

Auteur : Fred Merceur
Réf : DISCOM&RI/BLP/FM/2013-007
Version : 1.0
Date : Mai 2013

Statistiques d'usage des documents accessibles en Libre Accès via Archimer

Période 2005-2012

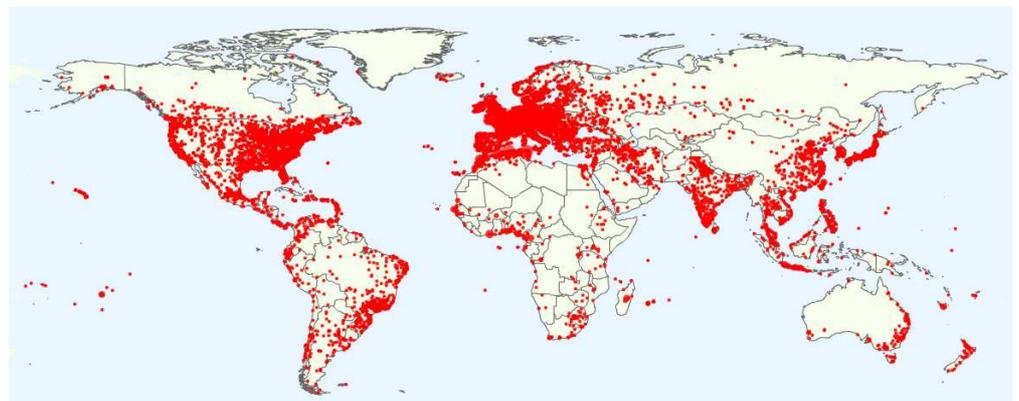


Table des matières

1. Introduction	2
2. Données statistiques.....	2
2.1. Dépôts.....	2
2.2. Téléchargements.....	4
2.3. Chemin d'accès	11
2.4. Internautes.....	14
3. Méthodologie.....	19

1. Introduction

Ce document propose un ensemble de données statistiques sur l'usage des documents déposés en Libre Accès dans Archimer, l'Archive Institutionnelle de l'Ifremer.

Seuls les téléchargements du texte intégral des documents ont été pris en compte. La visualisation des fiches de métadonnées n'a pas été comptabilisée.

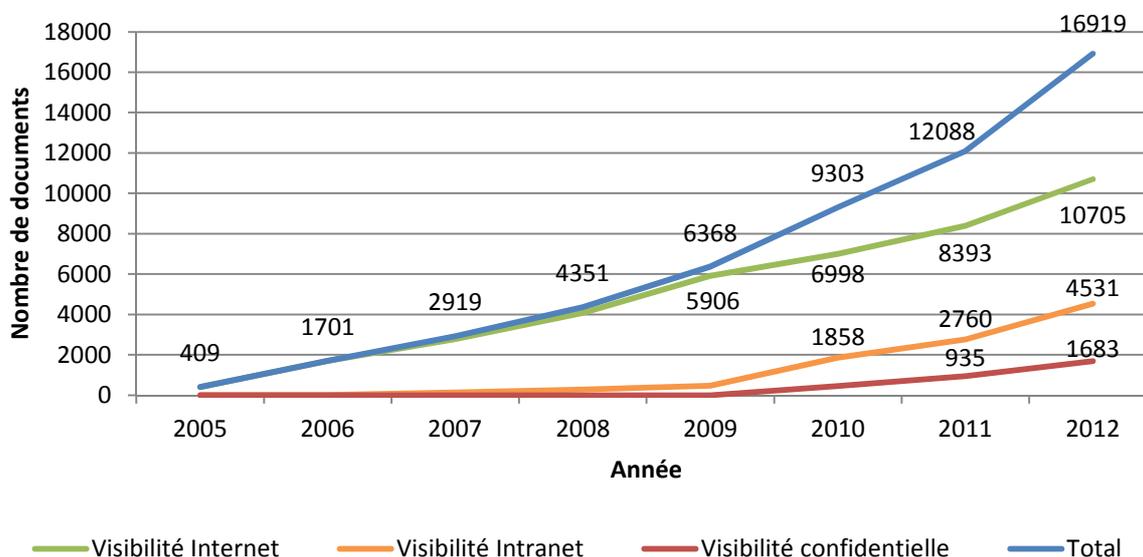
Pour éviter que l'activité d'administration de l'Archive (dépôt, contrôle, validation ...) ne fausse les résultats, les téléchargements réalisés depuis le réseau informatique de l'Ifremer n'ont pas été pris en compte.

Pour plusieurs données statistiques, des focus sur trois types de documents spécifiques sont proposés : les publications indexées dans le Web Of Sciences® (WOS), les thèses et les rapports¹. Les publications indexées dans le WOS² et les thèses sont particulièrement intéressantes du fait de leur homogénéité en termes de taille et de langue notamment.

2. Données statistiques

2.1. Dépôts

Figure 1 : Progression du nombre de documents³ disponibles dans Archimer



¹ Rapport de contrats, rapport scientifique, compte-rendu (d'essais, de campagnes, de métrologie...), note (de synthèses, de veille prospective ou technologique...) ...

² Nous avons différencié les publications indexées dans le WOS du reste des articles disponibles dans Archimer car elles sont souvent accessibles gratuitement via Archimer mais aussi, sous réserve d'abonnement, via le site de leurs éditeurs (contrairement, par exemple, aux articles de plusieurs journaux anciennement publiés par l'Ifremer, qui sont exclusivement accessibles via Archimer)

³ Tout type de documents confondus : publication, acte de colloque, rapport, thèse, HDR, avis, expertise, ouvrage, chapitre d'ouvrage, poster.

Figure 2 : Histogramme du nombre de documents³ déposés dans Archimer par année de dépôt

