

La science sur le Web. Numérisation de textes et dédifférenciation du savoir chez Google Scholar & Cie

Johannes Angermüller¹

Université de Mayence, Allemagne

À travers l'étude de portails et de moteurs de recherche dédiés à la science et à la littérature scientifique, comme Social Sciences Research Network, Google Scholar ou Academia.edu, cette contribution apporte un éclairage sur les conséquences de la numérisation du discours scientifique. Les technologies numériques de recherche et de stockage rendent les textes scientifiques accessibles aux lecteurs non spécialistes. Ensemble, les textes numériques forment une archive numérique mondiale, à l'intérieur de laquelle on peut communiquer au-delà des séparations qui existent entre les différents domaines de connaissance, c'est-à-dire entre les champs scientifiques spécialisés, mais aussi entre discours scientifiques et non scientifiques. De cette manière, alors que le discours scientifique se transforme en un panoptique numérique, les frontières entre l'économie, la politique et la science s'estompent progressivement.

MOTS-CLÉS : MÉDIA NUMÉRIQUE, COMMUNICATION SCIENTIFIQUE ET UNIVERSITAIRE, SPHÈRE SCIENTIFIQUE ET SPHÈRE PUBLIQUE, MOTEUR DE RECHERCHE EN LIGNE, GOUVERNEMENTALITÉ

Science on the Web. Digitalization of Texts and Dedifferentiation of Knowledge. This contribution discusses the impact of scientific research portals like the Social Sciences Research Network, Google Scholar and Academia.edu on scientific discourse. While digital data mining tends to render accessible scientific texts to non-scientific readers, they create a global archive of digital texts. By means of its indexes and rankings, digitalized scientific discourse crosses different fields of specialization and merges with non-scientific spheres. Thus, the boundaries between the economy, politics and science become fragile.

KEY WORDS : DIGITAL MEDIA, SCIENTIFIC COMMUNICATION, SCIENTIFIC/PUBLIC SPHERE, ONLINE RESEARCH PORTALS, GOVERNEMENTALITY

¹ angermue@uni-mainz.de

Pour produire du savoir, les scientifiques recourent, en règle générale, à des livres, à des articles ou à d'autres documents écrits, stockés et repérables dans les archives. Leur savoir est considéré comme scientifique s'il est soumis aux regards critiques d'une communauté d'auteurs spécialisés. Grâce à l'écrit, les productions scientifiques peuvent avoir une résonance dans un vaste espace de communication scientifique. Et grâce au passage du papier aux unités d'information numériques, aux « bits » et aux « bytes », beaucoup plus de lecteurs ont accès à d'encore plus vastes recueils d'écrits scientifiques. Que signifie cependant, pour le processus de production du savoir, le fait que nous n'allions plus seulement dans les bibliothèques pour parcourir laborieusement et consciencieusement les versions papier, mais que nous puissions, en un seul clic, consulter sur nos écrans les produits numériques de l'archive globale du savoir ?

Partant des travaux de Michel Foucault sur l'histoire de la gouvernementalité (2004), je conçois la communication numérique comme une technologie d'information qui ne facilite pas seulement la gestion, le stockage et la recherche de textes, mais qui bouleverse aussi profondément l'organisation sociale de la production de savoir scientifique. Les technologies numériques de l'information font office de « technologies de pouvoir » qui transforment les conditions sociales de la communication scientifique. Grâce aux nouvelles possibilités techniques qu'offre par exemple Google, les textes scientifiques constituent une archive numérique globale, explorée et parcourue à l'aide de machines. Cependant les ordinateurs sont loin d'être des lecteurs neutres, car ils permettent aux textes scientifiques d'être également accessibles aux lecteurs non spécialistes. Et ceci est valable pour les machines tout comme pour les acteurs économiques et politiques, qui utilisent souvent un savoir binaire pour agir dans leurs champs respectifs. Peut-on dire, par conséquent, que la présence des non-spécialistes dans le discours scientifique augmente ? La fin du cloisonnement entre des champs de recherche toujours plus spécialisés aurait-elle sonné ?

Cette contribution comporte deux parties. La première revient sur les pratiques de numérisation des savoirs et sur leurs effets dans le champ des sciences humaines et sociales : comment rendre compte de la science comme une institution sociale si elle est comptée, mesurée, calculée, jusqu'à devenir l'objet des procédures d'évaluation qu'elle a elle-même mises en place ? Le scientisme, croyance en la supériorité de la science – ou du *discours sur la science* – sur toute autre manière d'appréhender la réalité, pourrait bien trouver un terrain de développement favorable dans le règne de la rationalisation chiffrée, désormais étendue à l'ensemble des domaines de la société – y compris le champ scientifique. Pour questionner l'influence de la numérisation sur la constitution même des communautés scientifiques, je m'appuierai principalement sur trois exemples : le portail Social Sciences Research Network (<http://www.ssrn.com>), fondé en 1999 ; le moteur de recherche dédié à la littérature scientifique Google Scholar (<http://scholar.google.com>), apparu en 2004 ; et le « Facebook scientifique » (<http://www.academia.edu>), créé en 2008. La deuxième partie envisagera plus avant les conséquences sociales du traitement des textes électroniques, du point de vue de leur décloisonnement interne, mais aussi externe : le fait que des lecteurs non spécialistes puissent, grâce à la numérisation, avoir accès aux écrits des chercheurs, a-t-il un impact sur la relation entre les domaines scientifiques et d'autres domaines ? Car

si les technologies de communication numérique et les hybrides « textes-chiffres » qu'elles produisent permettent de réguler et de canaliser la production du savoir *dans* le champ scientifique, elles favorisent aussi un processus d'extrapolation du savoir *hors* du champ scientifique. Un double mouvement de décloisonnement semble ainsi à l'œuvre, touchant à la fois la différenciation des savoirs scientifiques entre eux, et la perméabilité grandissante des logiques scientifiques et non scientifiques.

