

***Statistiques de consultations internet de documents scientifiques :
l'impact des internautes du Maroc***

Pier Luigi ROSSI

Résumé

L'analyse des statistiques de consultations internet de documents scientifiques mis en ligne par l'IRD et ses partenaires montre qu'une part significative des consultations de ces documents (disponibles en texte intégral au format pdf) viennent du Maroc. Le "lectorat" marocain constitue une part importante du "lectorat" de l'Internet avec des fréquences comparables à celles du « lectorat » français. Ce lectorat semble très intéressé par les documents scientifiques disponibles en libre accès sur l'Internet, notamment ceux en langue française.

Les problématiques évoquant « une fracture numérique » au Maroc semblent marginales par rapport aux réels besoins de disponibilité d'information scientifique, notamment celle produite par des institutions marocaines.

A l'époque de l'internet, chaque Institution, chaque Université se doit de rendre accessibles ses propres productions. Ceci garantit visibilité, impact, notoriété ainsi qu'une plus grande légitimité, nationale et internationale, de la recherche scientifique et des Institutions qui la produisent.

ملخص: إحصائيات حول تصفح الوثائق العلمية على شبكة الإنترنت: تأثير مستخدمي الإنترنت في المغرب

أظهر تحليل إحصائيات حول تصفح الوثائق العلمية على شبكة الإنترنت بثها معهد البحث للتنمية (IRD) وشركائه على صفحات الشبكة أن المغرب يمثل نسبة مهمة فيما يتعلق بتصفح هذه الوثائق (متوفرة في نص كامل، حجم PDF). فنسبة القراء المغاربة لهذه الوثائق على شبكة الإنترنت تبقى مهمة بين بقية القراء وهي تضاهي نسبة تردد القراء الفرنسيين. يبدو من خلال هذه الإحصائيات أن القراء المغاربة يهتمون بشكل كبير بالوثائق العلمية المتاحة على الشبكة ولاسيما المكتوبة منها باللغة الفرنسية.

وهكذا فإن إشكالية "الفجوة الرقمية" في المغرب تبدو مسألة هامشية مقارنة مع الاحتياجات الحقيقية للمعلومات العلمية المتاحة بما في ذلك تلك التي تنتجها على الخصوص المؤسسات المغربية.

في زمن الإنترنت يجب على كل مؤسسة وجامعة أن تتيح الوصول إلى إنتاجها الشخصي عبر الإنترنت لأن هذا يضمن لها نوعاً من الوضوح في الرؤية والشفافية، والتأثير والخصوصية فضلاً عن اكتساب البحث العلمي والمؤسسات التي تنتجها مصداقية كبيرة على الساحة الوطنية والدولية.

Abstract : Statistics internet consultations scientific document: the impact of the Internet in Morocco

Statistical analysis of internet access to scientific documents posted by the IRD and its partners shows that a significant proportion of consultations of these documents (available in full text in pdf format) are from Morocco. The Moroccan "audience" is an important part of the "audience" of the Internet with similar frequencies to those of French "readership". The readers seem to be very interested in the available scientific literature freely available on the Internet, including those in French.

Issues evoking a "digital gap" in Morocco seem marginal compared to the real needs of availabilities of scientific information, including that produced by Moroccan institutions.

At the time of the Internet, each institution, each university must provide access to its own productions. This ensures visibility, impact, fame and greater legitimacy, national and international, of scientific research and institutions producing it.

I- Introduction

Dans le cadre des partenariats que l'IRD entretient avec plusieurs institutions présentes dans les pays d'Afrique francophone, nous avons mis en service un portail internet qui héberge des collections de documents scientifiques produits par ces partenaires.

Le serveur BEEP (Bibliothèques électroniques en partenariat), accessible à l'adresse www.beep.ird.fr, essaye ainsi d'apporter une solution concrète aux difficultés conjoncturelles qui limitaient l'opérabilité des serveurs des institutions localisées dans les pays en développement : connexions de mauvaise qualité, indisponibilité des serveurs, coupures électriques fréquentes, mise en œuvre complexe des projets informatiques¹.

Actuellement le serveur BEEP héberge onze collections² : trois d'institutions du Burkina Faso, deux relevant d'une Institution de Maurice, cinq relevant d'institutions du Sénégal et une constituée d'articles concernant les problématiques de l'érosion des sols dans les pays en développement.