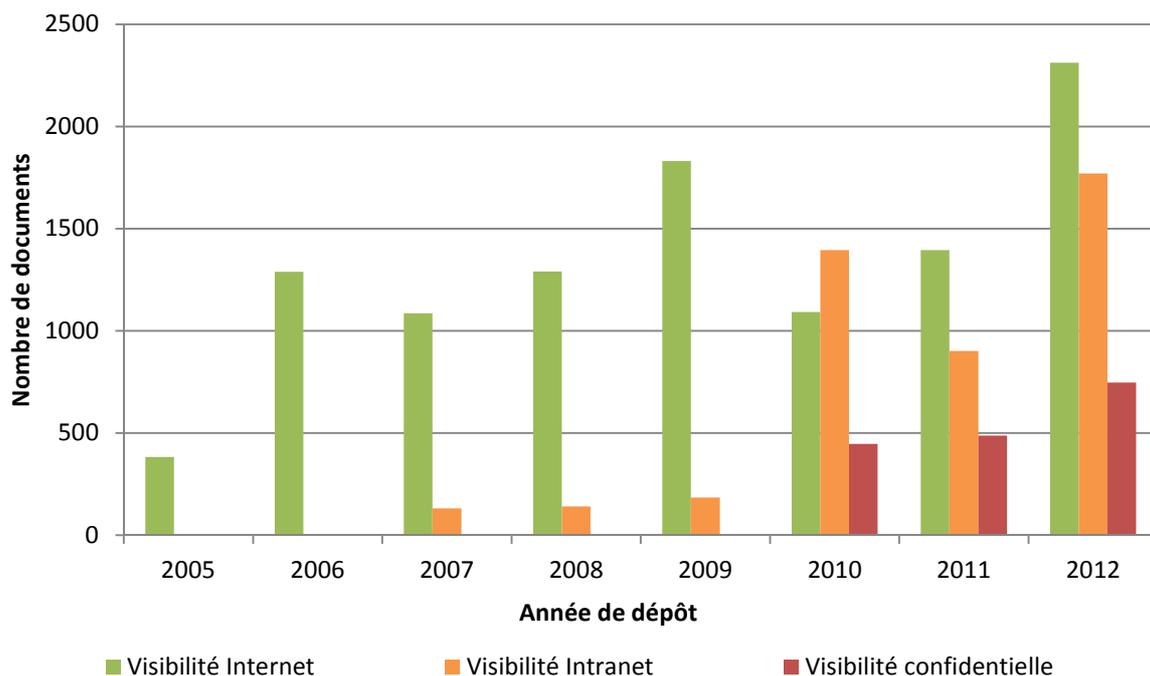
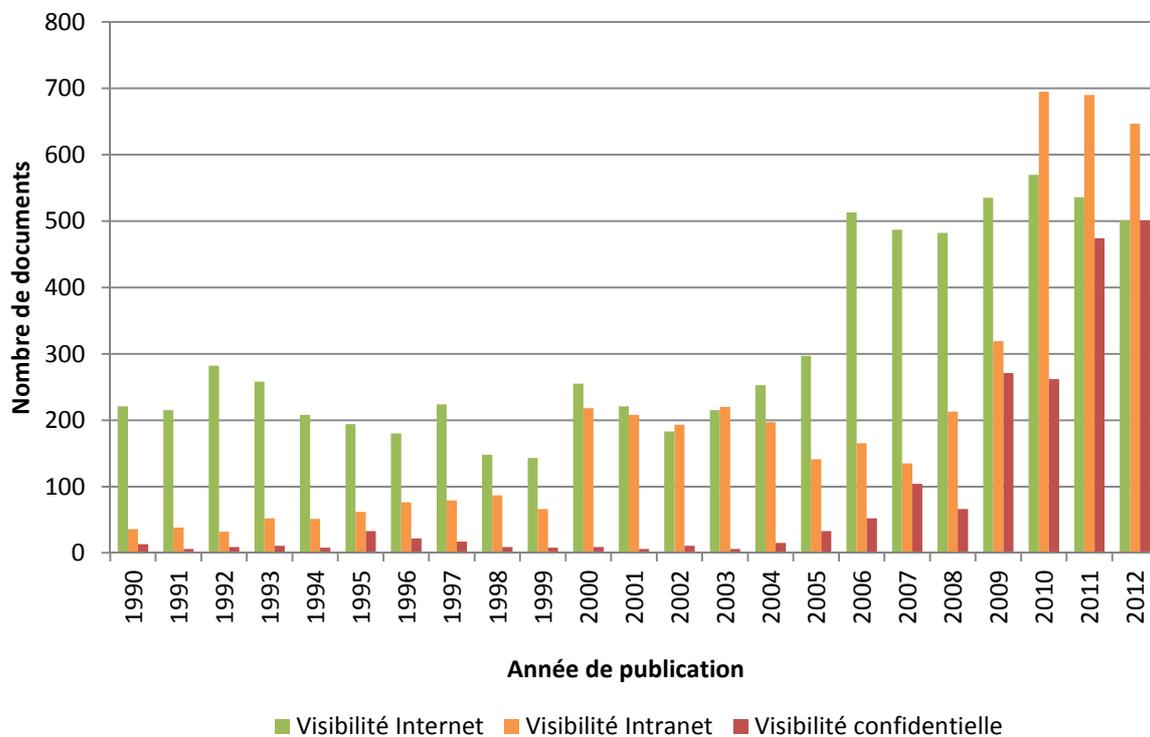


Figure 3 : Histogramme des années de publication des documents³ déposés dans Archimer