La numérisation de la communication scientifique

Les ordinateurs ont fondamentalement changé le quotidien du travail scientifique. Au milieu des années 1980, le « PC » entame sa marche triomphale. Grâce à lui, on peut désormais rédiger des textes chez soi, sur son écran, les enregistrer sur disquette et les imprimer. Avec les logiciels d'Apple et de Microsoft, les ordinateurs commencent à se répandre dans les ménages, dès le début des années 1990. Au milieu de cette décennie, Internet commence à mettre en réseau les ordinateurs des particuliers. Bientôt la correspondance épistolaire laisse place aux e-mails. Grâce aux moteurs de recherche comme Yahoo ou Google, mais aussi aux catalogues numériques des bibliothèques, il est possible, depuis la fin des années 1990, d'accéder en ligne aux produits du discours scientifique. Et depuis le début de ce siècle, les banques de données facilitent de plus en plus la rédaction de textes (vérificateur d'orthographe, dictionnaires, logiciels de gestion de références bibliographiques, intégration de catalogues, *etc.*).

Les ordinateurs ont sans aucun doute simplifié et accéléré la production du savoir scientifique. Mais si les textes scientifiques, désormais attachés en pièces jointes aux e-mails, ne circulent plus par la poste, ils doivent cependant toujours être lus et rédigés par des humains... Les conséquences exactes du média numérique sur le travail scientifique restent donc à clarifier. Car les ordinateurs ne servent pas seulement aux auteurs à noircir du papier plus rapidement, ou à fouiller plus vite dans la littérature scientifique : avec les nouveaux médias, les règles de la communication scientifique elles-mêmes se transforment. Il convient donc de questionner les changements structurels qui sont à l'œuvre, pour les règles de la production scientifique elle-même : quelles sont les conséquences sur la structure du travail scientifique depuis que nous ne feuilletons plus à la main les catalogues des bibliothèques, mais que nous les parcourons au moyen d'outils numériques ?

Pour montrer l'influence du média numérique sur la formation de communautés scientifiques, je m'appuierai sur trois exemples tirés du discours scientifique. Alors que l'aptitude à saisir le sens d'un texte et à le comprendre constitue la force du lecteur humain, les machines, elles, surpassent ces lecteurs et lectrices par leur capacité à compter, mesurer et corrélérer les textes. Les machines traduisent des textes en chiffres, et ce à une vitesse fulgurante, avec une précision et une fiabilité extrêmes. Par conséquent, grâce à ces chiffres qui apparaissent pour ainsi dire comme des effets secondaires du média numérique, la production scientifique peut être représentée par des répartitions, par des moyennes ou par des scores, ce qui constitue une différence significative par rapport à la communication sur papier.

Le Social Sciences Research Network (SSRN)

Considérons en premier lieu le Social Sciences Research Network (<http://www.ssrn.com>), créé en 1994 et utilisé au départ par des économistes nord-américains. Le SSRN se présente comme une archive ouverte de textes numériques et en particulier de versions préliminaires d'articles, soumis à publication dans des revues. Le SSRN est actuellement constitué de 17 réseaux thématiques disciplinaires différents, qui agrègent plus de 200 000 textes, produits par plus de 100 000 auteurs, occasionnant plus de 500 000 téléchargements par mois. Conformément au principe de libre-échange et de partage des données du Web 2.0, chacun peut s'inscrire sur SSRN et télécharger tous les textes disponibles en ligne. À la différence des bibliothèques classiques, le portail enregistre les réactions des utilisateurs : SSRN répertorie et comptabilise ainsi le nombre de « clics » pour chaque article, qu'il classe alors selon cet indicateur de fréquence. À partir de ces chiffres, SSRN génère une multitude de scores. Un texte très souvent téléchargé remonte dans la hiérarchie de l'archive ; si les téléchargements chutent, il retombe vers le bas. Grâce à un score élevé, un article peut être trouvé facilement, conformément à l'effet Matthieu : « Celui qui a recevra encore » (Merton, 1968). Les indicateurs de téléchargement ainsi calculés constituent la base d'une multitude de tableaux, dans lesquels apparaissent les textes et les auteurs qui ont le plus de « succès » (par semaine, par mois, par an..., classés par chercheur, par département ou par université). Sur la base de ces scores de téléchargements sont générés des hit-parades, dont les valeurs sont actualisées en temps réel, comme les cotes de Wall Street. C'est de cette manière qu'est représentée et constituée numériquement la hiérarchie des universités nord-américaines à la tête de laquelle se trouvent les institutions de l'Ivy League (cf. image 1).

