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

Période 2005-2013

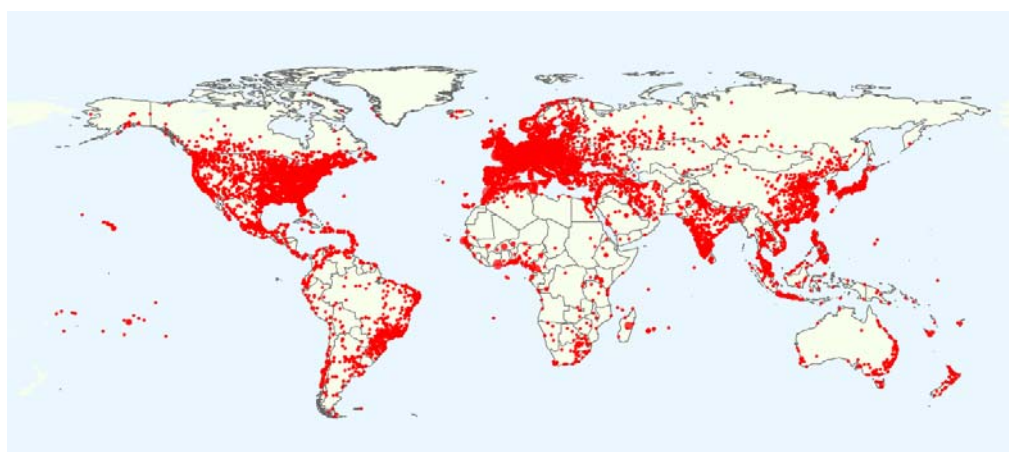


Table des matières

1. Introduction	2
2. Statistiques d'utilisation.....	3
2.1. Téléchargements.....	3
2.2. Chemin d'accès	10
2.3. Internautes.....	13
3. Dépôts	18
4. Méthodologie.....	20

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.

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)

2. Statistiques d'utilisation

2.1. Téléchargements

Figure 1 : Progression de la somme annuelle des téléchargements des documents³

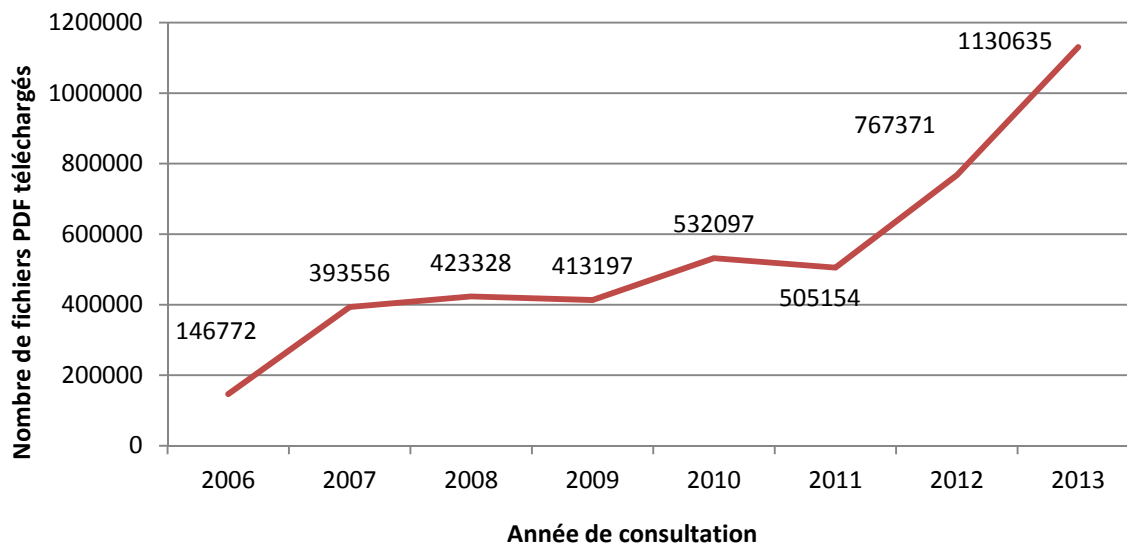
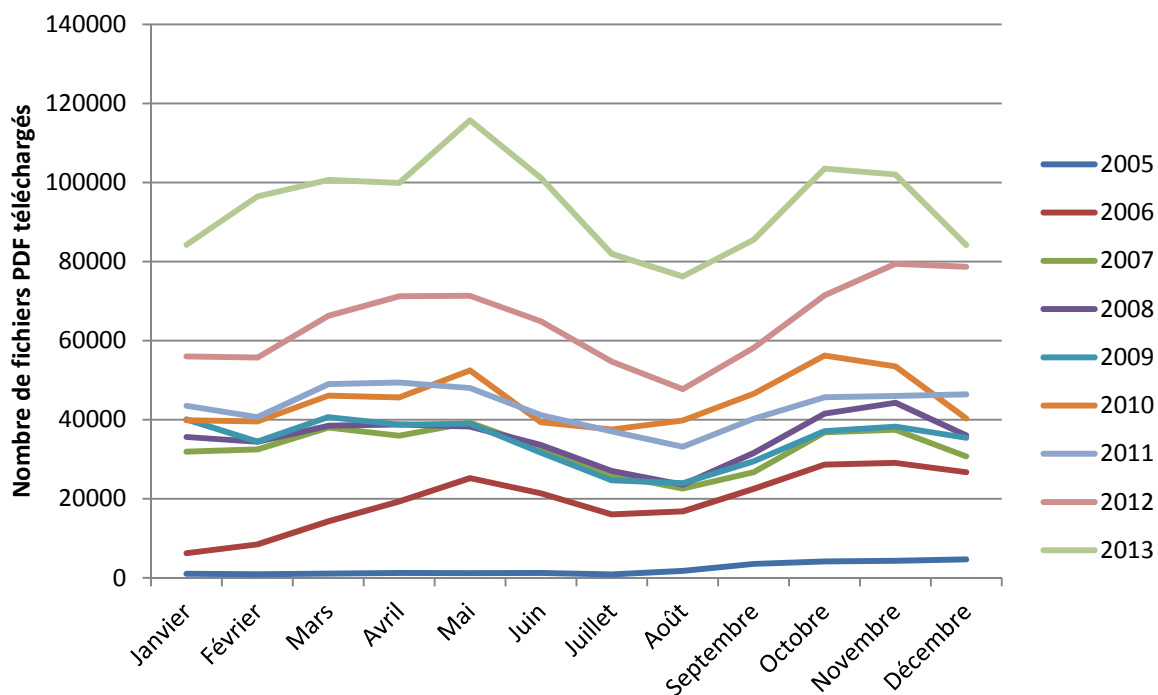
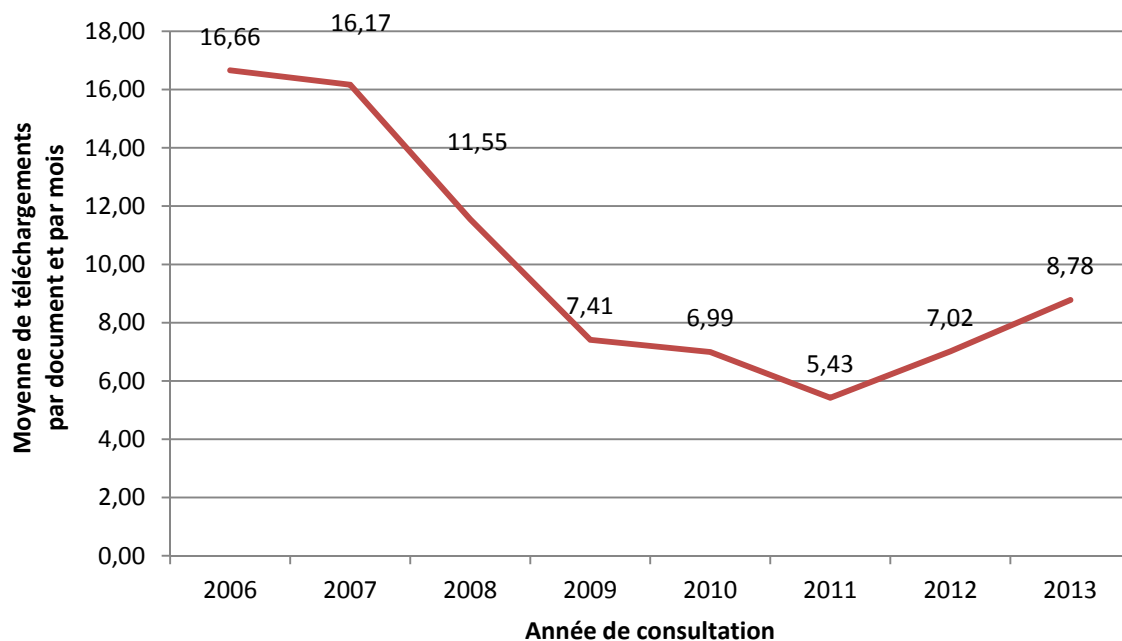