2.2. Téléchargements

Figure 4 : Progression de la somme annuelle des téléchargements des documents³

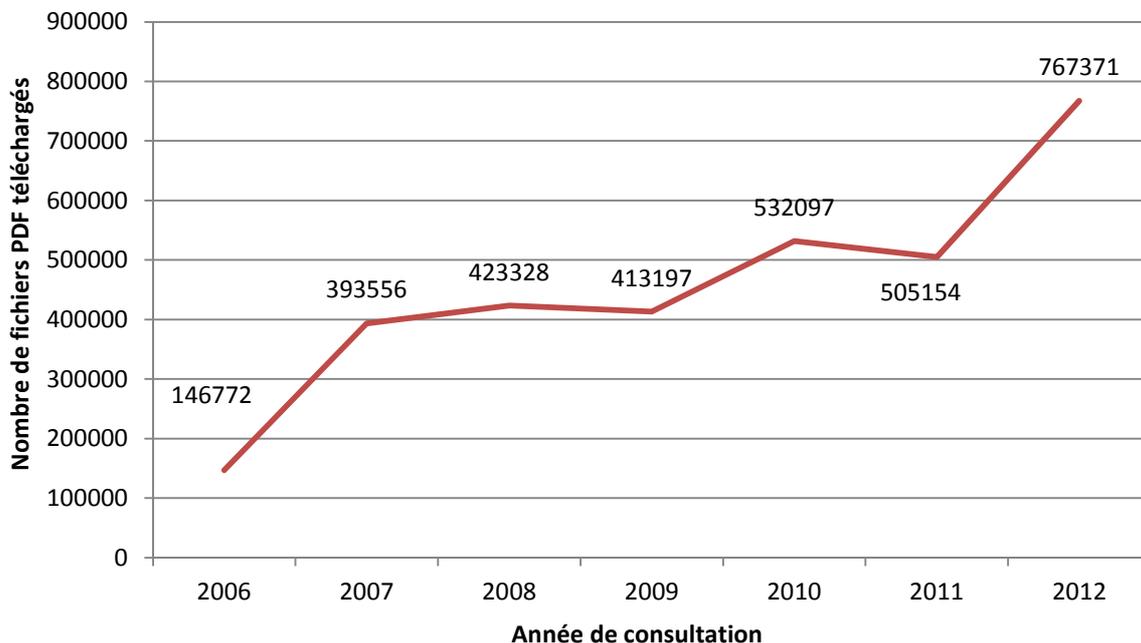


Figure 5 : Evolution de la somme mensuelle des téléchargements des documents³

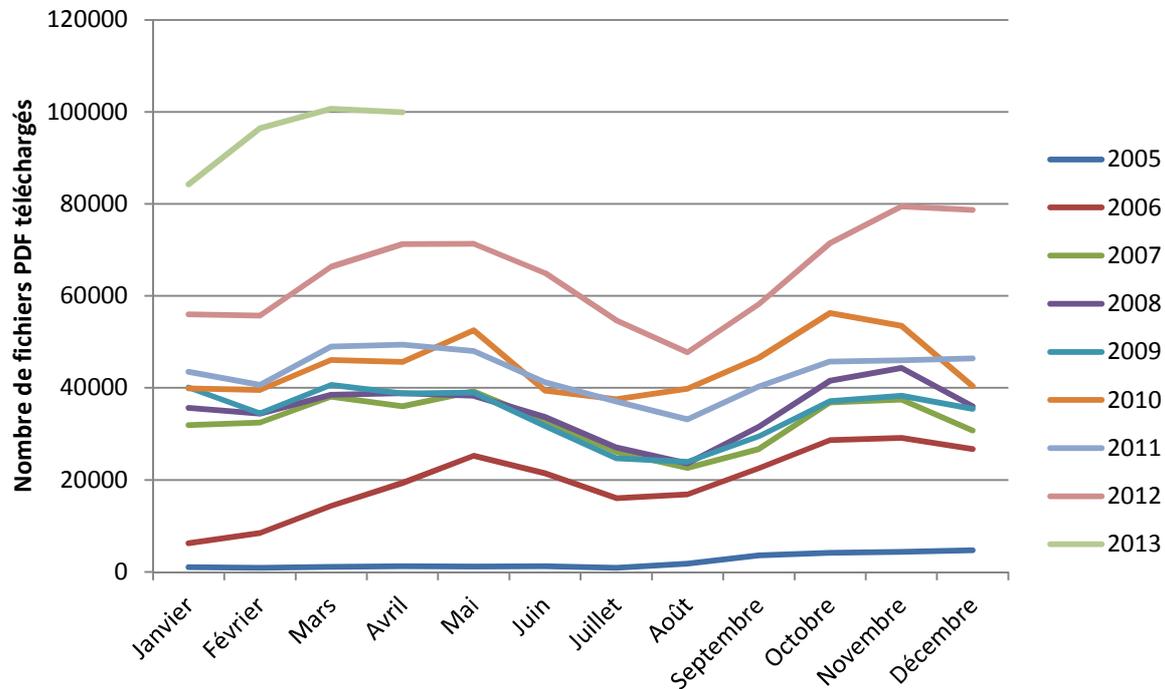
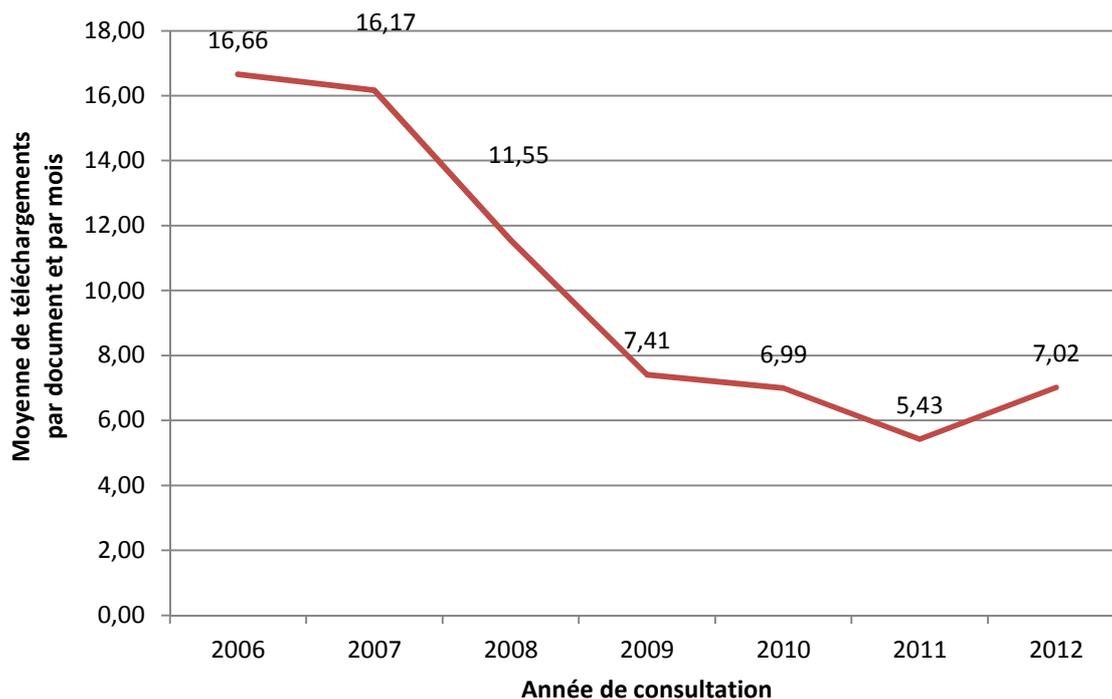


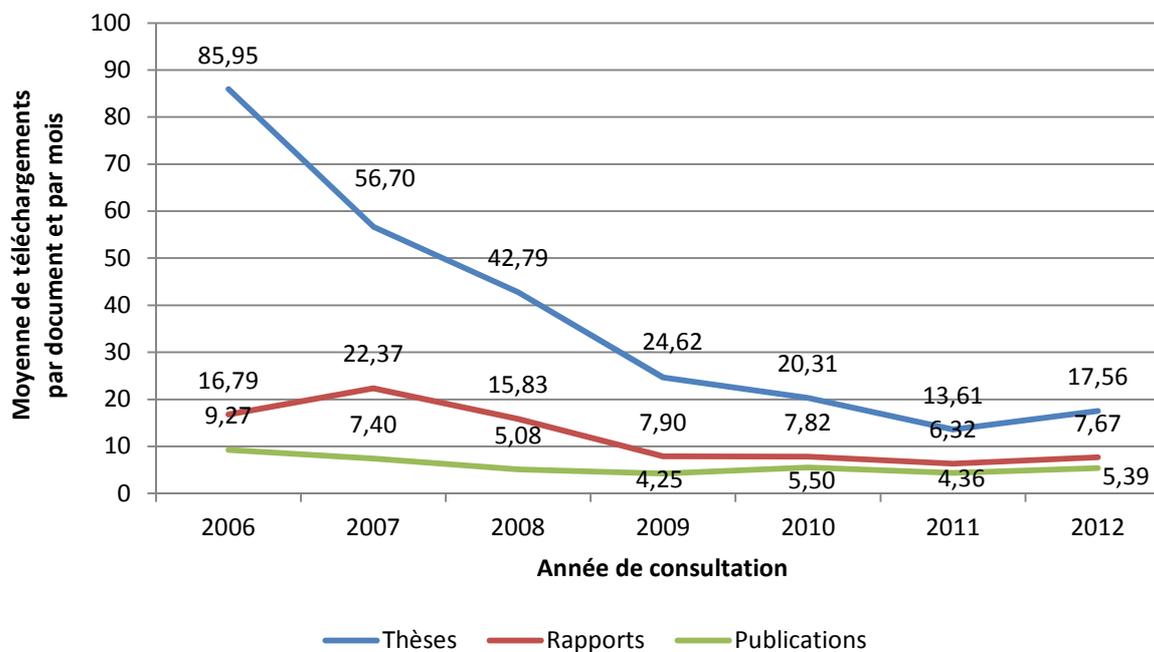
Figure 6 : Evolution de la moyenne mensuelle des téléchargements par document³



Remarque : les moyennes mensuelles de téléchargement lors d'une année n sont calculées ainsi :

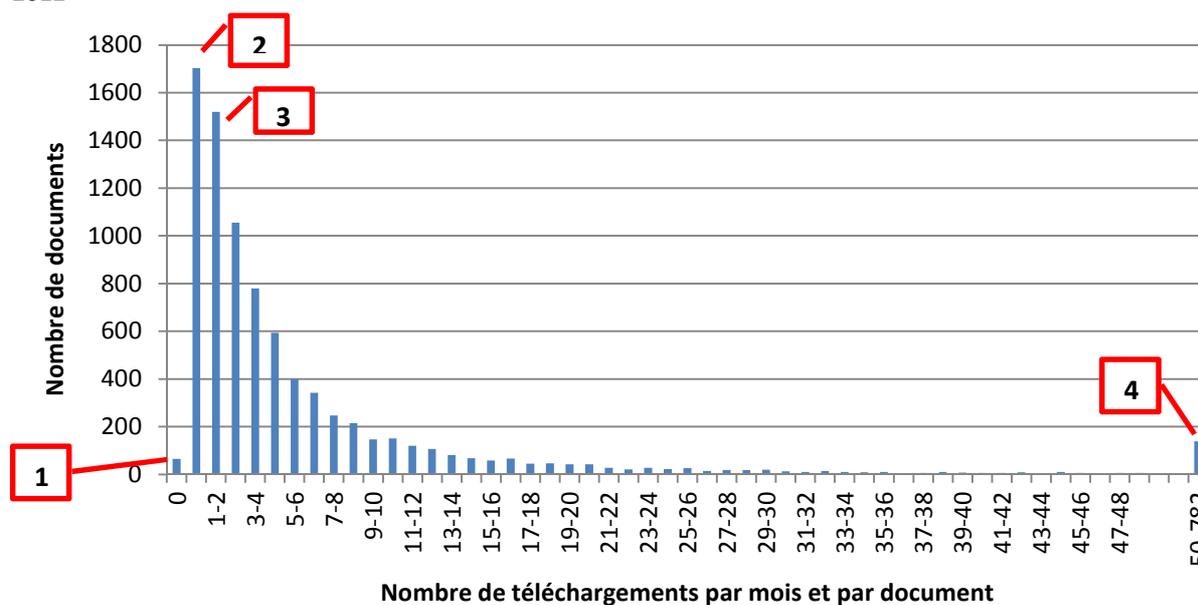
$$\frac{\text{Téléchargements enregistrés lors de l'année } n \text{ des documents déposés avant la fin de l'année } n - 1}{(\text{Nombre de documents déposés avant la fin de l'année } n - 1) \times 12}$$

Figure 7 : Evolution de la moyenne mensuelle des téléchargements des publications⁴, des thèses et des rapports



⁴ Publications indexées dans le WOS

Figure 8 : Histogramme des moyennes mensuelles de téléchargements des documents³ au cours de l'année 2012



La Figure 8 se lit de la façon suivante :

1. 65 des 8393 documents déposés en Libre Accès avant le 31 décembre 2011 dans Archimer n'ont jamais été téléchargés au cours de l'année 2012
2. 1703 de ces 8393 documents ont été téléchargés de 1 à 11 fois au cours de l'année 2012. La moyenne mensuelle de téléchargements de ces 1703 documents se situe donc entre 0 et 1.
3. 1520 de ces 8393 documents ont été téléchargés de 12 à 24 fois au cours de l'année 2012. La moyenne mensuelle de téléchargements de ces 1520 documents se situe donc entre 1 et 2.
4. 139 de ces 8393 documents ont été téléchargés en moyenne entre 50 et 782 fois par mois au cours de l'année 2012.