Search Organizations		Last 12 Months					All Time					Authors		
Rank	Institution	Total New Downloads (for all papers)	# of New Papers	New Downloads per paper (for all papers)	Total # of Downloads (for all papers)	# of Papers	Total Downloads Per Paper (for all papers)	# of Authors	Total Downloads Per Author	New Downloads per author				
1	Harvard Business School	138910	(1) 276 (5)	108 (15)	823288	(1) 1677 (5)	540 (4)	256 (3)	3216 (8)	543 (16)				
2	New York University - Leonard N. Stern School of Business	126588	(2) 1522 (2)	42 (231)	549857	(4) 3501 (1)	182 (188)	237 (4)	2320 (18)	534 (17)				
3	University of Chicago - Booth School of Business	125932	(3) 209 (9)	92 (30)	689718	(2) 1662 (6)	504 (10)	174 (9)	3964 (4)	724 (5)				
4	University of Pennsylvania - The Wharton School	102371	(4) 354 (4)	57 (127)	439034	(5) 2238 (3)	244 (116)	283 (2)	1551 (48)	362 (46)				
5	Massachusetts Institute of Technology (MIT) - Sloan School of Management	89467	(5) 185 (11)	85 (37)	413204	(6) 1262 (10)	392 (32)	201 (6)	2056 (24)	445 (31)				
6	University of Virginia - Darden Graduate School of Business Administration	84256	(6) 2203 (1)	35 (274)	249381	(13) 2485 (2)	103 (307)	97 (41)	2571 (13)	869 (2)				
7	Yale School of Management	78988	(7) 145 (17)	84 (38)	561290	(3) 1150 (12)	597 (5)	85 (57)	6603 (1)	929 (1)				
8	London Business School	71542	(8) 218 (8)	70 (70)	304210	(8) 1254 (11)	299 (75)	161 (12)	1890 (31)	444 (32)				
9	Erasmus Research Institute of Management (ERIM) - Joint Research Institute of Rotterdam School of Management (RSM) and Erasmus School of Economics (ESE), EUR	70887	(9) 556 (3)	43 (219)	248813	(12) 1818 (4)	152 (225)	371 (1)	673 (171)	191 (155)				
10	University of Texas at Austin - Red McCombs School of Business	58842	(10) 153 (16)	84 (38)	298198	(9) 877 (18)	428 (25)	135 (18)	2209 (20)	436 (34)				
11	IESE Business School - University of Navarra	58641	(11) 108 (33)	148 (1)	194321	(18) 432 (57)	492 (13)	107 (32)	1816 (34)	548 (15)				
12	Columbia University - Columbia Business School	57293	(12) 198 (10)	51 (169)	317035	(7) 1333 (9)	284 (81)	157 (13)	2019 (27)	365 (45)				

Image 1 : Le « Top des Business Schools » d'après Social Sciences Research Network. Le score est calculé en cumulant tous les téléchargements des auteurs d'une institution. Les articles des membres de la Business School d'Harvard ont été, d'après le graphique, téléchargés 138.910 fois, dans les 12 derniers mois, soit 823.288 fois en tout. Ces chiffres sont ensuite ramenés à la moyenne par article et par auteurs. Ainsi, pour le top 10, les chiffres des téléchargements oscillent entre 152 et 640, ce qui représente, pour chaque auteur, entre 191 et 859 téléchargements.

Une deuxième stratégie pour rendre compte des réactions des lecteurs consiste à compter les citations. Cette stratégie est d'ailleurs actuellement active dans la version bêta de SSRN. Comme pour CiteSeerX (<http://citeseerx.ist.psu.edu/>), archive ouverte agrégeant plus d'1,4 million de textes et 27 millions de citations notamment issues du domaine des sciences de l'information et des médias, les textes qui doivent faire partie de l'archive sont alors traités à l'aide d'un crawler qui reconnaît et compte les références bibliographiques qu'ils contiennent. Ces références sont reconnues à l'aide d'algorithmes complexes, et le nombre de leurs citations dans d'autres textes est calculé. Des scores sont ainsi fabriqués, qui permettent de faire apparaître plus facilement les textes les plus fréquemment cités. Les utilisateurs peuvent aussi ne consulter que les références bibliographiques. En outre, SSRN offre la possibilité de consulter ces références selon un classement hiérarchique, c'est-à-dire en commençant par la référence la plus citée ou la plus téléchargée.

Concernant le calcul des fréquences de téléchargement et du nombre de citations, SSRN fabrique des chiffres qui peuvent être lus comme des indicateurs « objectifs ». Dans le domaine du discours scientifique sur papier, les hiérarchies symboliques reposent toujours sur une réputation « subjective », c'est-à-dire sur un savoir que les auteurs ont accumulé sur leurs propres pairs, au fil des années ; tandis qu'avec l'archive numérique, les hiérarchies symboliques peuvent être représentées numériquement, et ce en appuyant simplement sur un bouton. Ce qui compte, désormais, c'est ce qui peut être compté.