Figure 2 : Evolution de la somme mensuelle des téléchargements des documents



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

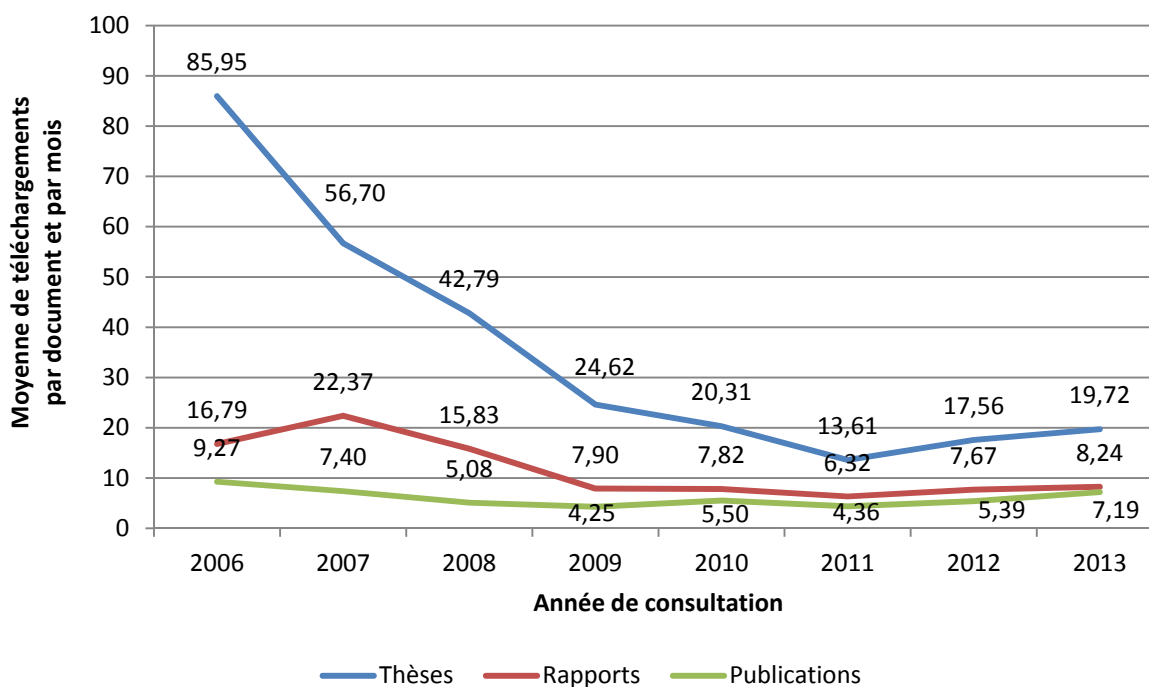
Figure 3 : Evolution de la moyenne mensuelle des téléchargements par document



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

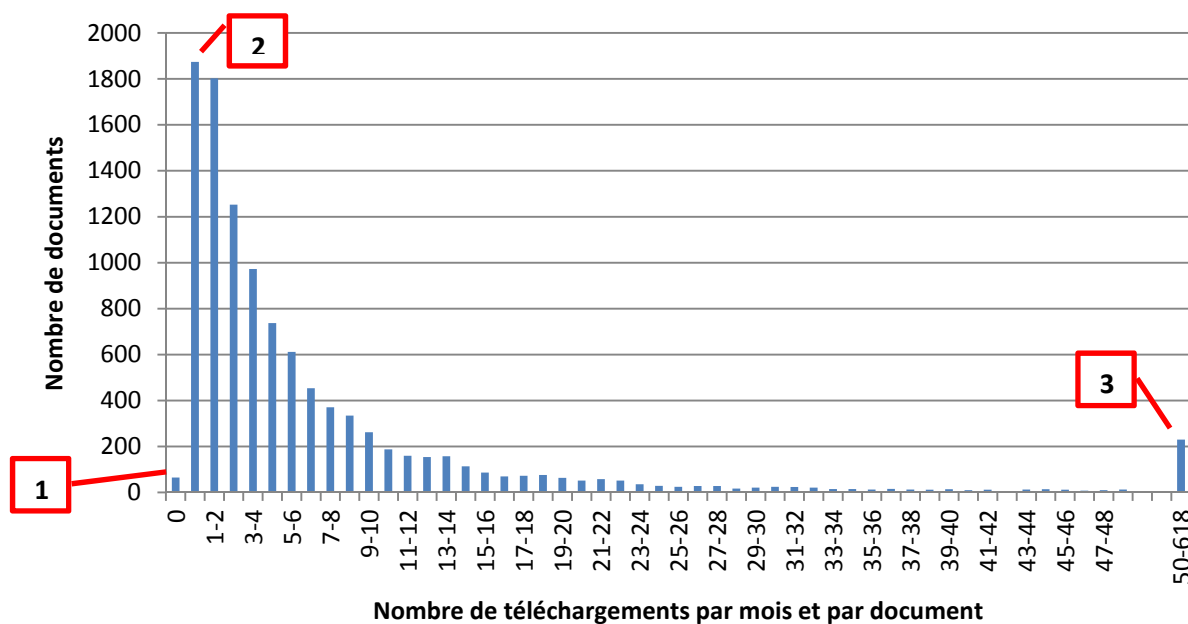
$$\frac{\text{Téléchargements enregistrés lors de l'année } n \text{ des documents déposés avant la fin de l'année } n - 1}{(\text{Nombre de documents déposés avant la fin de l'année } n - 1) \times 12}$$

Figure 4 : Evolution de la moyenne mensuelle des téléchargements des publications⁴, des thèses et des rapports



⁴ Publications indexées dans le WOS

Figure 5 : Histogramme des moyennes mensuelles de téléchargements des documents ^{Erreur ! Signet non défini.} au cours de l'année 2013



La Figure 5 se lit de la façon suivante :

1. 65 des 10732 documents déposés en Libre Accès avant le 31 décembre 2012 dans Archimer n'ont jamais été téléchargés au cours de l'année 2013
2. 1874 de ces 10732 documents ont été téléchargés de 1 à 11 fois au cours de l'année 2013. La moyenne mensuelle de téléchargements de ces 1703 documents se situe donc entre 0 et 1.
3. 230 de ces 10732 documents ont été téléchargés en moyenne entre 50 et 618 fois par mois au cours de l'année 2013.