Figure 9 : Histogramme des moyennes mensuelles de téléchargements des publications⁴ au cours de l'année 2012

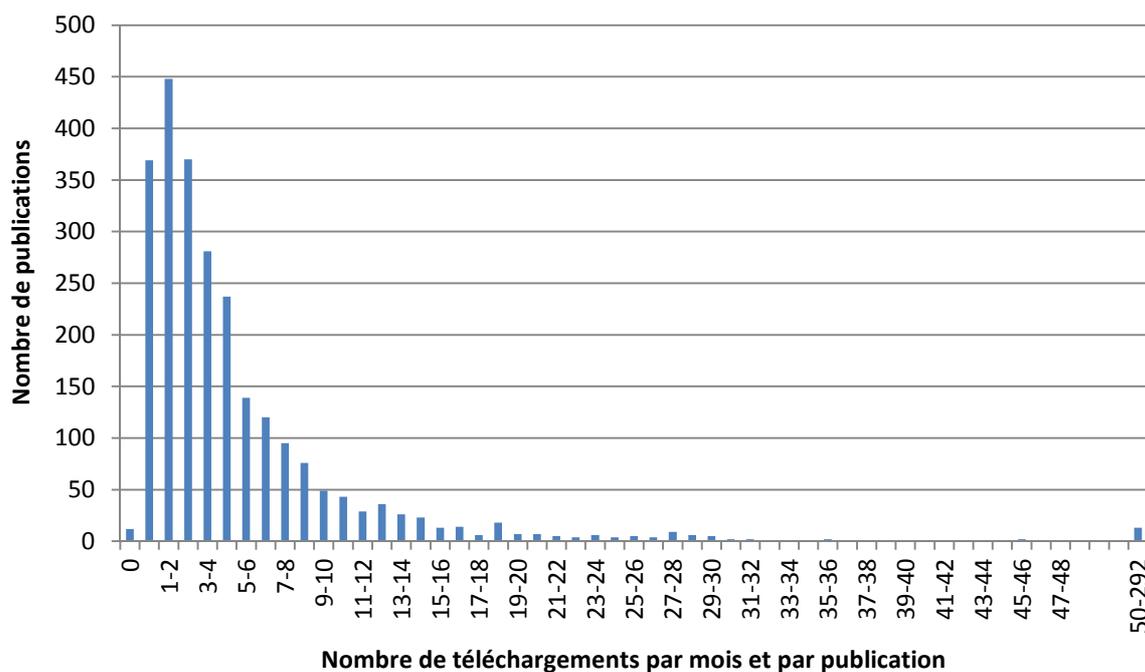


Figure 10 : Histogramme des moyennes mensuelles de téléchargements des thèses au cours de l'année 2012

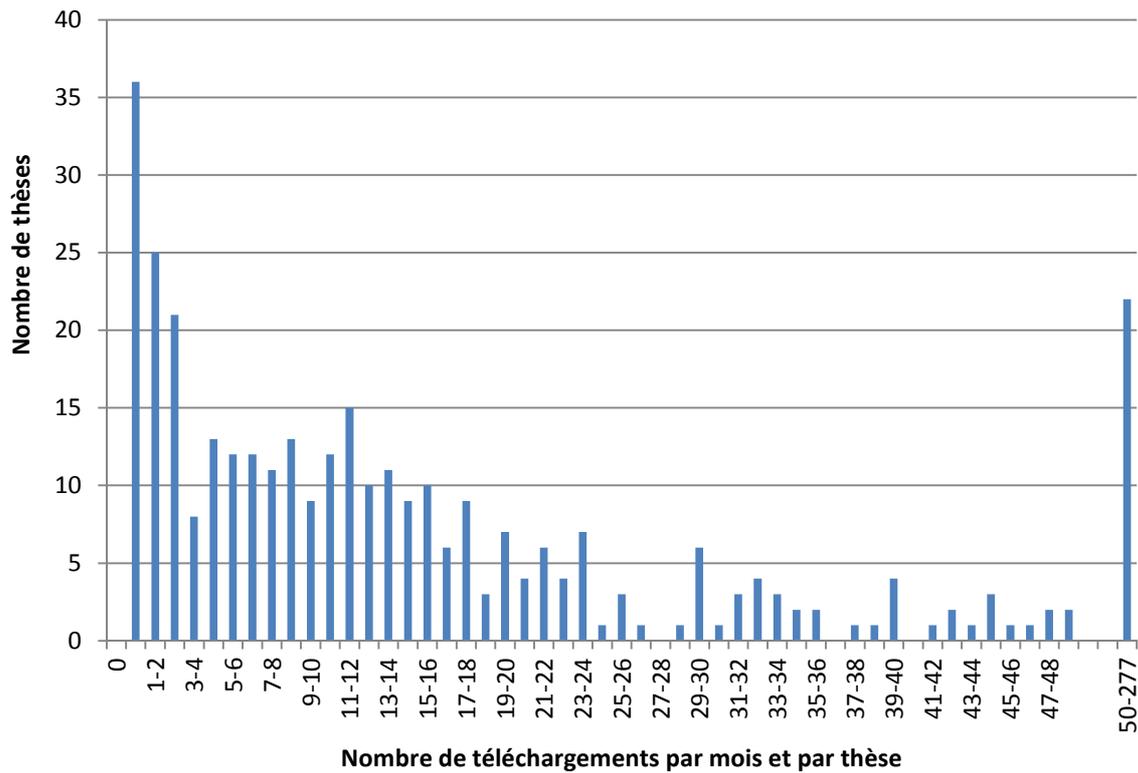


Figure 11 : Histogramme des moyennes mensuelles de téléchargements des rapports au cours de l'année 2012