S'il est contestable que les chiffres de SSRN puissent prétendre à l'objectivité d'indicateurs obtenus de manière très contrôlée, force est cependant de constater qu'une importante communauté scientifique, très active, s'est formée autour de ce portail. Certains auteurs ont même commencé à rédiger des contributions spécialement pour SSRN. Dans ce contexte, la question n'est plus de savoir s'il peut y avoir des indicateurs représentant la qualité de la recherche. La question est plutôt de savoir comment, grâce aux nombreux chiffres de l'archive numérique, certains produits acquièrent plus de visibilité, voire de pertinence que d'autres pour le lecteur.

Google Scholar

Mis en place en novembre 2004, Google Scholar (<http://scholar.google.fr/>), le moteur de recherche scientifique du groupe Google, est un autre exemple de la numérisation du discours scientifique. À la différence du moteur de recherche standard, Google Scholar ne recherche que les textes scientifiques. Google Scholar répertorie tous les textes scientifiques accessibles, mais aussi les revues qui composent des grandes collections de cd-rom, et de nombreux livres protégés par des droits d'auteur. Par conséquent, la recherche de littérature scientifique devient avec Google Scholar une sorte de recherche de l'intégralité des textes de l'archive scientifique mondiale. Google Scholar envisage pratiquement tous les documents scientifiques disponibles au format électronique et non, comme le *Web of Science* de l'*Institute for Scientific Information* (ISI), uniquement des sélections de revues de certaines aires linguistiques ou de certaines disciplines. Des accords avec d'importantes maisons d'édition, comme Springer, Elsevier ou Sage, permettent à Google d'explorer les contenus de leurs revues et de leurs livres sans que l'utilisateur y ait un accès complet. Ces accords ont rendu possible l'indexation de ressources protégées par des droits d'auteur. Google Scholar est ainsi le résultat de l'intégration de composants numériques différents : il agrège aussi bien, par exemple, des documents de particuliers disponibles librement sur Internet, des manuscrits de maisons d'édition utilisés à des fins commerciales, que les livres numériques de Google Books, qu'ils soient protégés par des droits d'auteur ou libres de droits.

La prouesse technologique réside dans le fait que Google Scholar parvient, grâce à des algorithmes complexes, à différencier les textes scientifiques des textes non scientifiques, et à les trier selon des critères de pertinence. Les textes dont la thématique n'est pas pertinente sont filtrés, voire relégués au bas du classement. Le moteur de recherche établit en outre, tout comme CiteSeerX, l'identité

bibliographique d'un texte scientifique, et peut reconnaître les références qui y sont utilisées et les ranger sous la même étiquette que le texte original. De cette manière, on peut, avec Google Scholar, savoir comment les textes renvoient les uns aux autres. Les références apparaissent sous forme d'hyperliens, grâce auxquels on peut accéder aux textes originaux. À la différence des indicateurs scientométriques classiques de l'ISI, qui compte les citations dans les revues sélectionnées, on peut, par un simple clic sur les chiffres établis, faire apparaître directement les citations répertoriées – tant qu'il ne s'agit pas de documents de maisons d'édition protégés par des droits d'auteur. Les utilisateurs et utilisatrices ne voient donc pas seulement le nombre des citations et des références qui permettent de mesurer et de hiérarchiser l'espace de communication ; ils peuvent aussi se faire une idée des contextes spécifiques d'utilisation des textes dans le discours scientifique. Les données chiffrées de Google sont, de ce fait, bien plus qu'un accessoire arithmétique. Elles participent au calcul de cotes et de scores grâce auxquels Google Scholar met ses résultats à la disposition des utilisateurs et utilisatrices (cf. image 2).

Web Images Vidéos Maps Actualités Groupes Gmail plus ▼ Connexion

Google scholar foucault science discours [Rechercher] Recherche avancée Scholar Préférences Scholar

Rechercher sur le Web Rechercher les pages en français

Scholar Tous les articles Articles récents Résultats 1 - 10 sur un total d'environ 16 900 (0,11 s)

[PDF] ► **Surveiller et punir**
M Foucault - 1975 - education-frederique.com
... MÉDECINE, la SEXUALITÉ, l'ÉTHIQUE, les SCIENCES DE L ... que chaque époque produit un discours dominant censé ... POUVOIR et SAVOIR Pour Foucault Le pouvoir n'est ...
Cité 1527 fois - Autres articles - Version HTML

[LIVRE] **Histoire de la sexualité: La volonté de savoir**
M Foucault - 1976 - Gallimard
Cité 653 fois

[LIVRE] ... et les choses: une archéologie des sciences humaines: une archéologie des sciences ...
M Foucault - 1966 - Gallimard
Cité 712 fois - Autres articles - Les 3 versions

[PDF] ► **Foucault**
M Foucault, P Mazumdar - 1998 - images.hachette-livre.fr
... 575 Foucault: Droit de mort et pouvoir sur la vie ... d'in-sister ici sur la rupture qui s'est alors produite dans le régime du discours scientifique et sur ...
Cité 52 fois - Autres articles

[LIVRE] **Oublier Foucault**
J Baudrillard - 1977 - Editions Galilée
Cité 92 fois

[LIVRE] **Au bord de la falaise: l'histoire entre certitudes et inquiétude**
R Charlier - 1998 - A. Michel
Cité 91 fois - Autres articles

Image 2 : Exemple d'une page de résultats sur Google Scholar. Les références correspondant aux mots-clés « Foucault », « discours », « science » sont affichées en fonction de la fréquence à laquelle elles sont citées. Deux de ces références renvoient à des textes numériques accessibles en ligne. Un clic sur l'onglet « Cité » conduit à chaque fois aux titres qui contiennent ces références.