BEEP fournit à tous ses partenaires des services à valeur ajoutée. L'élaboration d'outils efficaces pour atteindre la meilleure indexation par les principaux moteurs de recherche et le déploiement de passerelles OAI³ assurent une grande visibilité aux contenus qui sont hébergés⁴. Les analyses statistiques des consultations portant sur les contenus des collections (les fichiers pdf) permettent d'élaborer des indicateurs stratégiques pointus sur la nature et l'origine des recherches⁵.

¹ Rossi Pier Luigi. Electronic libraries in partnership: BEEP for Africa. European Conference on African Studies (ECAS4), Uppsala, 15-18/06/2011, 6 p. <http://www.nai.uu.se/ecas-4/panels/141-156/panel-145/Pier-Luigi-Rossi-full-paper.pdf>.

² Toutes les collections sont constituées de documents en texte intégral au format pdf. Seule exception une des deux collections de Maurice.

³ Open Archives Initiative : <http://www.openarchives.org/>.

⁴ Voir Rossi P.L., 2011.

⁵ Rossi Pier Luigi, Thiaw Anastasie. .Log analysis and text mining on internet access to dissertations of the Dakar sports faculty (INSEPS). SCOLMA - 50th anniversary conference, Oxford, 25-26 June 2012.

Nous avons utilisé les statistiques des consultations des collections hébergées par BEEP pour connaître et analyser l'impact des internautes du Maroc par rapport aux documents disponibles sur le serveur.

II- Analyse des consultations

L'analyse des consultations des documents hébergés par BEEP est basée sur l'exploitation des fichiers de log du serveur Apache⁶ du site. Les extractions sont effectuées exclusivement sur les fichiers pdf dont l'accès comporte le code 200 (requête traitée avec succès)⁷. Un traitement de dédoublonnage permet d'éliminer les lignes générées par des phénomènes de spam⁸.

Ces données sont ensuite segmentées par rapport aux accès des robots (notamment ceux du robot de Google) et par rapport à ceux des utilisateurs. Ceci concerne tous les fichiers pdf du serveur. Les mêmes principes de segmentation sont appliqués aux fichiers de log de chaque collection.

La résolution des adresses IP de chaque requête des utilisateurs permet de connaître la répartition géographique des consultations pour l'ensemble du site ainsi que pour chaque collection. Pour réaliser ces traitements un script PHP qui intègre « MaxMind GeoIP Country Database »⁹ a été développé.

Nous avons analysé les statistiques de consultation des collections qui se composent du plus grand nombre de documents ayant une couverture thématique pouvant concerner de larges publics¹⁰.

⁶ <http://httpd.apache.org/>

⁷ http://en.wikipedia.org/wiki/HTTP_code

⁸ Les accès sur les documents pdf ont été calculés en éliminant des fichiers de log les accès récursifs sur les fichiers pdf. Deux accès d'une même adresse ip sur le même fichier sont comptabilisés si l'accès se réalise avec une différence de temps supérieure à 5 minutes.

⁹ Voir : <http://www.maxmind.com/app/country>.

¹⁰ Cinq collections ont été prises en considération par rapport aux statistiques de consultation : BRE (Bulletin du Réseau érosion), THIES (Mémoires de l'Ecole polytechniques de Thiès), PTCI (Mémoires de DEA du Programme de troisième cycle interuniversitaire d'Economie de Dakar), INSEPS (Mémoires de l'Institut National Supérieur de l'Education Populaire et du Sport de Dakar), EISMV (Thèses et mémoires de l'Ecole Inter-états des Sciences et Médecine Vétérinaires, Dakar).

Le tableau 1 montre les pourcentages de consultation des collections par rapport à la localisation géographique des accès. Nous avons pris en considération pour ce tableau les consultations ayant comme origine la France, les trois pays du Maghreb (Algérie, Maroc, Tunisie) et regroupé les accès des pays restants (« Autres »). Pour avoir un terme de comparaison par rapport à de « possibles publics », dans le tableau figure également le nombre des effectifs des étudiants pour l'année scolaire 2011/12 des quatre pays identifiés. Les pourcentages sont calculés par rapport aux consultations du mois de décembre 2012.