Figure 6 : Histogramme des moyennes mensuelles de téléchargements des publications au cours de l'année 2013

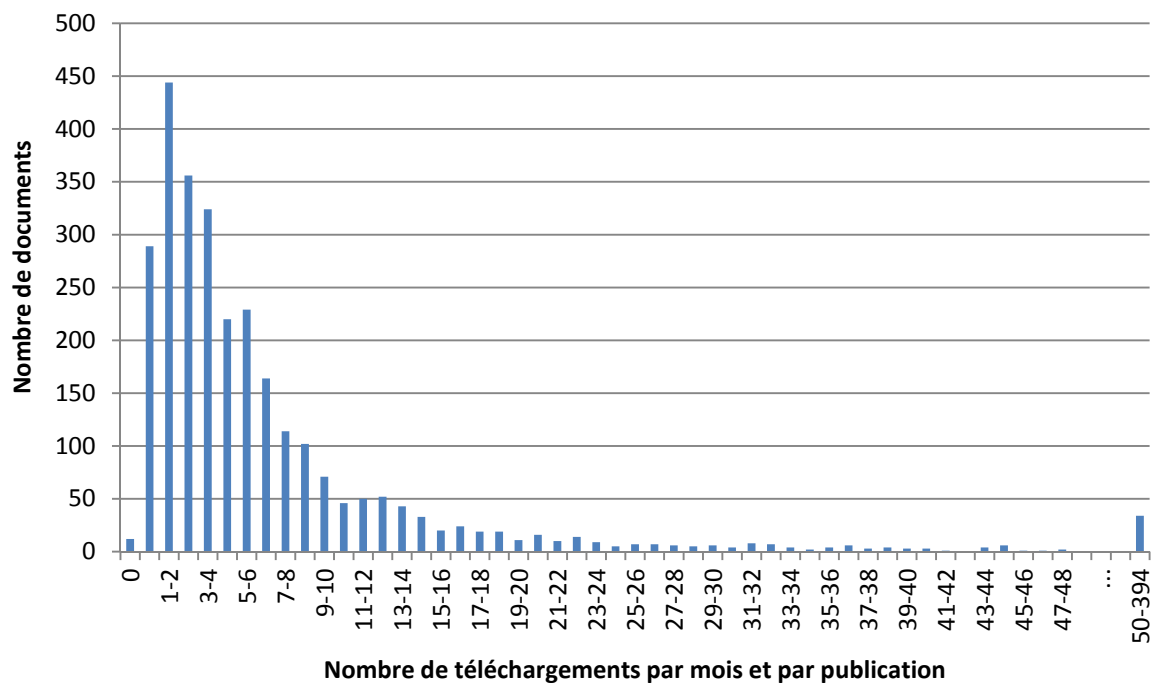


Figure 7 : Histogramme des moyennes mensuelles de téléchargements des thèses au cours de l'année 2013

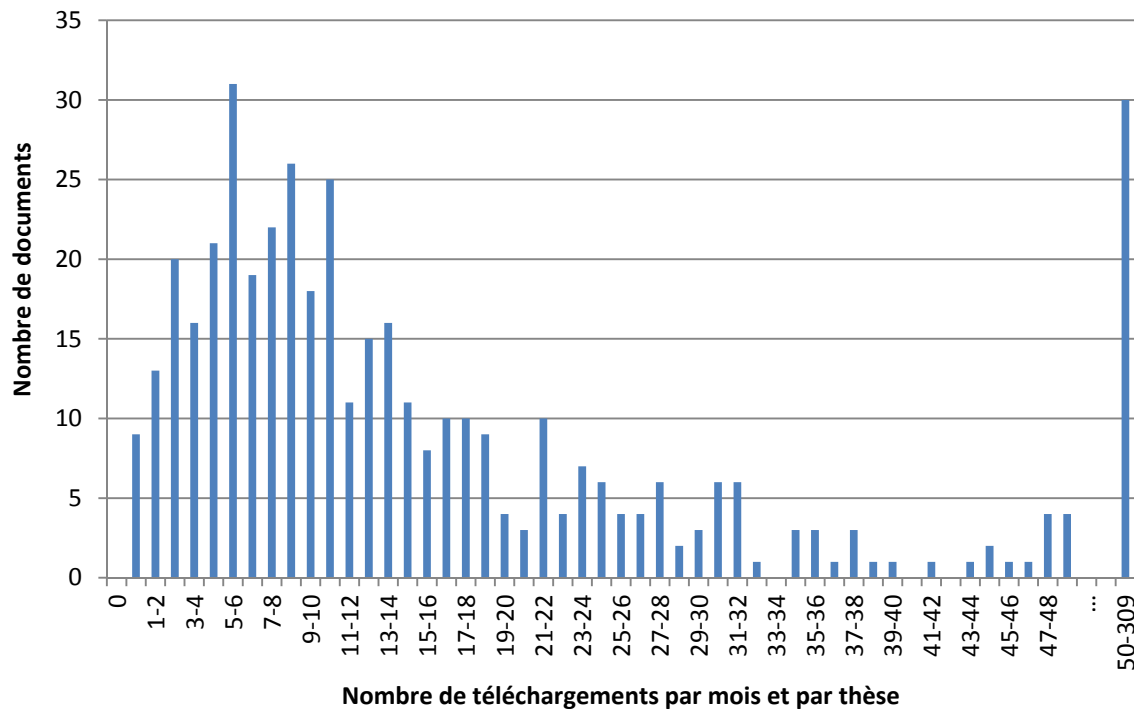


Figure 8 : Histogramme des moyennes mensuelles de téléchargements des rapports au cours de l'année 2013