Ainsi, Google Scholar grouille de données chiffrées, qui définissent l'ordre d'apparition des textes scientifiques mis à la disposition des utilisateurs. Quelles sont les différences par rapport aux indicateurs commerciaux fournis par le *Web of Science* de l'ISI ? Comme pour ces indexations de citations, il s'agit pour Google Scholar d'une technologie qui fonctionne principalement comme une boîte noire. De quelle autre manière pourrait d'ailleurs agir Google, qui poursuit malgré tout un but lucratif ? Cependant, Google propose les résultats de ses recherches gratuitement et en libre accès. Google Scholar offre ainsi des représentations de la réalité scientifique qui, dans le quotidien pratique de la recherche, peuvent être utiles. Alors que les données chiffrées d'ISI sont élaborées à l'intention des commissions, des administrations et d'autres pôles de décision, celles de Google Scholar sont aussi intéressantes, mais autrement, puisqu'elles sont relativement transparentes (on peut consulter rétrospectivement les passages cités) et fusionnent avec les produits scientifiques (au lieu de lire un texte donné, on lit un texte, tout en obtenant l'information qu'il est cité x fois). Les hybrides « textes-chiffres » que produit Google Scholar ont tendance à dépasser la dichotomie qui existe entre la capacité de contrôle et de planification d'une part, et la capacité de réflexion d'autre part. Les chiffres contribuent ainsi à la différenciation du discours scientifique, dans la mesure où leur lisibilité quasi-universelle augmente la perméabilité entre les champs du savoir. Des discours, auparavant séparés, s'interpénètrent grâce à des plateformes ou à des systèmes informatiques de plus en plus intégrés. Ainsi, les catalogues de bibliothèques intègrent parfois les entrées de Wikipédia, qui contiennent elles-mêmes des renvois aux catalogues de livres, qui sont à leur tour reliés aux logiciels de gestion bibliographique locale, qui ont eux aussi un accès aux catalogues des bibliothèques. Google Scholar offre en outre des liens directs au catalogue de la bibliothèque de l'université la plus proche. De cette manière, les contenus des banques de données scientifiques peuvent passer d'une plateforme à l'autre.

L'archive électronique se construit donc au-delà des frontières établies par une discipline ou par une langue. Avec les archives ouvertes et les moteurs de recherche, le règne des statistiques, des moyennes, des probabilités fait son entrée dans les sciences humaines et sociales. Si les sciences sociales, notamment avec la statistique comme science du gouvernement, ou les analyses quantitatives, n'étaient pas étrangères à la culture du chiffre (Dear, 1995), c'est seulement depuis les années 60 que la production scientifique est de plus en plus représentée sous forme de chiffres. Ainsi, après avoir étendu les outils de la raison statistique à toujours plus de terrains de la vie sociale (Desrosières, 1993), la science commence-t-elle à se prendre pour objet avec ses propres moyens. Après la scientométrie de l'ISI, les technologies numériques généralisent la mesure, l'observation, voire la surveillance, sans que cela puisse pour autant être centralisé par des acteurs. La représentation chiffrée des hiérarchies symboliques, calculée à la virgule près, donne naissance à un panoptique, dans lequel aucun domaine spécialisé, aussi réduit soit-il, aucune niche, aussi petite soit-elle, n'échappe au regard numérique. Ainsi, à la différence du savoir numérique de l'ISI, dont la portée est en règle générale limitée à l'action administrative (évaluation des produits, recrutement des producteurs, répartition des ressources...), les chiffres

qui prolifèrent dans les médias numériques se glissent dans le processus de recherche. C'est grâce à ces chiffres que les producteurs peuvent s'orienter dans une archive numérique globale qui leur serait inaccessible sans les classements et représentations numériques. Et c'est grâce à ces chiffres que les producteurs peuvent s'auto-surveiller en comparant les scores de chacun. Google Scholar se révèle être une technologie avec laquelle les chercheurs deviennent gouvernants d'eux-mêmes, en analysant, constamment et en temps réel, l'impact des citations d'autres chercheurs – ou des leurs.