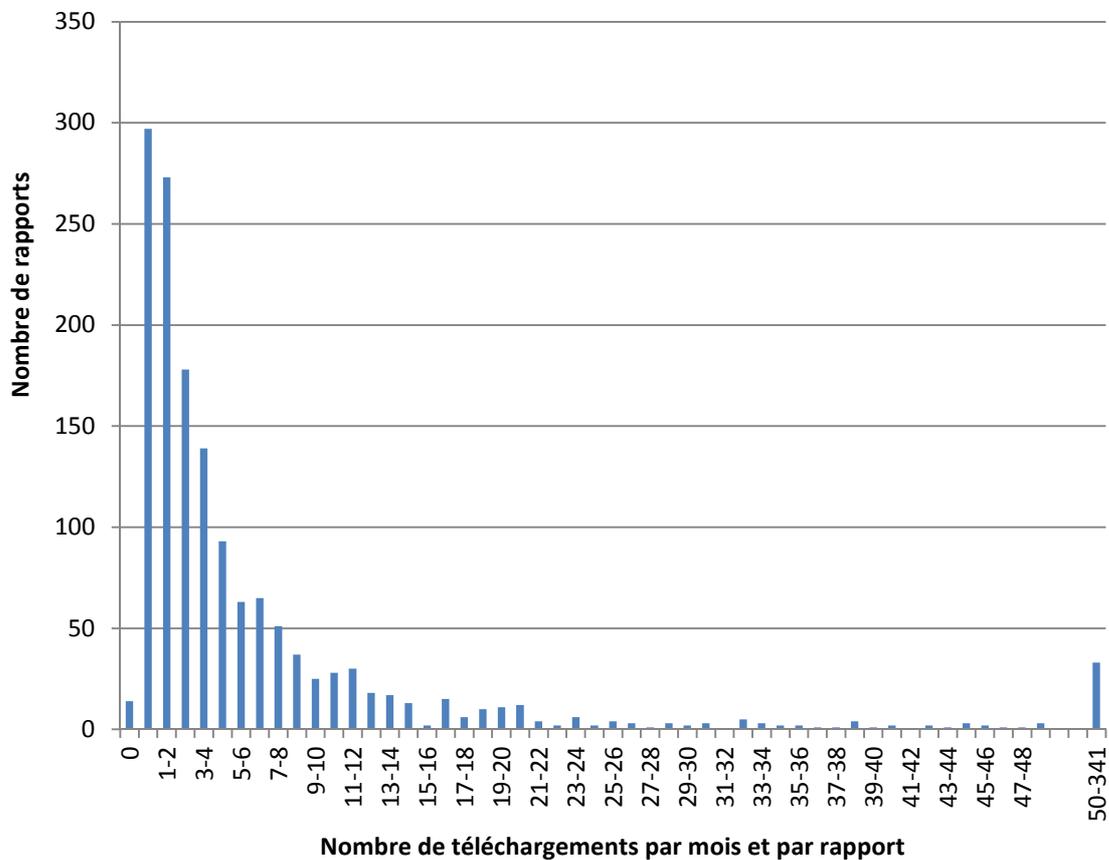


Figure 12 : Evolution de la moyenne mensuelle des téléchargements des thèses par année de dépôt

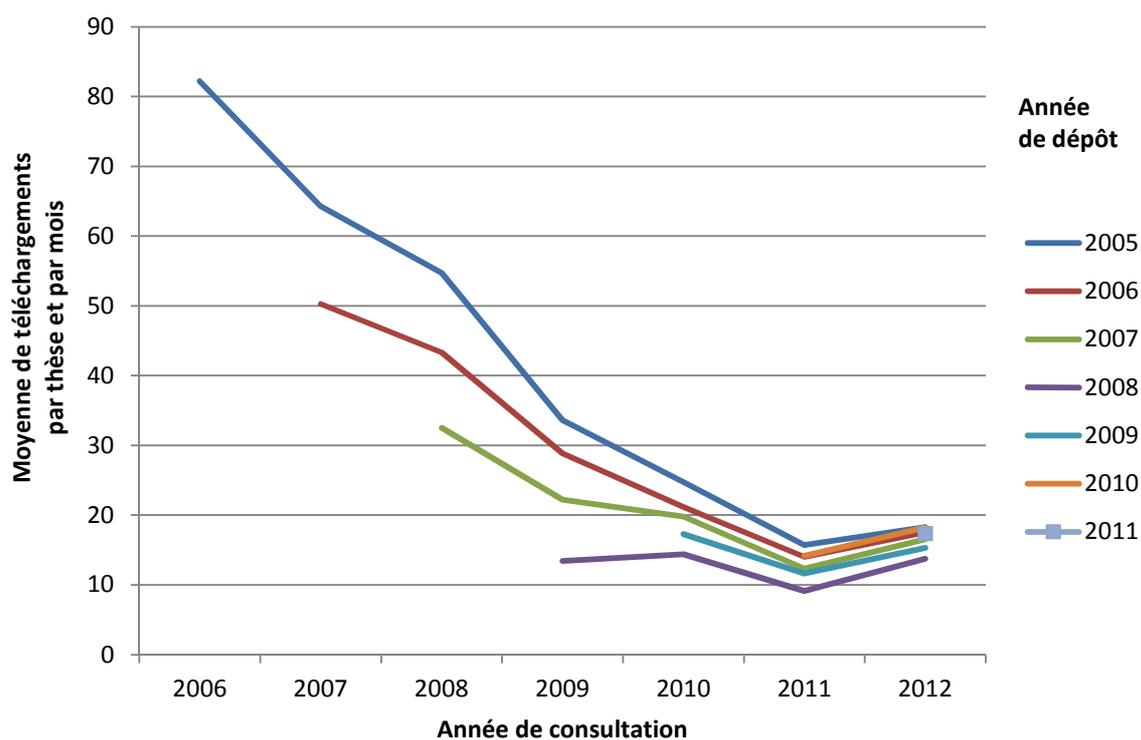


Figure 13 : Evolution de la moyenne mensuelle des téléchargements des publications⁴ par année de dépôt

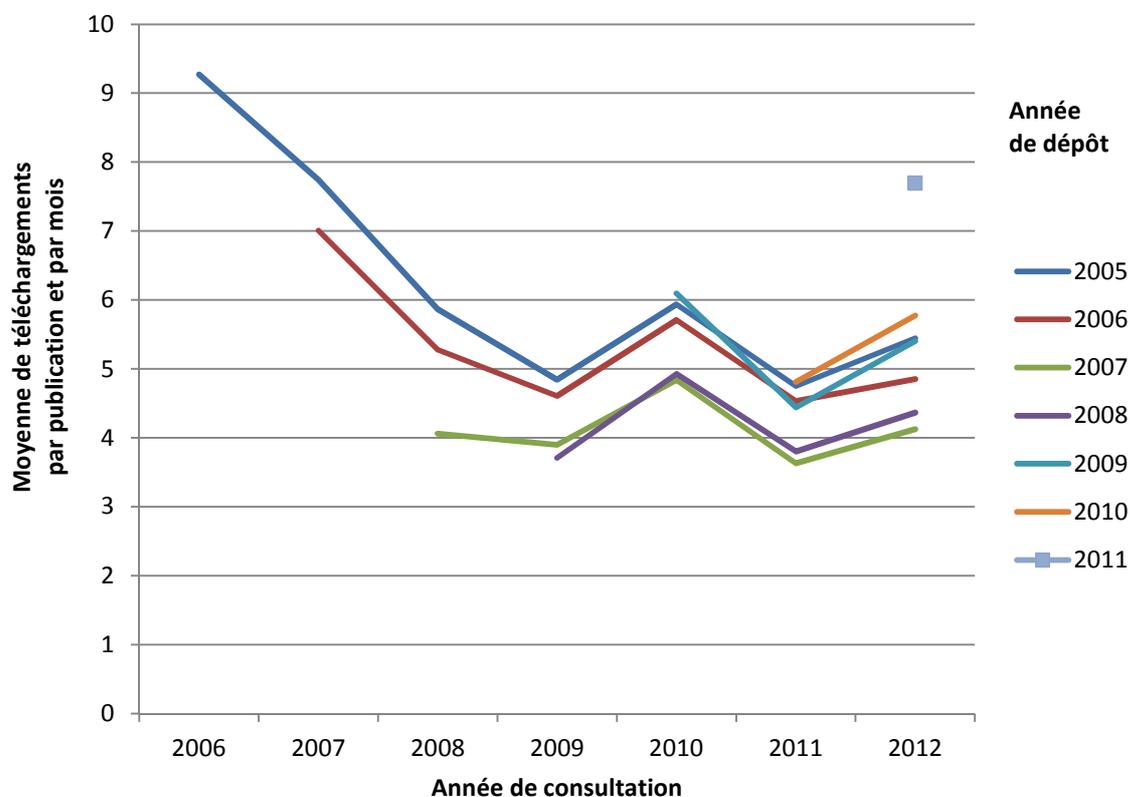


Figure 14 : Evolution de la moyenne mensuelle des téléchargements des rapports par année de dépôt