Tableau 1 : Statistiques de consultations (%) de cinq collections du serveur BEEP

	BRE	THIES	PTCI	INSEP S	EISM V	Etudiants U. 2011/12
Algeria	23.0	19.9	5.8	10.8	21.4	1.247.000
France	19.4	26.7	18.2	43.0	24.4	2.347.532
Morocco	17.0	13.5	8.0	6.2	4.6	430.000
Tunisia	6.0	9.8	11.1	5.4	7.2	339.619
Autres	34.6	30.1	56.9	34.7	42.3	

Les résultats montrent que pour toutes les collections le pourcentage des consultations ayant comme origine géographique le Maroc est important. Les pourcentages sont très significatifs pour les collections ayant des thématiques concernant l'érosion des sols (BRE) et les sciences de l'ingénieur (THIES).

La figure 1 montre la répartition de la localisation géographique des accès pour un document de la collection de l'Ecole polytechnique de Thiès¹¹. Pour ce document on constate que les accès localisés au Maroc constituent 15,2 % des accès pour l'année 2012.

¹¹ Ce document est le deuxième document le plus consulté de la collection THIES. Il a été choisi pour cette étude dans la mesure où il comporte le plus grand nombre de consultations, par rapport aux autres documents de la collection, ayant comme origine le Maroc. Il a été consulté 11.861 fois en 2012, soit environ 32 fois par jour. La moyenne journalière des consultations ayant comme origine le Maroc est de 5 par jour.

Par rapport aux 1791 accès comptabilisés en 2012 ayant comme origine le Maroc, nous avons pu extraire les questions des utilisateurs pour 94 % des accès, dont 86 % viennent de questions posées dans « Google ».

L'analyse linguistique des questions¹² (lemmatisation et extraction des termes significatifs : substantifs, adjectifs, verbes) associée à l'analyse de co-occurrences des termes significatifs permet de construire un graphe¹³ qui met en évidence trois clusters des termes associés :