Academia.edu

Face à SSRN et à Google Scholar, grâce auxquels de vastes collections de textes peuvent désormais être explorées, on note récemment l'importance grandissante des technologies du Web 2.0, qui regroupe de larges populations d'auteurs. Pour se servir de ces portails, les utilisateurs doivent tout d'abord s'inscrire et s'enregistrer ; il est alors possible de les compter et de les contrôler, si ce n'est en tant qu'individus, au moins à l'échelle de la population. Dans le domaine de la science, beaucoup de ces communautés virtuelles reposent sur un engagement volontaire et sur une initiative idéaliste – on pense aux sites interactifs et listes électroniques (par exemple <http://www.analysedudiscours.net/>). Concernant ces initiatives, la frontière entre l'initiative citoyenne et la motivation stratégique n'est pas toujours facile à établir. On peut prendre comme exemple Academia.edu (<http://www.academia.edu/>), le « Facebook des scientifiques », créé en 2007 par un chercheur qui voulait trouver des collègues travaillant sur le même domaine de recherche que lui. Pour se servir d'Academia.edu, les utilisateurs doivent s'abonner à certains domaines, rubriques disciplinaires ou sous-disciplinaires. L'astuce d'academia.edu réside dans le fait que ces rubriques ne sont pas définies *a priori*. Selon les demandes, on peut même en créer de nouvelles, qui peuvent être séparées et constituer des disciplines différentes, ou bien être regroupées sous une même étiquette. En plus des fonctions classiques d'une page Web, comme l'affichage d'informations sur les modalités pour contacter la personne, d'un court CV et de quelques textes téléchargeables, Academia.edu offre la possibilité de « suivre » certains collègues que l'on choisit, c'est-à-dire comme sur Facebook, de devenir « amis » avec eux. De cette manière, certains auteurs peuvent se contacter de manière informelle et regrouper leurs travaux sous des rubriques partagées (cf. image 3).

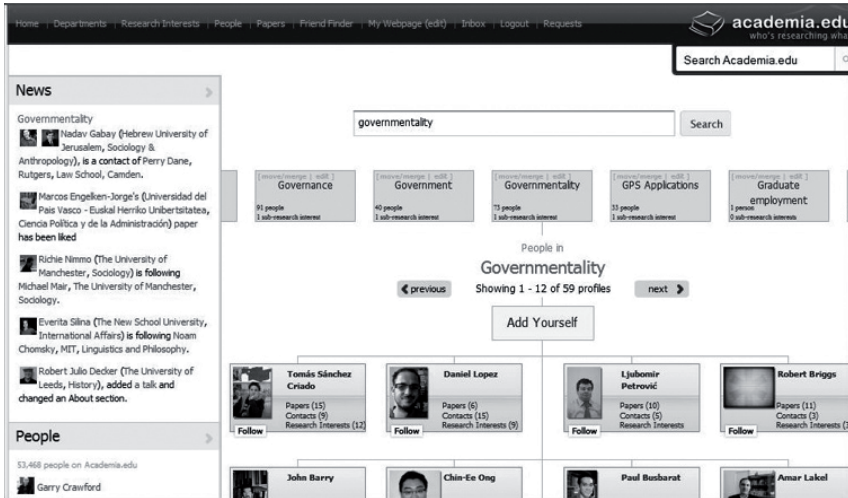


Image 3 : Sur Academia.edu, un bandeau comporte les boîtes grises correspondant aux rubriques thématiques définies par les utilisateurs, dans lesquelles ces derniers peuvent s'inscrire. Ici aussi on trouve de nombreuses données chiffrées qui indiquent le nombre d'utilisateurs inscrits dans une rubrique, le nombre d'articles, de contacts, ainsi que les centres d'intérêt scientifique.

Il est encore trop tôt pour dire comment ces projets Web 2.0, qui émanent de la « base », influencent la formation de communautés scientifiques. Si nombre de ces initiatives vont sans doute sombrer dans l'anarchie propre à l'autogestion, on peut supposer que l'influence des technologies numériques va venir du haut de la hiérarchie et agir sur les grandes populations d'auteurs, et ce dans le domaine de la gestion intra-universitaire (par exemple en ce qui concerne la rédaction de rapports de recherche) comme de la gestion politico-administrative des universités (par exemple concernant l'établissement de budgets globaux et de contrats d'objectifs et de moyens) et du financement de la recherche (en particulier sur le plan européen). On peut penser aux portails de recherche dans lesquels les universitaires doivent rendre compte de leurs activités. La mission officielle consiste ici à encourager l'échange entre la sphère scientifique et la sphère publique. Une autre tâche, plus officieuse celle-là, est sans aucun doute de surveiller de manière chiffrée les activités des chercheurs. Les portails dédiés à la recherche affichent des détails précis sur les publications et les projets des membres de grandes écoles et d'institutions scientifiques (<http://www.forschung-sachsen-anhalt.de/>). Ces données sont traitées statistiquement de multiples manières pour être ensuite compilées. Là aussi, les informations enregistrées dans le contexte de pratiques d'évaluations internes aux universités et grandes écoles, peuvent facilement passer dans une autre banque de données. Les annonces de publications, faites sur les différents portails publics dédiés à la recherche, sont directement reprises par les catalogues des

bibliothèques. Les projets de recherche apparaissent automatiquement sur les portails d'associations de chercheurs ou sur les portails qui subventionnent la recherche. À l'inverse, les catalogues des bibliothèques sont de plus en plus reliés à Wikipédia, qui, à son tour, renvoie à des catalogues de livres. Face à l'interpénétration des plateformes numériques, les auteurs ont de moins en moins de contrôle sur leur propre profil. Ce sont, tour à tour, les gestionnaires et les programmeurs des dispositifs techniques, les nouveaux technologues du pouvoir, qui contrôlent la diffusion des informations sur les auteurs et leurs écrits.

