scholar cristallise ainsi les positions antagonistes en terme de validation des promoteurs de l'open access d'une part et de ceux d'une validation traditionnelle de l'autre. Sans entrer dans ce débat il est en tout cas certain que la pertinence

---

<sup>17</sup> Le service payant de fourniture de la British Library (British Library Direct : <http://direct.bl.uk>) vient de passer un accord avec Google pour permettre à celui-ci de mettre en place dans Google Scholar des liens directs vers les neuf millions d'articles numérisés par la British Library (<http://www.iwr.co.uk/information-world-review/news/2151270/google-scholar-becomes-direct>)

<sup>18</sup> SPARC-OAForum Digest #556

<sup>19</sup> <http://citeseer.ist.psu.edu/>

(« relevance ») d'un document lui permettant d'être affichée en tête d'une page de résultats n'a que peu de choses en commun avec la pertinence scientifique d'un article au regard de l'ensemble de la production scientifique dans un champ de connaissance donné.

Une problématique connexe se présente alors sous l'angle des usages d'un tel service. Le risque est ici celui d'entraîner par effet de calque, des usagers non-experts sur des corpus qui nécessitent une expertise. Si l'accès de chacun – expert ou non-expert - à de tels corpus est une bonne chose en soi le déplacement d'habitus cognitifs différenciés est en revanche problématique : la lecture d'une liste de résultats dans l'interface Google.com repose sur une loi de notoriété. Celle d'une page de résultats dans Google Scholar repose sur une loi d'autorité (avec les limites exprimées plus haut). Or un public non-expert n'est pas censé disposer d'une acculturation lui permettant de décoder les indices bibliométriques fournis sur l'interface de Google Scholar. Il n'est donc pas improbable que les habitus en vigueur sur le moteur (sur lequel une écrasante majorité d'utilisateurs ne lisent jamais la deuxième page de résultat ou ne prennent même pas la peine de descendre en dessous de la « ligne de flottaison » de leur navigateur) il n'est donc pas improbable que ces habitus s'appliquent à des corpus scientifiques, excluant de facto du champ de la recherche tout ce qui ne figurera pas dans les 5 ou 6 premiers résultats du moteur (FRI 06), une logique de « tête de gondole » l'emportant sur celle de la simple pertinence. Au delà des problématiques de diffusion et d'accès à l'information scientifique, et dans un contexte ici décrit pour lequel il y a fort à parier que Google se positionnera très prochainement comme l'acteur incontournable en ce domaine, la massification des usages induits doit porter à l'attention de chacun l'importance d'une pédagogie accrue desdits usages.

La question que pose donc l'intérêt de Google pour la diffusion de la connaissance scientifique n'est au final que très peu éloignée de celle qui agite le débat suite à l'annonce par le même Google de la constitution d'une bibliothèque universelle. Des logiques d'usage, s'effaçant elles-mêmes de plus en plus devant des logiques marchandes (liens sponsorisés notamment), peuvent-elles prendre le pas sur des logiques classificatoires raisonnées dans le contexte de la diffusion et de l'accès aux connaissances scientifiques ?

## 6. CONCLUSION.

La seule logique des forces commerciales en présence – éditeurs et moteurs – ne saurait suffire à anticiper sur le résultat de ce conflit d'intérêt à l'échelle planétaire. Il est certain qu'il faudra attentivement regarder du côté des pratiques et des usages des chercheurs eux-mêmes, qu'ils soient auteurs ou lecteurs.

Concernant la gratuité et la disponibilité des ressources, l'arrivée de Scholar est pour l'heure une bonne chose pour les tenants de l'OAI. Dans le même temps et presque paradoxalement il faut craindre qu'elle accélère ou qu'elle fige les logiques de concentration dans l'édition scientifique commerciale, si ce secteur prétend demeurer en capacité de rivaliser avec l'effet de masse introduit par Google. Car pour Scholar comme pour l'ensemble des services déployés, Google se sert de son moteur comme d'un attracteur permettant d'amener un public captif considérable vers de nouveaux services.

Elle risque enfin d'obliger les acteurs universitaires à se repositionner en reprenant la main sur la maintenance et l'archivage de leurs fonds scientifiques pour ne pas prendre le risque d'une dépossession ou d'un détournement des logiques de consultation qui ne se feront plus que sur le site de la compagnie américaine avec en toile de fond l'application de logiques non plus classificatoires mais majoritairement marchandes.

Les enjeux concernent donc les logiques d'évaluation de la science, ses pratiques et ses modalités d'appropriation individuelles et collectives. En d'autres termes ses conditions de pratique et d'existence.

Pour autant, Google n'a pas en la matière bataille gagnée au regard des logiques propres au secteur de l'édition commerciale scientifique d'une part, et à celui des logiques intrinsèques du mouvement de l'open access d'autre part. Ces deux acteurs ont pour avantage d'avoir à la fois « arisen from and developed within the academic community it exists to serve. » (FRI 06) Et il est loin d'être acquis que la logique de régie publicitaire qui est au cœur de la stratégie de la firme de Mountain View emporte l'adhésion des chercheurs toujours plus attentifs à la qualité de service et très attachés au maintien de tout ou partie des logiques de validation caractéristiques des différentes communautés scientifiques. Cette « mise hors de portée marchande » - pour les éditeurs commerciaux - que nous

pointions en introduction s'accompagnera nécessairement d'une re-marchandisation publicitaire des corpus ainsi agrégés au sein d'outils comme Scholar et il serait dommage et surtout dommageable de ne pas poser clairement la question d'une nouvelle marchandisation de cet ensemble de biens culturels et de produits de la recherche publique.

