



HAL
open science

Quelques observations bibliométriques sur la culture informationnelle

Alexandre Serres

► **To cite this version:**

Alexandre Serres. Quelques observations bibliométriques sur la culture informationnelle. 2008. <hal-02383265>

HAL Id: hal-02383265

<https://hal.science/hal-02383265v1>

Preprint submitted on 27 Nov 2019

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.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Quelques observations bibliométriques sur la culture informationnelle
--

Février 2008

Note : Cette étude n'a jamais été publiée ; elle a été réalisée pour la préparation du chapitre pour l'ouvrage « Les problématiques émergentes en Sciences de l'information »

Que représentent les expressions « *culture informationnelle* », « *culture de l'information* », « *maîtrise de l'information* », « *information literacy* », dans l'indexation de la littérature scientifique sur ces domaines ? Combien de résultats trouve-t-on sur ces requêtes, dans différents outils spécialisés ? Quelle est la part respective de chacune de ces expressions ? Combien de thèses ont été soutenues et quelle est l'évolution dans le temps ? Combien de citations entre publications peut-on trouver sur ces expressions ? Quelles sont les évolutions du nombre de publications, depuis une dizaine d'années, sur l'ensemble des publications ? Quelle est la part des publications françaises dans l'ensemble de la littérature du domaine ?

Pour tenter de répondre à ces questions, nous avons mené, en octobre et décembre 2007, quatre sortes d'analyses « bibliométriques » :

- une comparaison des résultats trouvés sur différents outils, à partir de requêtes identiques ;
- une observation de l'évolution du nombre de thèses et mémoires, soutenus en France sur ces thématiques ;
- une recherche dans Article@inist, sur l'évolution du nombre d'articles publiés et référencés sur cette base de données ;
- enfin une étude longitudinale sur l'évolution du nombre de publications et de citations, menée avec un outil de scientométrie libre et gratuit, Harzing Publish or Perish¹, qui exploite les résultats de Google Scholar.

Une remarque préalable nous semble nécessaire : si ce travail d'enquête a été mené aussi rigoureusement que possible, il ne prétend nullement constituer une véritable recherche bibliométrique, systématique et approfondie, comme celle réalisée récemment par deux chercheurs indiens, Mohamad Nazim et Moin Ahmad². Au départ, nous avons fait des recherches assez simples, dans le cadre de la rédaction d'un texte sur la culture informationnelle comme problématique émergente³, pour évaluer, précisément, le caractère émergent ou non de cette thématique dans les publications. Et une analyse entraînant une autre, nous avons au final accumulé un certain nombre de données, qu'il nous paraît intéressant de rendre publiques en l'état.

1/ Comparaison des résultats à partir de requêtes identiques sur différents outils

La première étude a porté sur les **sept sources et outils** suivants :

¹ **Harzing PoP** (*Publish or Perish*) est un outil à télécharger sur son ordinateur ; il est gratuit et accessible sur : <http://www.harzing.com/resources.htm#/pop.htm>

² Nazim, M., Ahmad, M., « Research Trends in Information Literacy : A Bibliometric Study ». *SRELS Journal of Information Management*, 2007, vol. 44, n° 1, pp. 53-62. Disp. sur : http://eprints.rclis.org/archive/00012388/01/Research_trends_in_IL.doc

³ Il s'agit d'une contribution à un ouvrage collectif pourtant sur « Les problématiques émergentes en Sciences de l'information », coordonné par Fabrice Papy (enseignant-chercheur à Paris 8) et publié chez Hermès (parution prévue à l'automne 2008).

- **Google Scholar** : module spécialisé de Google sur l'information scientifique, et devenu aujourd'hui l'un des principaux points d'accès universels à la littérature scientifique mondiale, en dépit de nombreux points faibles (notamment l'absence d'indications sur les sources indexées) ;
- le **Sudoc** : catalogue collectif des bibliothèques de recherche françaises, recensant notamment la quasi-totalité des thèses soutenues en France ;
- **Article@inist**, le catalogue des articles et des monographies de l'INIST ;
- **TEL** (Thèses en ligne) : archive ouverte multi-disciplinaire pour les thèses ;
- **HAL** (Hyper Articles en Ligne) : archive ouverte multi-disciplinaire pour les articles, communications, etc. ;
- **ArchiveSic** : archive ouverte en Sciences de l'Information et de la Communication ;
- **Memsic** : archive ouverte pour les mémoires de 3^{ème} cycle (DEA, DESS, Master, mémoires professionnels, etc.) en Sciences de l'Information et de la Communication.