L'archive numérique mondiale et la dédifférenciation du savoir scientifique ?

La science communique à l'aide de textes conservés dans des archives ; le passage aux médias numériques n'y change rien. Nous sommes toujours de lire des textes, de nous y référer, et de prendre appui sur leur contenu pour prétendre à une vérité scientifique. Et nous devons toujours rédiger des textes en nous référant au savoir produit par d'autres et en citant nos sources (Angermüller, 2007). Grâce aux technologies numériques, les productions scientifiques sont effectivement susceptibles de toucher un nombre considérablement plus élevé de lecteurs et de lectrices, avec une rapidité, une précision et une universalité jusqu'à présent inédites. Qu'est-ce que cette augmentation des possibilités de communication change pour les règles de la production scientifique ?

Selon le courant « institutionnaliste » en sociologie de la science, la science moderne est le résultat d'un processus de différenciation. Dans la dynamique de ce processus, d'une part, la science affirme, vis-à-vis de l'extérieur, son autonomie face aux forces que sont la politique, l'économie, et les médias. D'autre part, à l'intérieur même de sa sphère, elle se divise en un nombre croissant de disciplines, de champs et de thèmes (Weingart, 2003). Avec le passage des livres et des bibliothèques de l'ère Gutenberg aux bits et aux bytes de l'époque numérique, ce processus rencontre peut-être une limite. Tandis que le rapport entre la sphère scientifique et la sphère publique est renégocié, des tendances de dédifférenciation et de « défrontiérisation » s'amorcent et vont à l'encontre de l'image que la science moderne se fait d'elle-même (Lash, 1990).

Ce que la communication numérique apporte de nouveau, c'est que des lecteurs humains trouvent une extension par les lecteurs non humains – par des machines qui sont certes incapables de comprendre le sens des textes, mais qui sont en revanche imbattables quand il s'agit de transformer le sens de ces textes en données chiffrées. Alors que des lecteurs humains lisent les textes en traçant leurs sources et en associant les contenus aux acteurs du discours intellectuel, les lecteurs non humains ne font qu'enregistrer des occurrences. Cela ne signifie pas pour autant qu'une technologie comme Google lit mal : elle lit juste *autrement*. Pour les machines, il n'existe que deux modalités

d'expression : événement ou non événement. Seule la référence attestée compte, et ce, indépendamment de son mode d'énonciation. Cette transformation des différentes modalités d'expression constitue la condition pour que les machines produisent quelque chose que les lecteurs humains ne peuvent pas produire : des statistiques, des comptes de vastes corpus, de la lecture algorithmique, *etc.*

Ainsi, les données chiffrées de l'archive numérique mondiale résultent d'un processus d'abstraction dans lequel les produits linguistiques sont séparés de leurs contextes épistémiques. Elles vont à l'encontre de la création de domaines scientifiques toujours plus spécialisés et autonomes, dans la mesure où elles replacent le produit scientifique isolé dans un ensemble toujours plus vaste. C'est la raison pour laquelle les stratégies disciplinaires et sous-disciplinaires, qui visent à la création de rubriques et de champs encore plus spécialisés, pourraient perdre de leur valeur. Si, avec l'expansion du domaine de la recherche, les producteurs essaient d'acquérir une légitimité incontestable en occupant des niches de plus en plus spécialisées, les représentations numériques de la production scientifique permettent, à un certain degré, de comparer des domaines de recherche différents. Grâce aux indicateurs numériques, tous peuvent s'observer, sur un terrain où tout peut être comparé avec tout. Le succès scientifique ne peut plus être envisagé séparément de l'ensemble de plus en plus global que forme l'archive numérique. En rendant l'immensité du savoir mesurable et comparable à grande échelle, les données chiffrées donnent naissance à de nouvelles hiérarchies symboliques dans le discours scientifique. Tout en haut des classements de Google Scholar ou de SSRN, apparaissent des auteurs interdisciplinaires, qui ont tendance à éclipser les mandarins disciplinaires. Google Scholar ne se soucie en effet nullement de savoir si un nom est très souvent cité au sein d'une seule et même discipline, ou s'il n'est que peu cité en soi, mais dans de nombreuses disciplines.

Les textes scientifiques ne sont cependant pas seulement destinés à être lus par des machines. Grâce aux données chiffrées, ils peuvent être lus par des lecteurs et des lectrices humains qui n'appartiennent pas à la communauté des auteurs. Ces données chiffrées sont en effet « parlantes » dans une multitude de champs disciplinaires, voire en dehors de la sphère scientifique. L'arithmétique est la *lingua franca* par excellence, qui ne permet pas seulement d'établir le contact entre les lecteurs humains et non humains, mais qui relie aussi des lecteurs humains entre eux. Les données chiffrées permettent aux scientifiques et aux non scientifiques d'entrer en contact – on pense ici aux possibilités que Google Scholar, par exemple, offre aux nombreux lecteurs non universitaires qui jusqu'à présent n'avaient aucun accès aux bibliothèques, ou ne souhaitaient pas en avoir. Ce nouveau public non scientifique a désormais son mot à dire, en plaçant en haut du classement grâce à ses clics et à ses téléchargements des thèmes qui présentent un intérêt de masse. Il ne faut pas négliger non plus que de toutes nouvelles possibilités puissent se présenter aux auteurs scientifiques lorsqu'ils touchent un public non scientifique. Si le discours scientifique peut désormais être lu par des lecteurs qui ne possèdent aucune notion scientifique, si les données chiffrées permettent à ces non scientifiques d'accéder à la hiérarchie symbolique du discours scientifique, c'est aussi la relation avec les pouvoirs politiques et économiques externes qui est en jeu (Gibbons *et al.*, 1994).