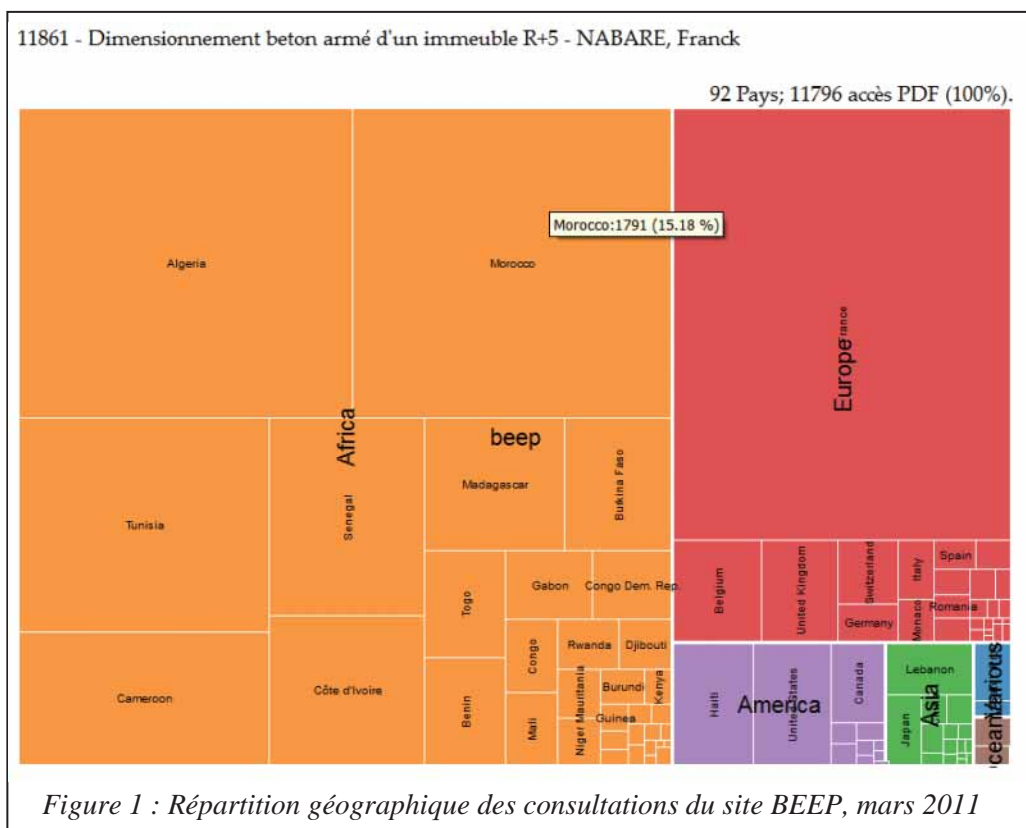


Figure 1 : Répartition géographique des consultations du site BEEP, mars 2011

¹² L'analyse linguistique des questions à été réalisée via un programme que nous avons réalisé qui intègre l'outil d'analyses linguistiques « TreeTagger » (<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>).

¹³ Ce graphe, réalisé avec « Gephi » (<https://gephi.org/>) et « Gexf-js » (<https://github.com/raphv/gexf-js>), est accessible en ligne à : <http://www.beep.ird.fr/graphe/index.html#beton.gexf>.

- un constituant les questions concernant le béton armé,
- un constituant les questions concernant le dimensionnement des bâtiments,
- un constituant les questions concernant la descente de charges.

Cette analyse des statistiques de consultations des collections de BEEP permet de montrer :

Une présence significative des publics marocains par rapport aux collections hébergées, notamment celles concernant les Sciences de l'Ingénieur.

L'impact des consultations ayant comme origine le Maroc par rapport à chaque document des collections : certaines publications sont consultées plusieurs fois par jour par des internautes du Maroc.

La cohérence des questions posées par rapport aux contenus des documents grâce à l'analyse des questions posées par les internautes.

III- Institutions au Maroc

Nous avons effectué une navigation sur les sites officiels de trois Universités du Maroc pour vérifier si ces institutions rendaient accessibles leurs propres publications scientifiques (ouvrages, articles de revues, productions scientifiques, ...).

Nous avons pris en considération les deux Universités de Rabat (Mohammed V-Agdal, Mohammed V - Souissi) ainsi que l'Université Hassan II de Casablanca.

Pour le site de l'Université d'Agdal¹⁴, il n'y a **aucune publication scientifique disponible**, ni aucun lien vers d'autres sites de l'Université. Par ailleurs, l'accès vers la bibliothèque¹⁵ est inopérant (code erreur de retour « 404 »).

¹⁴ <http://www.um5a.ac.ma/index.php/fr/2012-12-13-12-22-54/production-scientifique-et-valorisation-de-la-recherche/production-scientifique>. Page consultée le 26/03/2013.

¹⁵ http://www.um5a.ac.ma/index.php/fr/biblio.um5a.ac.ma/opac_css. Page consultée le 26/03/2013.

Pour le site de l'Université de Souissi¹⁶, sont indiqués **zéro ouvrage** et **zéro production scientifique**. Pour les revues, quatre titres sont indiqués mais **aucun numéro ou article** ne sont disponibles.

Pour l'Université Hassan II de Casablanca il existe une page concernant les archives ouvertes de cette Institution¹⁷. Lorsqu'on essaye de consulter les documents disponibles en suivant le lien qui est indiqué sur cette présentation, **aucune page n'est disponible** et on obtient le message « Le délai d'attente est dépassé ». Par ailleurs, sur le même site institutionnel, il existe une page concernant le répertoire des thèses soutenues à l'Université Hassan II de Casablanca¹⁸. Le contenu du fichier qui est donné en lien contient seulement le mot « TEST ».

Ce sondage réalisé sur les trois sites institutionnels des Universités de Rabat et de Casablanca montre qu'**aucun document** scientifique produit par ces Institutions **n'est disponible** via un accès Internet. De plus, sur ces sites, sont présents des liens vers des pages et/ou des présentations qui évoquent la possibilité de consulter des productions scientifiques relevant de ces Institutions sans que pour autant aucun document ne soit disponible¹⁹.