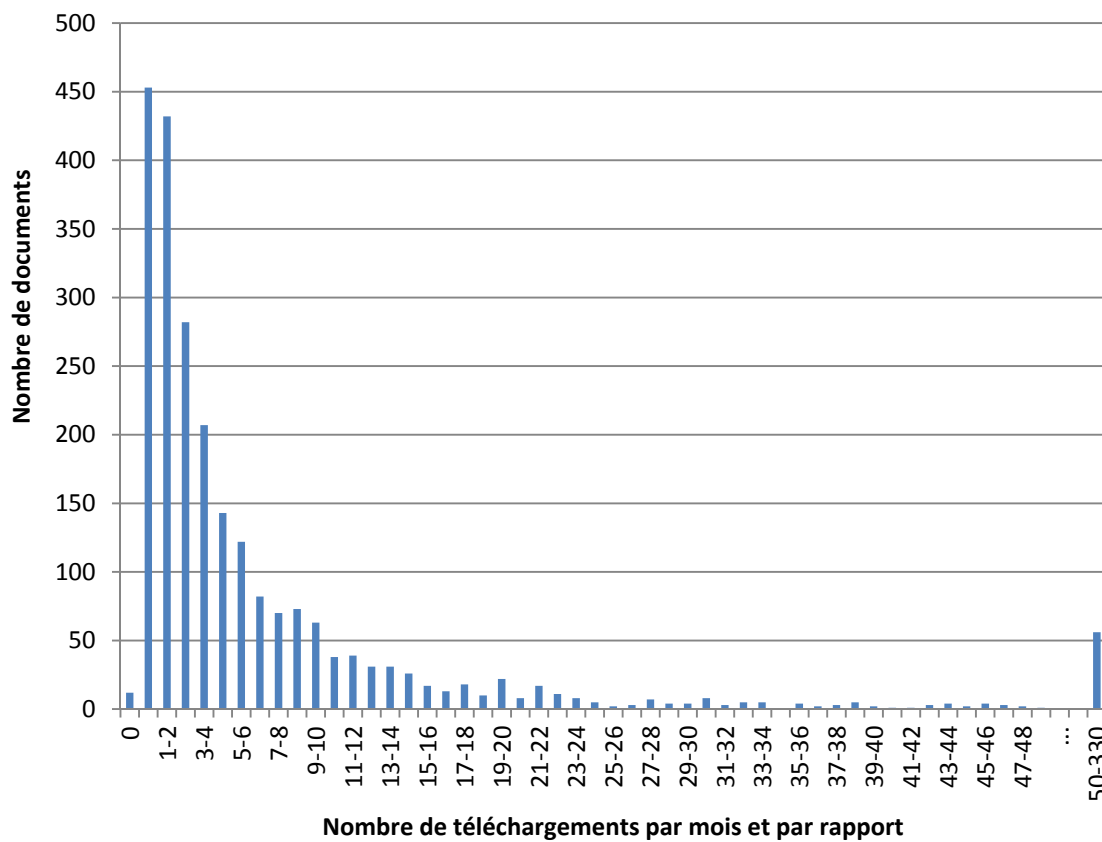


Figure 9 : Evolution de la moyenne mensuelle des téléchargements des thèses par année de dépôt

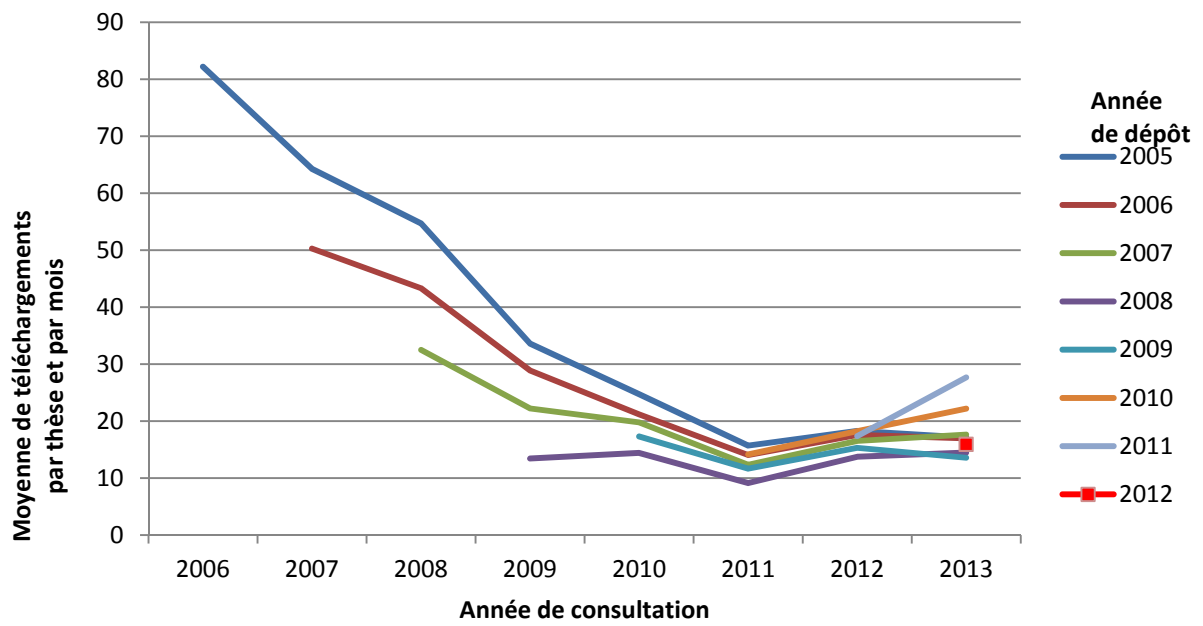


Figure 10 : Evolution de la moyenne mensuelle des téléchargements des publications par année de dépôt

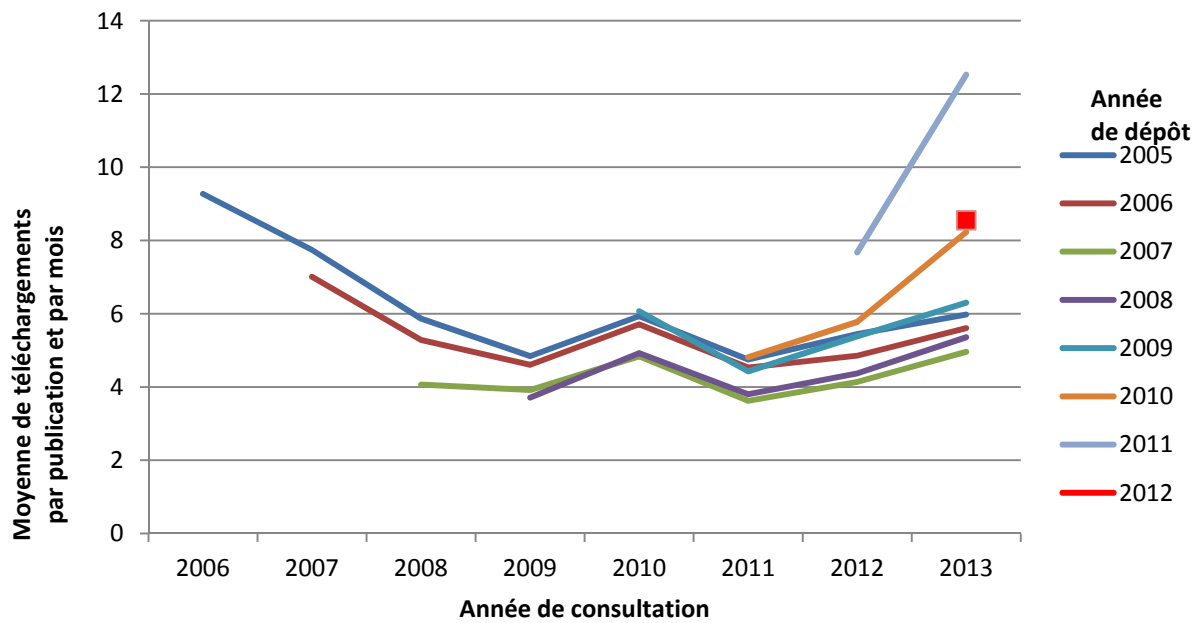


Figure 11 : Evolution de la moyenne mensuelle des téléchargements des rapports par année de dépôt