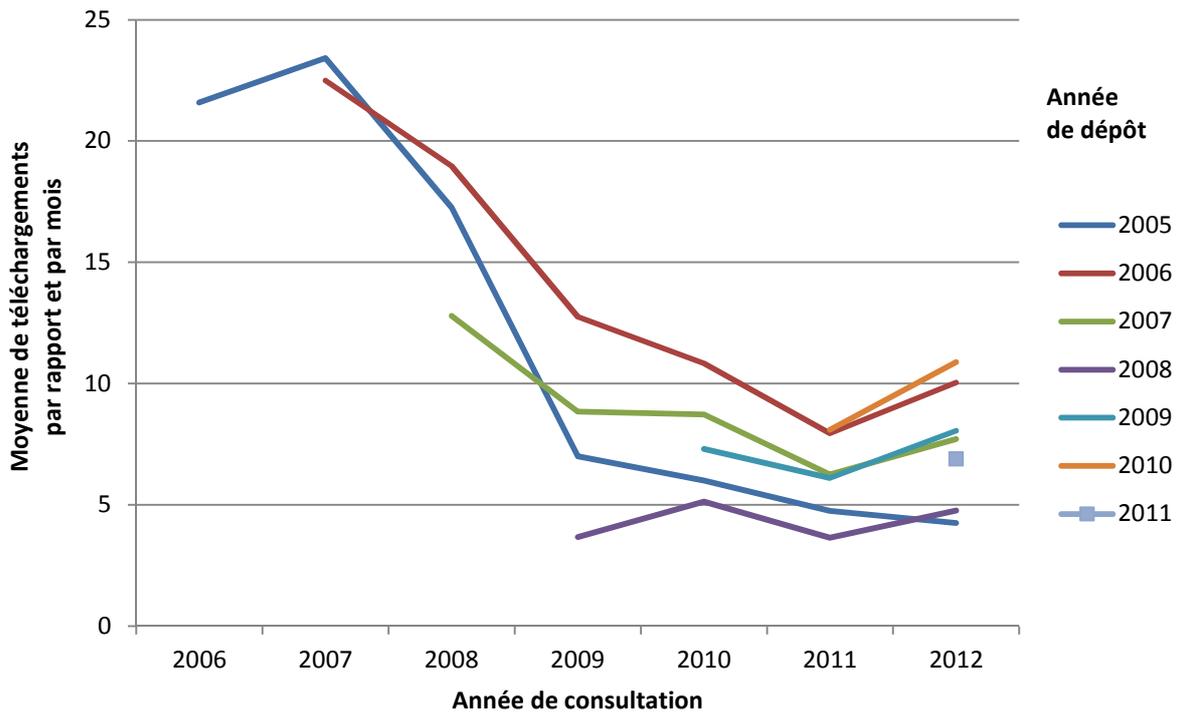
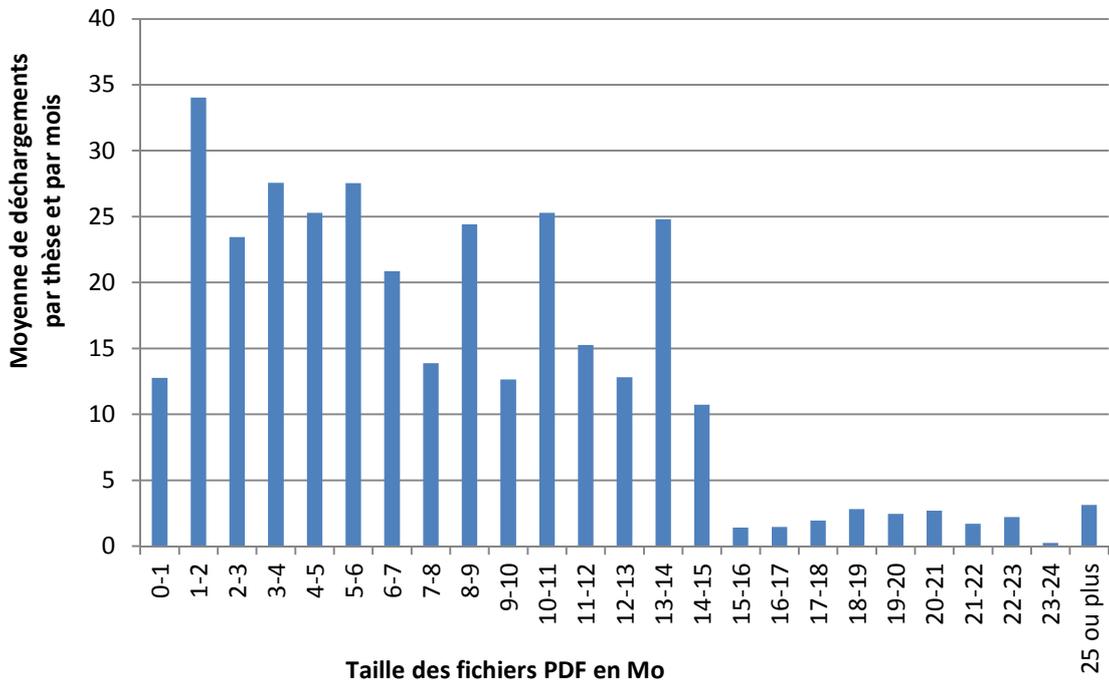
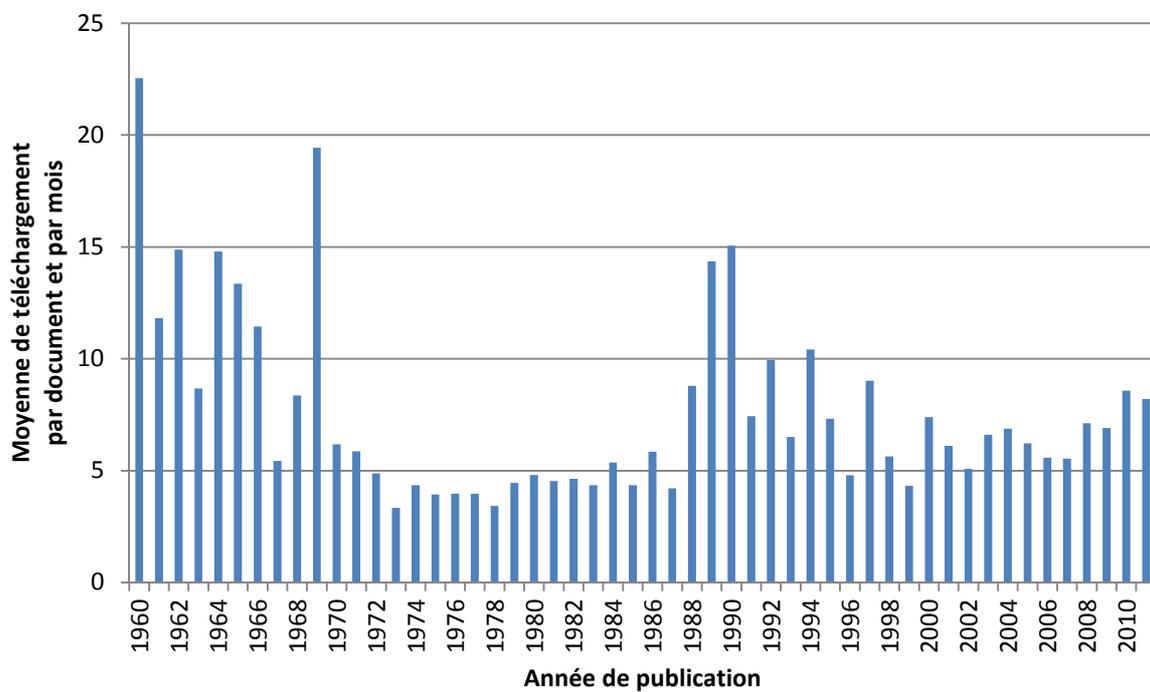


Figure 15 : Histogramme des moyennes mensuelles de téléchargements des thèses en fonction de la taille des fichiers au cours de l'année 2012



Remarque : La différence très marquée entre les téléchargements des fichiers de tailles inférieures et supérieures à 15 Mo pourrait être liée au fait que Google ne semblait pas indexer les fichiers volumineux (la limite semblait se situer autour de 10Mo à 15 Mo en fonction des années). Seules les fiches de métadonnées des gros fichiers étaient indexées. Cette limite semble avoir évolué en 2013 puisqu'il est désormais possible de trouver dans les index de Google des fichiers d'une taille supérieure à 60 Mo.