Cette hybridation entre savoir scientifique et savoir non scientifique dans le média numérique n'est pas sans conséquence sur la science en tant qu'institution. L'archive numérique mondiale ne s'est, en effet, pas construite en respectant les frontières disciplinaires. Elle ne reflète pas non plus l'influence structurante que l'État exerçait traditionnellement sur l'enseignement supérieur. Les portails dédiés à la recherche scientifique comme SSRN, Google Scholar et Academia.edu sont des technologies qui émanent d'initiatives privées et qui interfèrent de manière transversale avec les disciplines, renforçant peut-être le phénomène des *Studies*, ces champs de recherche au carrefour des disciplines. Elles annoncent un stade postmoderne dans l'évolution de la science, dans la mesure où elles ont pour objectif d'administrer un terrain de plus en plus décloisonné de la production scientifique. On assiste ainsi à la naissance d'un véritable dispositif de technologies du savoir qui soumettent la production scientifique aux regards des pouvoirs administratifs et politiques (Power, 1997). La production scientifique commence à ressembler désormais aux technologies de production qu'on connaît dans la grande industrie (taylorisme, qualité totale, entrepreneurialisation de l'État...).

Si cette évolution touche à l'autonomie de la science face aux instances hétéronomes de la politique et de l'économie (Slaughter et Leslie, 1997), l'identité scientifique est peut-être en train de se transformer. De la rhétorique d'un savoir pur, inaccessible pour les lecteurs non scientifiques, on passe au régime de « l'excellence » (Readings, 1996) qui vise à mesurer et classer le savoir scientifique ainsi qu'à le représenter sous forme de chiffres. Ce régime d'excellence qui s'établit actuellement dans nombre de pays européens s'articule souvent avec des pratiques technocratiques dans la gestion des universités. Dans la mesure où l'on utilise des indicateurs numériques afin d'évaluer un savoir dont on ne sait rien, ces pratiques témoignent d'un pouvoir technocratique qui se réclame des outils scientifiques pour contrôler le domaine du savoir scientifique. Or, les sciences sociales ne doivent pas rester muettes devant cette évolution. Si leur tâche n'est pas de renforcer l'idéologie scientiste du pouvoir non scientifique, elles peuvent contribuer à la réflexion des effets qu'a la transition au pouvoir-savoir numérique dans le domaine du savoir scientifique. Ainsi, les hybrides « textes-chiffres » de la communication numérique modifient-ils la relation entre la sphère scientifique et la sphère publique tout en bouleversant la relation entre pouvoir et pensée. Si la réflexion et la gestion s'interpénètrent de plus en plus, ce n'est pas seulement le pouvoir et le savoir qui fusionnent, c'est aussi la différenciation du social dans des domaines spécialisés (science, économie, politique *etc.*) qui est en jeu. Peut-on dire alors que s'annonce, avec le passage aux technologies numériques, la fin de la différenciation de la science en champs de recherche de plus en plus spécialisés ? Il faudra encore davantage de recherches pour déterminer exactement les conséquences de la numérisation du discours scientifique sur la science en tant qu'institution.

Traduit de l'allemand par Stéphanie Roussel.

RÉFÉRENCES

Angermüller, Johannes, 2007. *Nach dem Strukturalismus. Theoriediskurs und intellektuelles Feld in Frankreich*, Bielefeld : transcript.

Dear, Peter Robert, 1995. *Discipline and Experience. The Mathematical Way in the Scientific Revolution*, Chicago : University of Chicago Press.

Desrosières, Alain, 1993. *La Politique des grands nombres : histoire de la raison statistique*, Paris : La Découverte.

Foucault, Michel, 2004. *Territoire, Population, sécurité*, Paris: Gallimard, Seuil.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzmann, P. Scott et M. Trow, 1994. *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London : Sage.

Lash, Scott, 1990. *The Sociology of Postmodernism*. London, New York : Routledge.

Merton, Robert K., 1968 : 56-63. « The Matthew Effect in Science », *Science*, 159 (3810).

Power, Michael, 1997. *The Audit Society. Rituals of Verification*, Oxford : Oxford University Press.

Readings, Bill, 1996. *The University in Ruins*. Cambridge, MA, London : Harvard University Press.

Slaughter, Sheila, Larry, L. Leslie, 1997. *Academic Capitalism. Politics, Policies, and the Entrepreneurial University*, Baltimore, London : The Johns Hopkins University Press.

Weingart, Peter, 2003. *Wissenschaftssoziologie*, Bielefeld : Transcript.