Les requêtes ont porté sur **cinq expressions** : *culture de l'information*, *culture informationnelle*, *maîtrise de l'information*, *information literacy*, *formation des utilisateurs* ou *formation des usagers*.

Deux recherches différentes ont été menées sur le SUDOC : l'une sur tous les types de documents, par une requête sur les Mots sujet, l'autre seulement sur les thèses, par des requêtes en mode avancé, soit sur le champ Résumé de thèses, soit sur les Mots sujet (pour Formation des utilisateurs).

**Requêtes faites le 31 octobre 2007, en mode simple, sur tous les champs :
(en gras, les résultats les plus importants)**

	<i>Google scholar</i>	<i>Sudoc</i>	<i>Article@ Inist</i>	<i>Sudoc (thèses)</i>	<i>Tel</i>	<i>Archive Sic</i>	<i>Hal</i>	<i>Memsic</i>	Total
<i>Culture de l'information</i>	355	318	544 (540 articles)	4	0	3	3	2	1229
<i>Culture informationnelle</i>	105	19	12	3	0	4	4	0	147
<i>Maîtrise de l'information</i>	735	296	57	14	0	5	6	0	1113
<i>Information literacy</i>	19 800	32	654	2	2	2	2	0	20494
<i>Formation des utilisateurs (ou des usagers)</i>	844	159	137	24	0	4	4	9	1181

Quelles observations peut-on tirer de ce premier jeu de requêtes ?

- la prédominance écrasante de l'expression *information literacy*, (20 494 résultats) par rapport aux diverses traductions françaises (1229 pour « culture de l'information »), ce qui souligne, *a contrario*, la faiblesse des publications scientifiques en français sur cette thématique. L'examen rapide des résultats, notamment sur Google Scholar, permet de vérifier aisément l'hégémonie absolue des publications anglo-saxonnes ;
- l'importance de la « *culture de l'information* », nettement préférée à « *culture informationnelle* » (1229 résultats contre 147), et qui l'emporte même sur la « *maîtrise de l'information* » (1113 résultats), pourtant considérée généralement comme la traduction française la plus courante de *information literacy* ;
- dans le SUDOC, la prééminence du mot-clé « *formation des utilisateurs* » pour les thèses (24

- thèses contre 14 indexées avec « maîtrise de l'information ») ;
- la grande faiblesse des archives ouvertes (HAL, Tel, ArchiveSic, Memsic) sur ces thèmes : 50 résultats toutes archives confondues⁴ ;
 - la domination absolue de Google Scholar sur les autres sources (hormis sans doute Article@Inist pour les articles sur la culture de l'information) ;

2/ Evolution du nombre de thèses et mémoires soutenus en France

Une série de requêtes identiques, portant sur les mêmes expressions (sauf *information literacy*, très peu utilisé dans le Sudoc), a été également menée dans le catalogue Sudoc, sur les thèses.

Recherches, faites le 31 octobre 2007 sur le SUDOC, sur l'évolution du nombre de thèses et mémoires professionnels :

- les requêtes ont été faites en Mode avancé, avec filtrage sur les « Types de publication », avec les expressions entre guillemets ; pour les expressions « maîtrise de l'information », « culture de l'information » et « culture informationnelle », elles ont porté sur le champ Résumé et sur le champ Mots du Titre, les résultats trouvés s'additionnant ; pour l'expression « formation des utilisateurs », les requêtes ont porté sur le champ Mot Sujet ;
- il faut noter que la recherche par expression (*i.e.* la recherche d'une chaîne de caractères, mise entre guillemets, comme « maîtrise de l'information ») semble fonctionner de manière assez aléatoire sur le Sudoc : quand le catalogue ne trouve pas l'expression, il cherche sur tous les termes, par un ET logique, ce qui génère évidemment beaucoup de bruit documentaire. Les résultats ci-dessous ont donc dû être retravaillés, pour éliminer le bruit (ou le silence) et il est difficile d'obtenir des résultats vraiment fiables et complets.