Figure 16 : Histogramme des moyennes mensuelles de téléchargements des documents³ en fonction de leur année de publication au cours de l'année 2012



2.3. Chemin d'accès

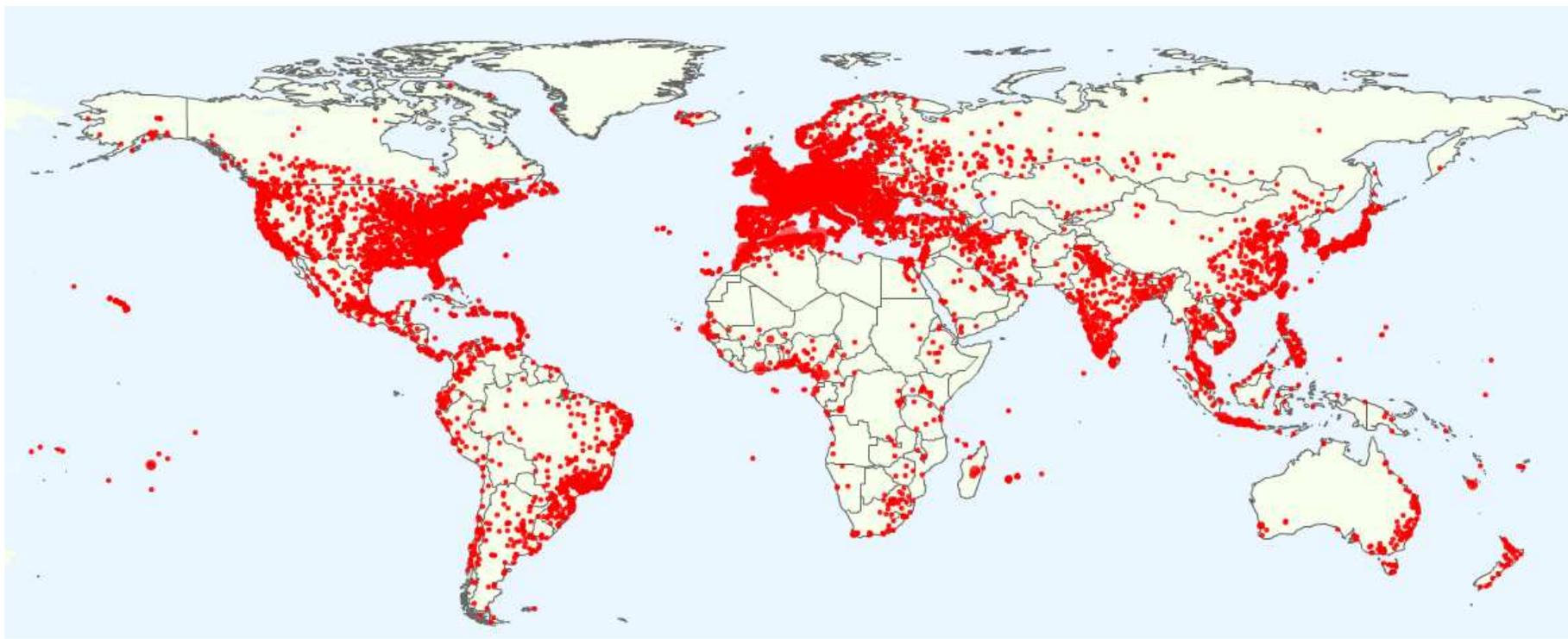
Tableau 1 : Sources d'accès aux documents téléchargés au cours du mois de Novembre 2012

Source des téléchargements	Nombre de téléchargements	Pourcentage des téléchargements
Google	57170	72,15%
Inconnu	9247	11,67%
Google Scholar	3222	4,07%
Archimer	1910	2,41%
Blogs, forums, lien dans des articles ⁵	710	0,90%
Gallica	581	0,73%
Bing	580	0,73%
Yahoo	504	0,64%
Ask	255	0,32%
Annuaire Ifremer	240	0,30%
WebMail	190	0,24%
Pages WEB Ifremer	178	0,22%
Wikipedia	150	0,19%
Facebook	105	0,13%
Yandex.ru	74	0,09%
Avano	68	0,09%
Baidu	61	0,08%
Sabiia	39	0,05%
eaufrance	39	0,05%
Bibliomer	37	0,05%
Twitter	34	0,04%
Driver	26	0,03%
Isidore	25	0,03%
123people	25	0,03%
ResearchGate	21	0,03%
opendoar	18	0,02%
worldcat.org	17	0,02%
academic.research.microsoft	13	0,02%
scientificcommons	10	0,01%
Autres	3684	4,65%

⁵ Exemples : connaissancesdesenergies.org, universcience.fr ..., bretagne-environnement.org, ...

2.4. Internautes

Figure 19 : Localisation des téléchargements enregistrés au cours de l'année 2012



Remarque : 10% des téléchargements 2012 ne sont pas localisés sur cette carte. Un point peut représenter jusqu'à 43 000 téléchargements.

Tableau 2 : Liste des principaux pays utilisateurs des documents rédigés en anglais

Pays	Nombre de téléchargements	Pourcentage des téléchargements
United States	31256	13,71%
France	23787	10,43%
India	15581	6,83%
United Kingdom	12569	5,51%
China	12012	5,27%
Germany	6714	2,94%
Spain	6571	2,88%
Canada	5714	2,51%
Italy	5263	2,31%
Indonesia	4955	2,17%
Iran, Islamic Republic of	4832	2,12%
Australia	4553	2,00%
Malaysia	4402	1,93%
Brazil	4209	1,85%
Japan	3957	1,74%
Mexico	3609	1,58%
Philippines	3376	1,48%
Portugal	3315	1,45%
Korea, Republic of	3263	1,43%
Algeria	3088	1,35%
Thailand	3081	1,35%
Netherlands	2702	1,18%
Tunisia	2490	1,09%
Turkey	2465	1,08%
Russian Federation	2400	1,05%
Taiwan	2280	1,00%
Belgium	2259	0,99%
Egypt	2138	0,94%
Norway	1840	0,81%
Chile	1751	0,77%

Tableau 3 : Liste des principaux pays utilisateurs des documents rédigés en français

Pays	Nombre de téléchargements	Pourcentage des téléchargements
France	296868	61,18%
Algeria	47266	9,74%
Tunisia	27651	5,70%
Morocco	26750	5,51%
Canada	11441	2,36%
Belgium	9141	1,88%
United States	6425	1,32%
Cote D'Ivoire	4078	0,84%
Cameroon	4078	0,84%
Senegal	3699	0,76%
Madagascar	3548	0,73%
Spain	3397	0,70%
Switzerland	3227	0,67%
United Kingdom	2703	0,56%
Italy	2130	0,44%
Benin	2088	0,43%
Germany	1723	0,36%
China	1594	0,33%
Portugal	1059	0,22%
Mauritania	949	0,20%
Brazil	935	0,19%
Burkina Faso	894	0,18%
Korea, Republic of	840	0,17%
Netherlands	764	0,16%
Russian Federation	754	0,16%
Congo	741	0,15%
Togo	711	0,15%
Gabon	673	0,14%
Mexico	633	0,13%

Tableau 4 : Liste des principaux utilisateurs académiques et privés des documents rédigés en anglais au cours du mois de Novembre 2012