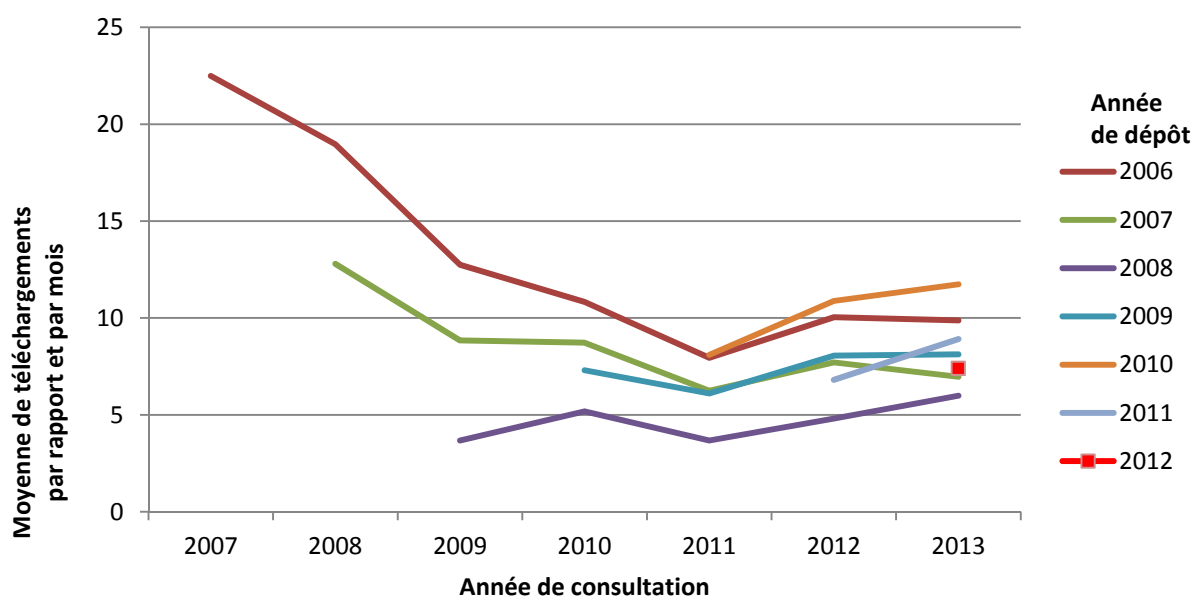
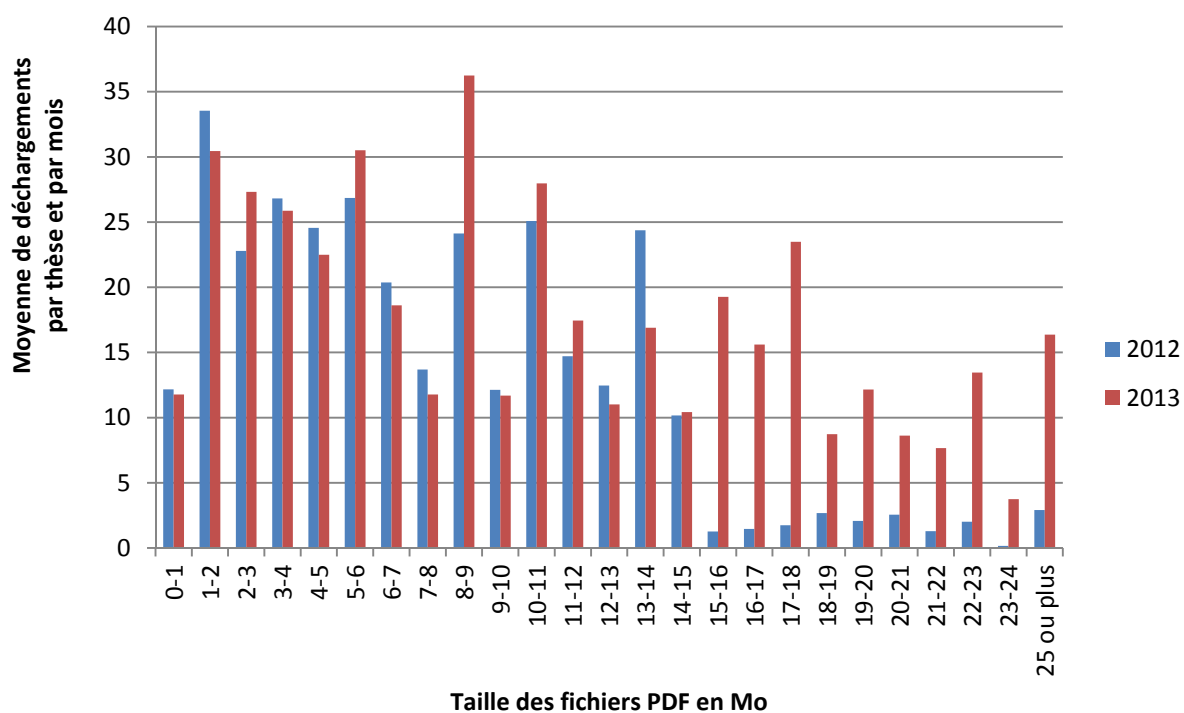
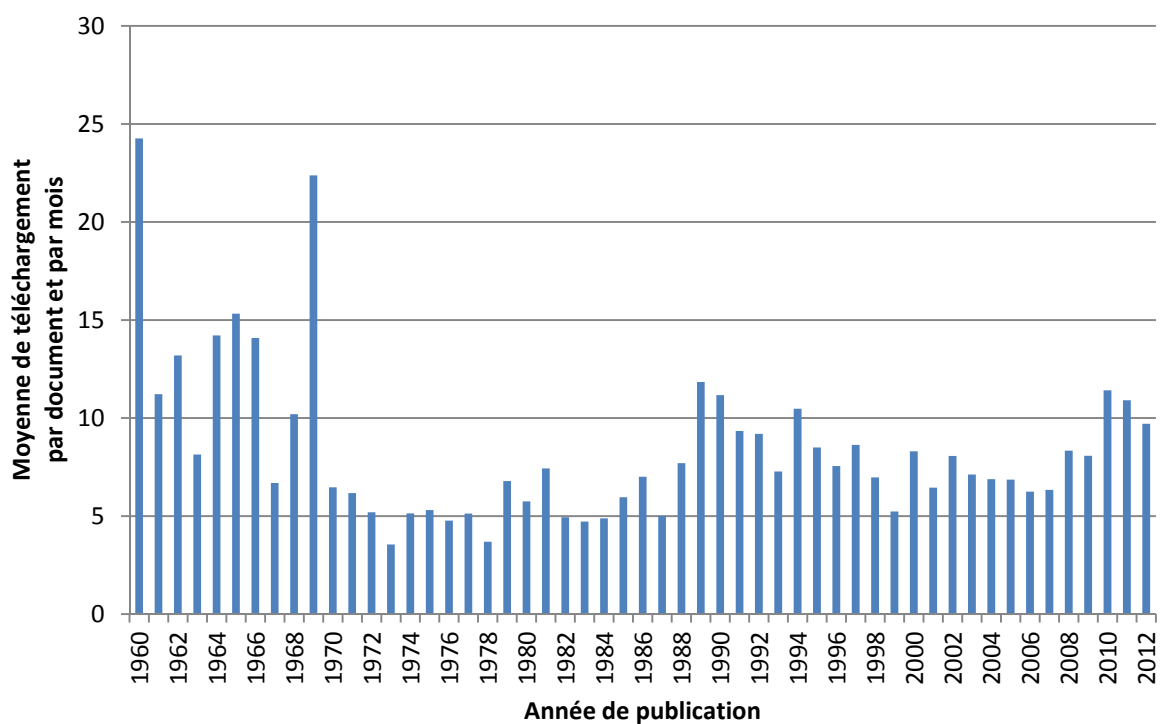