SUDOC : évolutions des thèses et mémoires	Total	< 2000	> 2000	2000	2001	2002	2003	2004	2005	2006
<i>Culture de l'information</i>	5	0	5	1		3			1	
<i>Culture informationnelle</i>	4	2	2					1		1
<i>Maîtrise de l'information</i>	19	7	12	2	2	0	4	3	1	
<i>Formation des utilisateurs</i>	24	2	22	1	1	2	5	13	0	

Quelques remarques :

- sur l'indexation et le choix des mots-clés : il est intéressant de noter que sur le total des 19 thèses trouvées avec le mot-clé « *maîtrise de l'information* », 14 thèses portaient sur la maîtrise de l'information dans l'entreprise, la veille ou l'intelligence économique, et trois sur la formation des utilisateurs dans les bibliothèques ; dans le Sudoc, le mot-clé retenu pour la maîtrise de l'information, au sens classique de *information literacy*, est d'abord « formation des utilisateurs »

4 Ce chiffre de 50 ne signifie nullement 50 publications distinctes, car un même texte est souvent indexé avec plusieurs de ces mots-clés dans les archives ouvertes.

- ;
- sur l'évolution du nombre : il faut relever la très nette augmentation du nombre de thèses et mémoires professionnels à partir de 2000, pour toutes les entrées, notamment la « formation des utilisateurs », qui passe de 2 avant 2000 à 22 depuis 2000, avec une nette progression depuis 2003 ; *idem* pour la culture de l'information, où les 5 résultats trouvés concernent des thèses ou mémoires soutenus depuis 2000.

3/ Evolution du nombre d'articles de revues dans *Article@Inist* (31 octobre 2007)

Requêtes		Total	< 2000	> 2000	2000	2001	2002	2003	2004	2005	2006	2007
<i>Culture de l'information</i>	Tts lang.	540	16	524	92	96	89	34	154	50	9	0
	Fr.	17	10	7	0	3	0	1	3	0	0	0
<i>Culture informationnelle</i>	Tts lang.	12	4	8	0	2 r.i.	1	0	0	4	1	0
	Fr.	9	4	5	0	2	0	0	0	2	1	0
<i>Maîtrise de l'information</i>	Tts lang.	57	21	36	2	4	5	2	2	19	1	1
	Fr.	49	17	32	2	4	5	2	2	16	1	0
<i>Information literacy</i>	Tts lang.	654	112	542	48	56	80	69	53	93	87	56
	Fr.	14	0	14	0	0	0	0	1	13	0	0

Notes :

- r.i. : références identiques
- Tts lang. : résultats en toutes langues
- Fr. : résultats en français

Observations des résultats :

Cinq observations principales peuvent être faites, à partir de ces recherches sur les articles de revues dans le catalogue de l'INIST :

- **l'importance des résultats avec *information literacy*** (654) et *culture de l'information* (540) ;
- le **choix de l'INIST d'utiliser *culture de l'information*** plutôt que *maîtrise de l'information* ;
- **l'utilisation extrêmement faible de *culture informationnelle*** (12) ;
- la **très nette majorité des résultats se situe après 2000**, surtout pour les articles sur la culture de l'information, avec un « pic » en 2004 ;
- enfin, le **nouveau constat de la faiblesse des résultats en français** (17 sur 540 pour culture de l'information, 14 sur 654 pour information literacy) ; mais il faut remarquer que les expressions *maîtrise de l'information* (49 sur 57) et surtout *culture informationnelle* (9 sur 12) concernent davantage les articles en français.