Bien qu'il soit aujourd'hui douteux qu'une solution négociée demeure possible tant les enjeux commerciaux sont considérables, il est en revanche plus que jamais nécessaire que les différentes communautés s'accordent sur des indicateurs qualitatifs et des métriques quantitatives prenant en compte les moteurs de recherche en général et Google en particulier comme autant d'acteurs incontournables de la diffusion d'une littérature scientifique, phénomène à la marge duquel la massification induite des usages ne peut être que bénéfique à chacun. « All players – open-access and subscription publishers, managers of open-access repositories, and librarians – should collaborate to find a way to identify high academic quality in search engine world. » (FRI 06)

## 7. BIBLIOGRAPHIE

- (ANT 04) ANTELMAN K., « Do Open-Access Articles Have a Greater Research Impact ? » Disponible sur internet : [http://www.lib.ncsu.edu/staff/kantelman/do\\_open\\_access\\_CRL.pdf](http://www.lib.ncsu.edu/staff/kantelman/do_open_access_CRL.pdf) (consulté le 6 Juin 2005)
- (ARC 04) ARCHAMBAULT E., VIGNOLA-GAGNE Etienne, « Science-Metrix Rapport final : l'utilisation de la bibliométrie dans les sciences sociales et les humanités ». Préparé pour le Conseil de recherches en sciences humaines du Canada. août 2004. Disponible sur internet : [http://www.science-metrix.com/pdf/Science-Metrix\\_Utilisation\\_bibliometrie\\_SSH.pdf](http://www.science-metrix.com/pdf/Science-Metrix_Utilisation_bibliometrie_SSH.pdf) (consulté le 6 Juin 2005)
- (BAU 04) BAUDOIN Lesya, HAEFFNER-CAVFAILLON Nicole, PINHAS Nicole, MOUCHET Suzy et KORDON Claude. « Indicateurs bibliométriques : Réalités, mythes et prospective ». Médecine sciences, Vol 20, n°10, Octobre 2004. Disponible sur internet: <http://www.erudit.org/revue/ms/2004/v20/n10/009339ar.pdf> (consulté le 7 Janvier 2006)
- (BEL 05) BELEW Richard K., avril 2005. « Scientific impact quantity and quality : analysis of two sources of bibliographic data. » Disponible sur internet : <http://arxiv.org/abs/cs.IR/0504036> (consulté le 7 Janvier 2006)
- (BIL 06) BILDER Geoffrey, « In Google We Trust », Journal of Electronic Publishing, vol. 9, n°1, Février 2006. Disponible sur internet : <http://www.hti.umich.edu/j/jep/> (consulté le 7 Janvier 2006)
- (BUT 04) BUTLER Decan, « Science searches shift up a gear as Google starts Scholar engine ». Nature, vol 432, 25 novembre 2004. Disponible sur internet : [http://www.nature.com/news/2004/041122/pf/432423a\\_pf.html](http://www.nature.com/news/2004/041122/pf/432423a_pf.html) (consulté le 7 Janvier 2006)
- (FRI 06) FRIEND Frederick J., « Google Scholar : Potentially Good for Users of Academic Information », Journal of Electronic Publishing, vol. 9, n°1, Février 2006. Disponible sur internet : <http://www.hti.umich.edu/j/jep/> (consulté le 7 Janvier 2006)
- (GAR 05) GARFIELD Eugene, « A Prospective View of Citation Indexing and Information Retrieval In the 21st Century », ASIS&T Los Angeles Chapter's Contributions to Information Science & Technology Award. Disponible sur internet : <http://www.garfield.library.upenn.edu/papers/asistlosangeles2005.pdf> (consulté le 7 Janvier 2006)



- (KUR 04) KURTZ M.J., « Restrictive access Policies Cut Readership of Electronic Research Journal Articles by a factor of Two. » Document de travail. Disponible sur Internet : <http://opcit.eprints.org/feb19oa/kurtz.pdf> (consulté le 6 Juin 2005)
- (LAW 01) LAWRENCE S., « Online or Invisible ? » *Nature*. 411(6837):521. Disponible sur internet : <http://www.neci.nec.com/~lawrence/papers/online-nature01> (consulté le 6 Juin 2005)
- (ODL 02) ODLYZKO A., « The rapid evolution of scholarly communication. » *Learned Publishing*. 15:7-19
- (SAL 05) SALAUN Jean-Michel, « Bibliothèques numériques et Google-Print. », Regards sur l'actualité, La Documentation Française. 01 décembre 2005 La Documentation Française. Disponible sur internet : [http://archivesic.ccsd.cnrs.fr/sic\\_00001576.html](http://archivesic.ccsd.cnrs.fr/sic_00001576.html) (consulté le 7 Janvier 2006)
- (SWA 05) SWAN Alma & BROWN Sheridan, « Open Access self-archiving: an author study », Mai 2005. Disponible sur internet : [http://www.keyperspectives.co.uk/OpenAccessArchive/2005\\_Open\\_Access\\_Report.pdf](http://www.keyperspectives.co.uk/OpenAccessArchive/2005_Open_Access_Report.pdf) (consulté le 7 Janvier 2006)
- (TES 04) Testa J., McVeigh M.E., 2004. *The Impact of Open-Access Journals : A Citation Study from Thomson ISI*. Préparé par Thomson ISI. Disponible sur internet : <http://www.isinet.com/media/presentrep/acropdf/impact-oa-journals.pdf> (consulté le 6 Juin 2005)
- (VEI 04) McVeigh, Marie E. (2004). « Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns ». Disponible sur internet : <http://www.isinet.com/forms/whitepapers/> (consulté le 6 Juin 2005)