Figure 12 : Histogramme des moyennes mensuelles de téléchargements des thèses en fonction de la taille des fichiers



Remarque : La différence très marquée entre les téléchargements des fichiers de tailles inférieures et supérieures à 15 Mo a disparu en 2013. Elle était sans doute 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 13 : Histogramme des moyennes mensuelles de téléchargements des documents en fonction de leur année de publication au cours de l'année 2013



2.2. Chemin d'accès

Tableau 1 : Sources d'accès aux documents téléchargés au cours du mois de Novembre 2013

Source des téléchargements	Nombre de téléchargements	Pourcentage des téléchargements ⁵
Google	54384	80,38%
Google Scholar	5065	7,49%
Archimer	1474	2,18%
Yahoo	890	1,32%
Bing	743	1,10%
Blogs, forums, liens dans des articles	490	0,72%
Annuaire Ifremer	409	0,60%
pages WEB Ifremer	287	0,42%
Isidore	264	0,39%
Wikipedia	243	0,36%
Baidu	169	0,25%
documentation.eaufrance	163	0,24%
facebook	144	0,21%
...		
Gallica	56	0,08%
Twitter	49	0,07%
Yandex	21	0,03%
ResearchGate	15	0,02%
WorldCat	12	0,02%
academic.research.microsoft	4	0,01%
Autres	2773	4,10%

⁵ Pourcentage par rapport aux requêtes dont la source est identifiable via la *Referring URL*

⁶ Ce tableau ne liste pas un grand nombre de systèmes qui pourraient s'apparenter à du SPAM. Certains systèmes réindexent le contenu de documents disponibles dans Archimer et les diffusent sous un format différent (ex : banque-pdf.fr ...). D'autres s'apparentent à des barres de recherche qui interceptent les recherches des internautes (ex : babylon, conduit ...) avant de les rediriger vers Google.

Figure 14 : Nuages de mots recherchés dans Google et Google Scholar à l'origine de téléchargements de documents rédigés en anglais

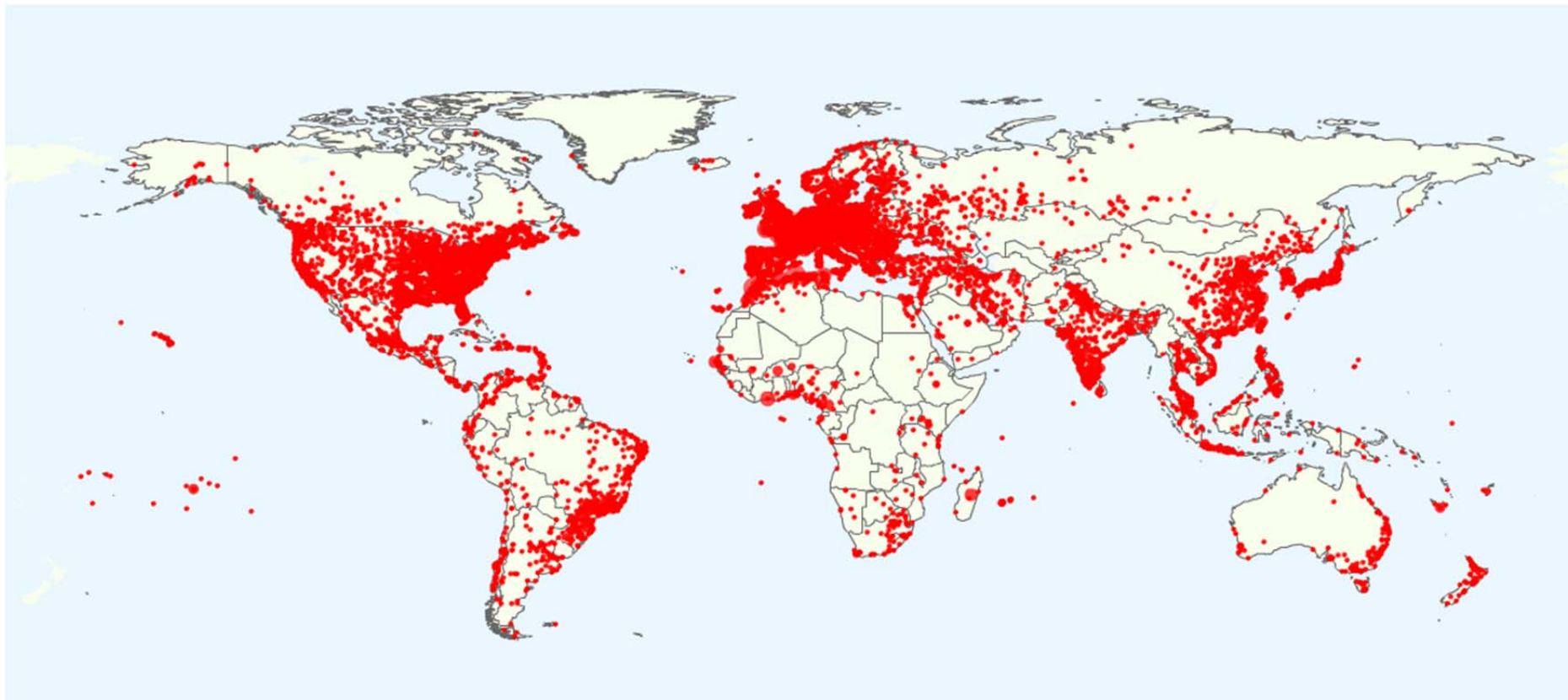


Traitement : Wordle (<http://www.wordle.net>)

Remarque : Google bascule progressivement tous ses domaines en https. Les recherches des internautes présentés dans ce nuage ne nous seront bientôt plus accessibles. La *Referring URL* de 30% des recherches dans Google restait exploitable en décembre 2013.

2.3. Internautes

Figure 16 : Localisation des téléchargements enregistrés au cours de l'année 2013



Remarque : 21% des téléchargements 2013 ne sont pas localisés sur cette carte. Un point peut représenter jusqu'à 38 000 téléchargements.