4/ Etude bibliométrique, sur *Harzing Publish or Perish*, menée le 23 décembre 2007 :

4.1 Evolution du nombre de publications et de citations sur les requêtes suivantes :

		Total sans dates	< 2000	> 2000	2000	2001	2002	2003	2004	2005	2006	2007	Total dat.
Culture de l'information	<i>Nb pub.</i>	408	60	133	11	19	24	25	21	20	10	3	193
	<i>Nb cit.</i>	2098	1204	998	147	566	103	77	62	33	10	0	2202
Culture informationnelle	<i>Nb pub.</i>	115	6	30	1	4	4	1	6	8	6	0	36
	<i>Nb cit.</i>	64	9	47	1	7	14	1	14	9	1	0	56
Maîtrise de l'information	<i>Nb pub.</i>	835	108	203	18	24	20	37	35	34	21	14	311
	<i>Nb cit.</i>	486	254	215	21	64	41	49	24	11	5	0	469
Information literacy	<i>Nb pub.</i>	(996)	(998)	7756	848	989	983	991	993	990	985	977	8754
	<i>Nb cit.</i>	13295	12046	22531	4536	4469	4157	3651	2802	1833	664	419	34577

Notes :

- *Nb pub.* : nombre de publications trouvées
- *Nb cit.* : nombre de citations trouvées par Harzing PoP
- *Total dat.* : total des résultats des publications datées, *i.e.* des documents publiés avant 2000 et après 2000 (selon les critères de Google Scholar) ;

Remarques méthodologiques :

- Harzing PoP travaille exclusivement à partir des résultats de Google Scholar et reflète donc les limites de cet outil (par exemple l'imprécision sur les ressources indexées), notamment sur un point précis, qui a légèrement perturbé les résultats : **Google Scholar ne permet pas de retrouver certains documents à partir de leur date de publication, notamment la plupart des documents disponibles sur les archives ouvertes**, qui apparaissent en général **non datés**. Ce point explique la dernière colonne de droite, qui récapitule le total des documents datés (somme des documents publiés avant 2000 et depuis 2000), ce chiffre pouvant être différent de celui de la première colonne de gauche, obtenu à partir d'une requête sans indication de dates.
- A ce premier problème s'ajoute un deuxième : **Google Scholar n'affiche et ne permet de traiter les résultats que dans une limite de 1000**, même s'il peut en annoncer beaucoup plus dans le chiffre total indiqué. Ainsi le total de résultats sur « *information literacy* » est de 996 sur Harzing PoP, correspondant bien aux 996 résultats affichés sur Google Scholar, alors que le nombre total est de 22 200 ! De ce fait, presque toutes les requêtes sur *information literacy* ont été biaisées par cette limitation et ne dépassent pas les 1000 résultats, alors que la somme des résultats par années montre clairement que le total (ici 8754) est largement supérieur au total des résultats trouvés sans indication de dates ;
- pour chaque requête, Harzing PoP donne, non seulement le **nombre de publications recensées** dans Google Scholar, mais aussi le **nombre de citations reçues par ces publications**, permettant ainsi de mesurer l'impact d'une publication particulière ou de l'ensemble trouvé. Ne sont rapportés

ici que le total statistique pour chaque requête, mais il faut préciser que l'outil donne la liste complète des résultats, avec pour chaque publication, son nombre de citations, son rang, etc. Plusieurs autres indicateurs scientométriques sont fournis, que nous n'avons pas exploités ici.

- Rappel : les chiffres indiqués en italiques (colonne *Total dat.*, et résultats de *information literacy > 2000*) proviennent, non pas directement de Harzing PoP, mais de l'addition des résultats correspondant ; les **chiffres indiqués entre parenthèses traduisent la limitation de Google Scholar à 1000 résultats** et peuvent donc être parfois très inférieurs au total réel de publications.

4.2 Synthèse des résultats :

	<i>Total publis.</i>	<i>Publis. « datées »</i>	<i>% de publis « non datées »</i>	<i>% de publis > 2000</i>	<i>Total citations</i>	<i>Cit. / par publi.</i>
<i>Culture de l'information</i>	408	193	53 %	69 %	2098	5,14
<i>Culture informationnelle</i>	115	36	69 %	83 %	64	0,56
<i>Maîtrise de l'information</i>	835	311	63 %	65 %	486	0,58
<i>Information literacy</i>	996	--	--	--	13295	13,35

Remarques méthodologiques :

- *Total publis.* : nombre total de publications trouvées en requête simple ;
- *Publis. datées* : nombre total de publications retrouvées par la « Date de publication » (sur le mode avancé de Google Scholar) ;
- *% de publis « non datées »* : pourcentage de publications non retrouvables, sur Google Scholar, par la date de publication ;
- *% de publis > 2000* : pourcentage de publications (par rapport au total des publications « datées »), parues depuis 2000
- *Total citations* : nombre total de citations (trouvées par Harzing PoP), reçues par l'ensemble des publications trouvées sur une requête ;
- *Cit. par publi.* : nombre moyen de citations reçues par chaque publication.