IV- Conclusions

En utilisant les statistiques de consultations des collections et des documents au format PDF hébergés sur le serveur BEEP (Bibliothèques électroniques en partenariat, www.beep.ird.fr) nous avons pu constater que les accès réels en provenance du Maroc sont très importants en nombre. Ceci est vérifié par rapport aux documents hébergés mais également par rapport aux nombres de consultations ayant comme origine les autres pays du Maghreb et la France ainsi que par rapport au nombre des « utilisateurs potentiels » présents dans ces pays (nombre d'étudiants universitaires).

¹⁶ http://www.um5s.ac.ma/index.php?option=com_docman&Itemid=165&lang=fr. Page consultée le 26/03/2013.

¹⁷ <http://www.biblio.uh2c.ac.ma/archives-ouvertes>. Page consultée le 26/03/2013.

¹⁸ http://www.uh2c.ac.ma/uh2c/theses_doctorat.php. Page consultée le 26/03/2013.

¹⁹ Le fait de créer des liens évoquant la possibilité de consulter des documents sans que pour autant ces liens soient opérationnels nous paraît nuisible en termes de communication et de crédibilité de sites institutionnels.

Nous supposons que le fait de mettre à la disposition des publics du Maroc des documents scientifiques en langue française favorise l'augmentation du nombre des consultations, en admettant que la langue du document est un des déterminants de la localisation géographique des consultations. Le graphique en « treemap²⁰ » de la figure 1 met en évidence ce phénomène.

Nous avons pu constater que trois sites institutionnels d'Universités du Maroc ne donnent accès à aucune production scientifique (articles, ouvrages, littérature grise,..) relevant de leurs propres « auteurs » (enseignants, chercheurs, étudiants), bien que des liens non opérationnels soient présents.

Il nous semble que ce manque constaté de réservoirs de productions scientifiques favorise la consultation de documents présents sur d'autres sites localisés hors du Maroc. Ce nomadisme de l'accès aux documents favorise, par ailleurs, la consultation d'une littérature scientifique de langue française.

A partir de ces constats, il nous semble que l'on ne peut pas faire référence à une « fracture numérique » au Maroc pour des publics d'internautes relevant du monde universitaire (étudiants, enseignants, chercheurs, professionnels). Nous croyons que l'on pourrait plutôt faire référence à un « creux numérique » de contenus accessibles via l'Internet en ce qui concerne les productions scientifiques universitaires du Maroc.

Il nous semble que ce « creux numérique » limite de façon significative la visibilité, l'impact, la légitimité des productions scientifiques des Universités et des Institutions scientifiques du Maroc. Par ailleurs, le fait de ne pas pouvoir disposer de statistiques de consultations des productions scientifiques qui existent physiquement dans plusieurs lieux de ces mêmes institutions (bibliothèques, administrations, bureaux des enseignants, ...) ne permet en aucun cas de produire des indicateurs de consultations ni d'élaborer des analyses stratégiques sur leur impact.

²⁰ Shneiderman Ben. Treemaps for space-constrained visualization of hierarchies. <http://www.cs.umd.edu/hcil/treemap-history/index.shtml>

Universite Mohammed V - Souissi
Presses de l'Institut Universitaire de la Recherche Scientifique

L'UNIVERSITE A L'ERE NUMERIQUE

E-FORMATION

E-RECHERCHE

E-GOUVERNANCE

Actes du colloque international
28 Février - 1 Mars 2013

Université Mohammed V- Souissi
Presses de l'Institut Universitaires de la Recherche Scientifique

L'UNIVERSITE A L'ERE NUMERIQUE
E-FORMATION, E-RECHERCHE, E-GOUVERNANCE

Actes du colloque international du 28 février - 1^{er} mars 2013

2013

Presses de l'Institut Universitaires de la Recherche Scientifique

IURS, avenue Allal El Fassi, Madinat el-Irfane, Rabat, BP 6287, Rabat-Instituts,
Maroc- Site web : iurs.um5s.ac.ma

Comité scientifique : Amina Aouchar, Amina Aderdour, Abdelkabir Bahani,
Hicham Hanchane, Zakaria Rhani

Dépôt légal : 2013MO1961

ISBN : 9954-538-18-0

Mise en page et Impression : Kawtar Print – Rabat - Tél : 0537 73 36 35

E-mail : kawtarprint@gmail.com

*Les opinions exprimées dans les articles qui paraissent dans le présent ouvrage sont
libres et strictement personnelles à leurs auteurs.*