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

Pays	Nombre de téléchargements	Pourcentage des téléchargements
United States	58869	15,4%
France	30040	7,9%
China	28794	7,6%
India	25354	6,6%
United Kingdom	22566	5,9%
Canada	10948	2,9%
Germany	9225	2,4%
Spain	9136	2,4%
Australia	8436	2,2%
Italy	8339	2,2%
Indonesia	8199	2,1%
Malaysia	6867	1,8%
Brazil	6813	1,8%
Japan	6589	1,7%
Philippines	6538	1,7%
Iran, Islamic Republic of	6311	1,7%
Thailand	5105	1,3%
Mexico	5098	1,3%
Korea, Republic of	4845	1,3%
Turkey	4827	1,3%
Portugal	4654	1,2%
Netherlands	4591	1,2%
Algeria	3725	1,0%
Czech Republic	3660	1,0%

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

Pays	Nombre de téléchargements	Pourcentage des téléchargements
France métropolitaine	347773	51,7%
Algeria	75225	11,2%
Morocco	45736	6,8%
Tunisia	39008	5,8%
Canada	18903	2,8%
Belgium	13484	2,0%
United States	11540	1,7%
Cote D'Ivoire	9789	1,5%
China	9654	1,4%
Cameroon	7940	1,2%
Madagascar	6650	1,0%
Senegal	6518	1,0%
Reunion	5675	0,8%
Switzerland	4962	0,7%
Guadeloupe	3923	0,6%
United Kingdom	3895	0,6%
Spain	3875	0,6%
Benin	3402	0,5%
Martinique	3342	0,5%
Germany	2758	0,4%
Italy	2483	0,4%
New Caledonia	2338	0,3%
French Polynesia	2118	0,3%

Tableau 4 : Liste des principaux utilisateurs académiques et privés des documents rédigés en anglais au cours du mois de Novembre 2013

Institution	Pays	Nombre de téléchargements	Pourcentage des téléchargements
Danish Network for Research and Education	Denmark	588	0,15%
Taiwan Academic Network	Taiwan	580	0,15%
University of Southampton	United Kingdom	570	0,15%
Research network University of Ghent	Belgium	566	0,15%
Universitaet Bremen	Germany	472	0,12%
Korean Education Network	Korea, Republic of	464	0,12%
Wageningen University and Research Centre	Netherlands	452	0,12%
Universitaet Kiel	Germany	436	0,11%
Universite de Bordeaux, Departement TIC	France	415	0,11%
The University of Plymouth	United Kingdom	412	0,11%
Universiti Putra Malaysia Campus Network	Malaysia	382	0,10%
Dalhousie University	Canada	376	0,10%
Universite de Brest	France	376	0,10%
University of California, San Diego	United States	364	0,10%
Universidade do Porto	Portugal	354	0,09%
Universidade de Sao Paulo	Brazil	325	0,09%
Beijing Baidu Netcom Science and Technology Co., L	China	317	0,08%
Navy Network Information Center (NNIC)	United States	306	0,08%
University of Stirling	United Kingdom	305	0,08%
Algerian Academic Research Network	Algeria	303	0,08%
Universite Pierre et Marie Curie	France	285	0,07%
University of Wales Bangor	United Kingdom	285	0,07%
Prince of Songkla University	Thailand	277	0,07%
Instituto Politecnico Nacional	Mexico	274	0,07%
Laboratorios LABEIN	Spain	271	0,07%
ESMISAB	France	269	0,07%
RRM - Reseau de la Recherche a Marseille	France	269	0,07%
Stiftung Alfred-Wegener-Institut fuer	Germany	267	0,07%
Universiti Sains Malaysia, Penang, Malaysia	Malaysia	264	0,07%
Universidade de Aveiro	Portugal	258	0,07%
Universite Montpellier II	France	257	0,07%
Bergen University, Norway	Norway	255	0,07%
TOTAL S.A.	France	255	0,07%
Universidad Nacional Autonoma de Mexico	Mexico	254	0,07%
University of Tasmania	Australia	251	0,07%
University of Hawaii	United States	246	0,06%
Universidade do Algarve	Portugal	244	0,06%
Instituto Espanol de Oceanografia	Spain	244	0,06%
Universitat de Barcelona	Spain	242	0,06%
Norwegian University of Science and Technology	Norway	236	0,06%

Tableau 5 : Liste des principaux utilisateurs académiques et privés des documents rédigés en français au cours du mois de Novembre 2013

Institution	Pays	Nombre de téléchargements	Pourcentage des téléchargements
Algerian Academic Research Network	Algeria	3716	0,55%
Universite de Bordeaux	France	1834	0,27%
Universite de Caen	France	1365	0,20%
Houari Boumediene's University of Science & Techno	Algeria	1175	0,17%
Universite de Nantes	France	1123	0,17%
Universite de Brest	France	1033	0,15%
Universite de La Rochelle	France	1009	0,15%
Universite Montpellier II	France	981	0,15%
INRA	France	926	0,14%
Tizi University	Algeria	759	0,11%
Reseau de la Recherche a Marseille	France	654	0,10%
Universite Pierre et Marie Curie	France	592	0,09%
ESMISAB	France	590	0,09%
Ministere De La Sante Et Des Affaires Sociales	France	550	0,08%
University of Littoral - Dunkerque	France	503	0,07%
Universite de PERPIGNAN	France	494	0,07%
Unievrsité de Rennes	France	491	0,07%
Gaz de France	France	456	0,07%
Universite de Liege (ULg)	Belgium	445	0,07%
Universite de Strasbourg	France	413	0,06%
Museum national d'histoire naturelle	France	408	0,06%
Universite des Sciences et Technologies de Lille	France	378	0,06%
Universite du Quebec a Rimouski	Canada	376	0,06%
Universite Laval	Canada	368	0,05%
Universite Cadi Ayyad - Marrekech -	Morocco	365	0,05%
TOTAL S.A.	France	361	0,05%
Universite du Sud Toulon - Var	France	355	0,05%
Universite de Rouen	France	340	0,05%
Universite de Lorraine	France	331	0,05%
Electricite De France	France	327	0,05%
CNRST (Centre National pour la Recherche Scientifi	Morocco	322	0,05%
CEMAGREF Groupement de Bordeaux	France	315	0,05%
Institut Universitaire de Technologie de Corte	France	314	0,05%
CNRST (Centre National pour la Recherche Scientifi	Morocco	307	0,05%
Universite de Nice Sophia Antipolis	France	297	0,04%
Universite de Poitiers	France	295	0,04%
Institut Polytechnique LaSalle-Beauvais	France	293	0,04%
Assistance Publique-Hopitaux de Paris	France	268	0,04%