Observations des résultats :

• Sur le nombre de publications :

- on relève une fois de plus la **domination écrasante de « information literacy »** sur les autres expressions : **au minimum 8754 résultats** (pour les autres expressions, on ne constate pas de différence entre la recherche indirecte par Harzing PoP ou la recherche directe par Google Scholar) ;
- parmi les trois expressions françaises, c'est le syntagme **« maîtrise de l'information »** qui **l'emporte nettement** sur les deux autres : 835 publications au total, soit deux fois plus que « culture de l'information » (408) et presque huit fois plus que « culture informationnelle » (115) ;
- il faut relever également la **différence importante entre « culture de l'information » (408) et « culture informationnelle » (115)**, beaucoup moins utilisée ;
- enfin il faut relever la **différence, également assez importante, entre le total des publications « datées » et celui des « non datées »** : ainsi, les pourcentages des publications « datées » (et prises en compte à ce titre par Google Scholar) sont respectivement de 31 %, pour « culture informationnelle », 47% pour « culture de l'information et 37% pour « maîtrise de

l'information »⁵. Comment interpréter ce différentiel ? Plusieurs facteurs sont à prendre en compte, car l'observation des résultats montre que les publications, considérées par Google Scholar sans date de publication, sont en réalité assez diverses : beaucoup de *pré-prints* (déposés dans des archives ouvertes, publiés sur des sites web...), textes de communications non encore publiés, mais aussi certains articles de revues dont la date ne semble pas prise en compte⁶, thèses ou mémoires, divers documents, etc. Cette étude bibliométrique met incidemment le doigt, sans le vouloir, sur l'un des points faibles de Google Scholar : sa difficulté à prendre en compte les dates de publication des documents scientifiques, dès qu'ils sortent d'un certain cadre, par ailleurs non explicité...

- au-delà des aspects parfois incohérents ou inexplicables de Google Scholar, ce différentiel entre publications « datées et non datées » atteste ce fait, intéressant à relever : **la part dominante** (plus de la moitié, voire les deux-tiers pour « culture informationnelle ») **de publications encore mal reconnues**, souvent professionnelles, et non « légitimées », *i.e.* hors du cadre institutionnel des revues scientifiques ; ce constat est particulièrement frappant **pour les résultats sur « culture informationnelle »**, ce qui est une autre manière, *a contrario*, de souligner le caractère nouveau, émergent, de cette thématique, par ailleurs largement « franco-française ».

- **Sur le nombre de citations :**

Certaines observations identiques peuvent être faites à propos des citations :

- notamment la **domination absolue du nombre de citations de « information literacy »** : au moins **34577 citations** sur le total des 8754 publications « datées » ; si l'on prend le chiffre trouvé par Harzing PoP, 13295 citations pour 996 publications (sans précisions de dates, dans la limite des 1000 résultats), l'outil bibliométrique fournit le rapport de 13,35 citations par publication, alors qu'il s'établit à 5,14 pour « culture de l'information » et tombe à 0,58 et 0,56 pour « maîtrise de l'information » et « culture informationnelle » ;
- en revanche, il faut relever une différence importante avec l'analyse précédente du nombre de publications : si les **publications de « culture de l'information » sont deux fois moins nombreuses que celles de « maîtrise de l'information »**, elles ont néanmoins **un impact bibliométrique très nettement supérieur**, avec un **total de citations quatre fois plus élevé** (2098 contre 486) ; ce volume important de citations s'explique, selon nous, par trois facteurs :
 - d'abord, la proportion moins importante de publications « non datées » (55 % contre 63 % pour la « maîtrise de l'information »), ce qui signifie que les articles de revues sont proportionnellement plus nombreux dans le lot de résultats « culture de l'information », et recueillent donc davantage de citations ;
 - ensuite, la part assez importante de publications anglo-saxonnes, dont plusieurs articles capitalisent un grand nombre de citations ;
 - enfin la présence de nombreux ouvrages, dont certains assez célèbres et abondamment cités⁷ ;
- les résultats des citations sur les publications « maîtrise de l'information » sont plus difficiles à analyser : ces publications reçoivent un nombre de citations très faible (0,58 citations par publication), alors qu'elles sont beaucoup plus nombreuses que les deux autres. L'examen des résultats sur Harzing nous fournit des éléments d'explication : publications plus hétérogènes que les autres dans leurs thématiques, moins d'ouvrages que « culture de l'information », beaucoup