Institution	Pays	Nombre de téléchargements	Pourcentage des téléchargements
Universite de Brest	France	415	0,18%
INRA	France	382	0,17%
ESMISAB	France	344	0,15%
Danish Network for Research and Education	Denmark	320	0,14%
Wageningen University and Research Centre	Netherlands	311	0,14%
Universite de Bordeaux	France	311	0,14%
Universiti Putra Malaysia Campus Network	Malaysia	305	0,13%
Universitaet Bremen	Germany	302	0,13%
Universite Pierre et Marie Curie	France	302	0,13%
Universitaet Kiel	Germany	291	0,13%
Universite de Caen	France	288	0,13%
University of Southampton	United Kingdom	275	0,12%
Taiwan Academic Network	Taiwan	258	0,11%
Universidade do Porto	Portugal	256	0,11%
The University of Plymouth	United Kingdom	250	0,11%
Bogor Agricultural University	Indonesia	245	0,11%
Université de Nantes	France	234	0,10%
Dalhousie University	Canada	233	0,10%
University of Tokyo	Japan	227	0,10%
Universite Montpellier II	France	223	0,10%
Stiftung Alfred-Wegener-Institut fuer	Germany	218	0,10%
Academic and research institutions	Norway	212	0,09%
TOTAL S.A.	France	207	0,09%
Ocean University Of Qingdao	China	203	0,09%
Universiti Sains Malaysia, Penang, Malaysia	Malaysia	198	0,09%
National Taiwan Ocean University	Taiwan	196	0,09%
RRM - Reseau de la Recherche a Marseille	France	195	0,09%
Ministry of Agriculture,Forestry and Fisheries	Japan	189	0,08%
Instituto Espanol de Oceanografia	Spain	188	0,08%
University of California, San Diego	United States	187	0,08%
Institut Recerca i tecnologia Agroalimentarias	Spain	187	0,08%
National Institute of Oceanography	India	187	0,08%
Universidade do Algarve	Portugal	181	0,08%
Ocean University of China	China	181	0,08%
Universite de La Rochelle	France	167	0,07%
University of Aberdeen, UK	United Kingdom	161	0,07%
Universitat de Barcelona	Spain	161	0,07%
University of Wales Bangor	United Kingdom	158	0,07%

Tableau 5 : Liste des principaux utilisateurs académiques et privés des documents rédigés en français au cours du mois de Novembre 2012

Institution	Pays	Nombre de téléchargements	Pourcentage des téléchargements
Algerian Academic Research Network	Algeria	2381	0,49%
Universite de Bordeaux	France	1340	0,28%
Universite de Caen	France	1071	0,22%
Universite de Nantes	France	1028	0,21%
INRA	France	1027	0,21%
Universite de Brest	France	998	0,21%
Universite de La Rochelle	France	876	0,18%
Universite Montpellier II	France	706	0,15%
Reseau de la Recherche a Marseille	France	692	0,14%
Universite Pierre et Marie Curie	France	651	0,13%
CNRST	Morocco	617	0,13%
ESMISAB	France	553	0,11%
Electricite De France	France	515	0,11%
Université de Rennes	France	504	0,10%
Ministere De La Sante Et Des Affaires Sociales	France	497	0,10%
Institut National de Recherches Agronomiques	France	396	0,08%
Museum national d'histoire naturelle	France	379	0,08%
Universite des Sciences et Technologies de Lille	France	358	0,07%
Commissariat a l'Energie Atomique	France	353	0,07%
Gaz de France	France	330	0,07%
University of Littoral - Dunkerque	France	315	0,06%
Universite de Liege (ULg)	Belgium	303	0,06%
Universite de Strasbourg	France	300	0,06%
CEMAGREF	France	299	0,06%
Universite du Quebec a Rimouski	Canada	289	0,06%
Universite de PERPIGNAN	France	278	0,06%
TOTAL S.A.	France	277	0,06%
Universite de Corse	France	268	0,06%
Montpellier Academy	France	260	0,05%
CIRAD	France	256	0,05%
Institut Polytechnique LaSalle-Beauvais	France	254	0,05%
Universite du Havre	France	252	0,05%
BRGM	France	246	0,05%
Universite Laval	Canada	238	0,05%
Assistance Publique-Hopitaux de Paris	France	237	0,05%
Universite de Nice Sophia Antipolis	France	235	0,05%
MINISTERE de L'INTERIEUR	France	230	0,05%
Universite du Maine	France	214	0,04%
Observatoire Oceanologique CNRS	France	213	0,04%
observatoire oceanologique de banyuls	France	212	0,04%
Ministere des Peches Maritimes	Morocco	200	0,04%
Universite de Tours	France	200	0,04%

3. Méthodologie

Les résultats présentés dans ce document sont issus de l'analyse des logs du serveur Web Apache de l'Ifremer. Toutes les nuits, un script informatique repère dans le fichier de logs, les lignes correspondantes aux téléchargements des fichiers PDF disponibles dans Archimer. Ces lignes se présentent sous la forme suivante :

```
xxx.xxx.xxx.xxx - - [22/May/2013:00:28:09 +0200] "GET /doc/00038/14916/12245.pdf HTTP/1.1" 200 409200
"http://scholar.google.it/scholar?q=phytoplankton+diversity+during+the+spring+bloom+in+the+northweste
rn+mediterranean+sea&hl=it&as_sdt=0&as_vis=1&oi=scholart&sa=X&ei=6_SbUcilKsTR7AbGxYGgCQ&sqi=2&v
ed=0CCwQgQMwAA" "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.31 (KHTML, like Gecko)
Chrome/26.0.1410.64 Safari/537.31"
```

Les informations ci-dessous sont extraites de chaque ligne :

- **L'adresse IP** (ex : xxx.xxx.xxx.xxx) permet d'obtenir des informations (FAI et coordonnées géographiques) sur l'identité de l'Internaute à l'aide de la base GeoIP⁶.
- **La date et l'heure du téléchargement** (ex : 22/May/2013:00:28:09).
- **L'URL du document téléchargé** (ex : /doc/00038/14916/12245.pdf) contient le numéro d'identification du document dans Archimer (ex : 14916).
- **Le code du résultat de la requête HTTP** (ex : 200)
- **La « Referring URL »** (ex : http://scholar.google.it/scholar?q=phytoplankton+diversity+during...) indique l'URL de la page qui est à l'origine du téléchargement. Elle correspond, dans notre exemple, à l'URL de Google Scholar avec, en paramètre, la question posée par l'internaute dans Google Scholar⁷ (ex : phytoplankton diversity during the spring bloom in the northwestern mediterranean sea). Si un internaute accède à un document via sa fiche de métadonnées (ex : <http://archimer.ifremer.fr/doc/00038/14916/>), le script remonte dans le fichier de log pour trouver la « Referring URL » de cette première requête. L'information « Referring URL » n'est pas systématiquement disponible. Les robots, notamment, exécutent souvent des requêtes HTTP sans fournir ce paramètre. Son absence peut donc aider à les identifier.
- **Le « User-Agent »** de la requête HTTP (ex : Mozilla/5.0 Windows NT 6.1 ...) fournit des informations sur le navigateur de l'internaute.

Les robots des moteurs de recherche et les aspirateurs de sites sont à l'origine de la très grande majorité des téléchargements. Plusieurs actions permettent de filtrer leurs activités :

- Les requêtes HTTP dont le « User-Agent » contient un des mots⁸ associés à des robots (BOT, SPIDER, CRAWLER ...) sont supprimées.
- Les téléchargements provenant de plusieurs FAI (ex : OVH, Microsoft Hosting ...) sont également exclues en l'absence de paramètre « Referring URL ».
- Un grand nombre de téléchargements exécutés à partir de la même adresse IP est systématiquement suspect : les requêtes sont alors vérifiées manuellement (au-delà de 100 requêtes par mois). Dans certains cas, ces téléchargements proviennent d'une université dont les appels sortant sont filtrés par un proxy, mais, dans la majorité des cas, il s'agit de robots dont les téléchargements sont alors supprimés.
- ...

⁶ <http://www.maxmind.com/fr/home>

⁷ Depuis 2012, les recherches dans Google ne sont plus visibles dans les requêtes HTTP quand les internautes sont connectés à leur compte GMail.

⁸ Une liste de robots à exclure est disponible sur le site du projet COUNTER : http://www.projectcounter.org/code_practice.html