3. Dépôts

Figure 17 : Progression du nombre de documents disponibles dans Archimer

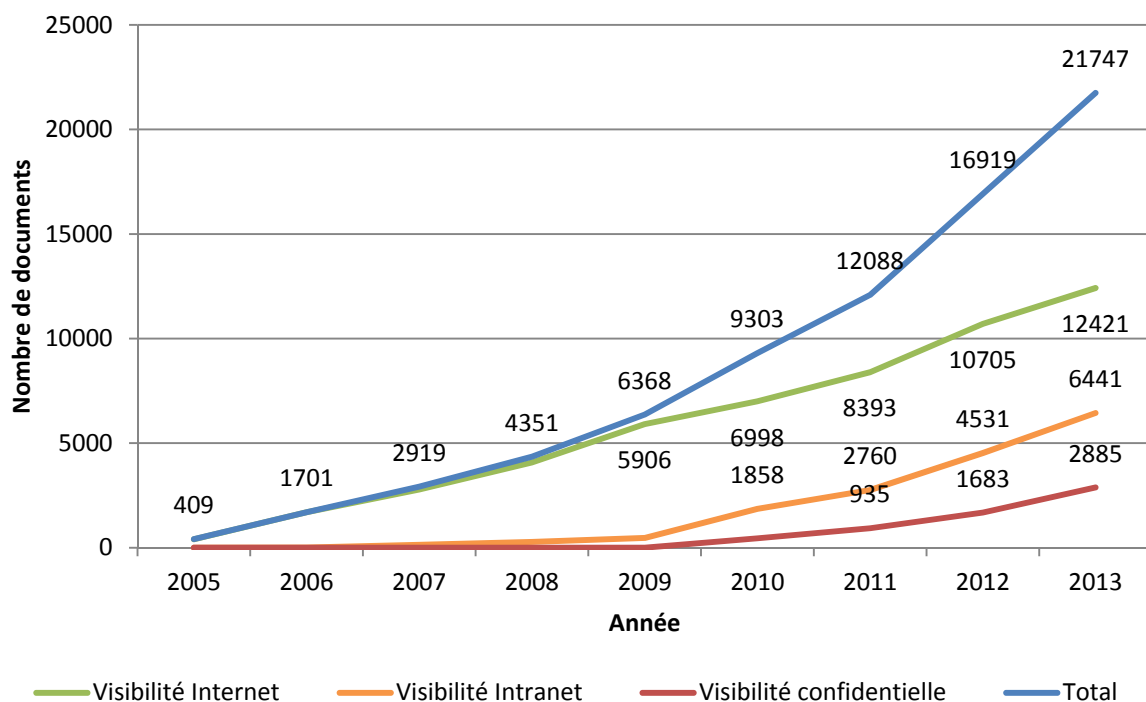


Figure 18 : Histogramme du nombre de documents déposés dans Archimer par année de dépôt

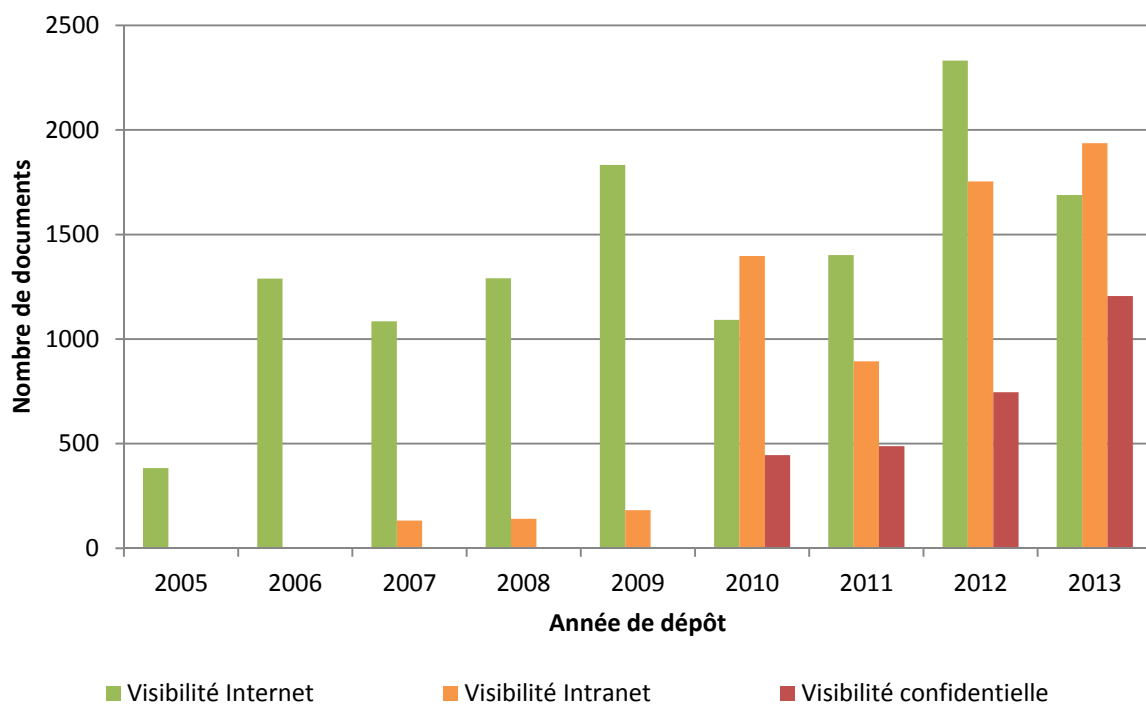
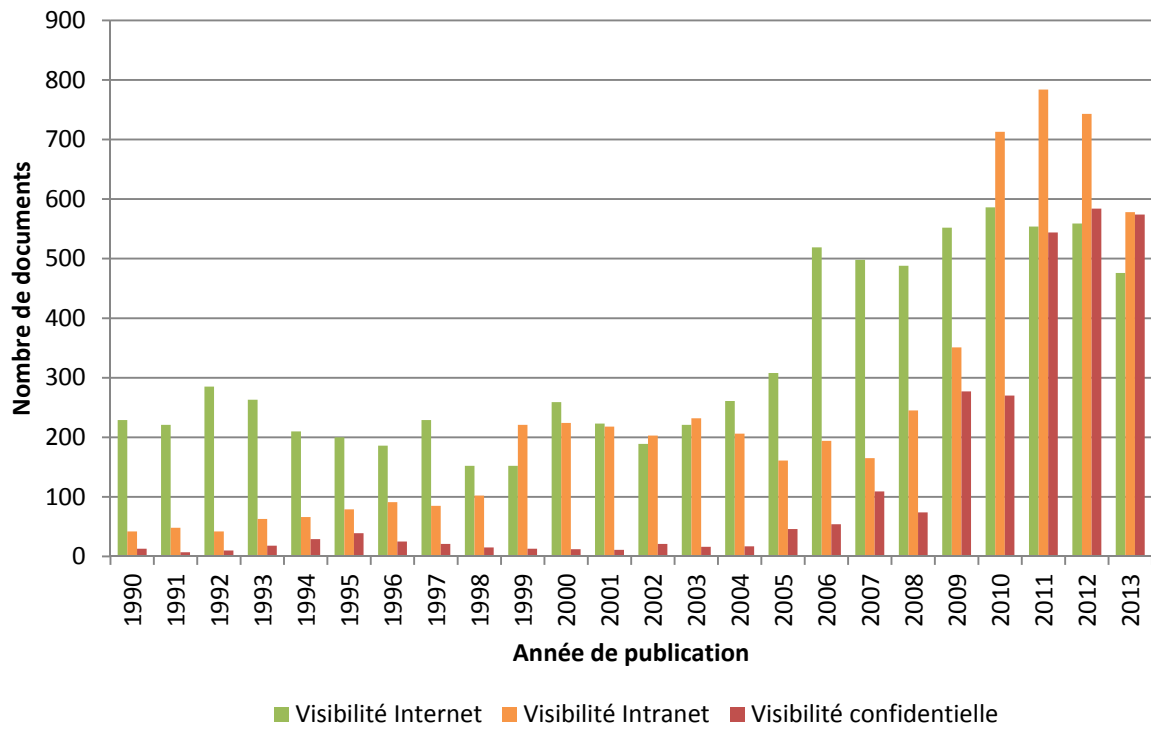


Figure 19 : Histogramme des années de publication des documents déposés dans Archimer



4. 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=scholar&sa=X&ei=6_SbUciKsTR7AbGxYGgCQ&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 GeolIP⁷.
- **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 Scholar (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.
- ...

⁷ <http://www.maxmind.com/fr/home>

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