5 Le calcul ne peut être fait, à partir des résultats de Harzing PoP, pour « information literacy », compte-tenu de la limitation à 1000 du lot de résultats.

6 C'est le cas par exemple de plusieurs articles du BBF (Bulletin des Bibliothèques de France), dont l'année de publication n'apparaît curieusement pas dans Google Scholar !

7 Comme le livre de Christine Bruce, « *The seven faces of information literacy* », cité 247 fois !, ou le livre de Pierre Lévy, « *Cyberculture* », qui recueille 181 citations ;

moins de publications anglo-saxonnes, beaucoup de publications « non datées » (63 %), même s'il n'y a pas de corrélation systématique entre le nombre de citations et l'absence de date de publication⁸, beaucoup de documents d'archives ouvertes...

- enfin, il faut relever **la très grande faiblesse des publications sur la « culture informationnelle »** : étant peu nombreuses, mal référencées, peu publiées, elles sont également **très peu citées** (0,56 citations par publications).

- **Sur l'évolution des publications dans le temps :**

Si l'on observe l'évolution du nombre de publications « datées »⁹ et du nombre de citations, deux éléments apparaissent nettement :

- la **progression importante des publications à partir de 2000** : pour tous les lots de résultats, la différence est très nette entre les résultats avant et après 2000 : 83 % des résultats pour « culture informationnelle » publiés depuis 2000, 69 % pour « culture de l'information »... Cette progression spectaculaire est sans aucun doute le principal signe du caractère émergent de ces thématiques, dans le paysage de la littérature spécialisée.
- Concernant les années proprement dites, il est difficile de tirer des enseignements significatifs de ces résultats, hormis le constat d'une certaine progression à partir de 2003.

- **Sur l'origine ou la langue des publications :**

L'observation directe des résultats sur Harzing (mais on peut également le vérifier sur Google Scholar) montre trois faits incontestables :

- d'abord la **suprématie des publications en anglais ou d'origine anglo-saxonne** apparaît encore plus absolue ici, notamment à l'examen des résultats sur *information literacy* ; ce n'est certes pas une surprise, mais les proportions (que nous n'avons pu mesurer statistiquement) sont impressionnantes : les deux tiers, au moins, des résultats sur *information literacy* (sur Harzing) sont issues de publications américaines, anglaises, australiennes...
- une **proportion très importante** (près d'un tiers) de résultats issus de **publications asiatiques**, chinoises ou japonaises ;
- enfin **la part quasi-insignifiante des publications de langue française**, sur *information literacy* : ainsi sur les 977 résultats recensés sur Harzing pour l'année 2007, aucune publication française !

Cette dernière observation sur la part réduite des publications scientifiques écrites en français, corroborée par les trois autres études, n'est certes pas une nouveauté, et elle ne fait que confirmer une tendance lourde, malheureusement visible dans d'autres domaines.

Dans le champ de la culture informationnelle, cette faiblesse bibliométrique de la littérature spécialisée française montre le chemin qui nous reste à parcourir, pour arriver à une meilleure reconnaissance de cette thématique émergente...

8 Ainsi une publication peut être plusieurs fois citée, même si elle est considérée comme « non datée » sur Google Scholar.

9 Toujours selon les critères de Google Scholar, bien entendu.